Enhanced Representation and Learning of Magnetic Resonance Image Signatures in Multiple Sclerosis

Yogesh Karpate

ANNÉE (2015)

THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne

pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : Traitement du Signal et Télécommunications

Ecole doctorale MATISSE

présentée par

Yogesh Karpate

préparée dans l'unité de recherche UMR CNRS 6074
Nom développé de l'unité : VisAGeS-INSERM U746
Composante universitaire : ISTIC

Enhanced Representation & Learning of Magnetic Resonance Image Signatures in Multiple Sclerosis

Thèse soutenue à Rennes le 14/09/2015

devant le jury composé de :

Olivier COLLIOT
Rapporteur
Jean-Philippe RANJEVA
Rapporteur
Koen VAN LEEMPUT
Examinateur
François ROUSSEAU
Examinateur
Patrick BOUTHEMY
Examinateur
Gilles EDAN
Examinateur
Christian BARILLOT
Directeur de thèse
Olivier COMMOWICK
Co-directeur de thèse

Acknowledgments

I am profoundly grateful to my supervisors, Olivier Commowick and Christian Barillot. Their tireless pursuit of excellence in research, teaching, advising, and every other aspect of their academic work is truly inspirational. I am indebted to my supervisors for priceless advice about selecting interesting problems, making progress on difficult ones, pushing ideas to their full development, writing and presenting results in an engaging manner.

Many thanks to my friends who have had nothing to do with work in this thesis, but worked hard to keep my relative sanity throughout. I will not list all of you here, but my gratitude to you is immense.

My parents have given me unbending support and constant encouragement. I thank them for all the sacrifices they have made to ensure that their children would have better education they never had at their time. To my sister, Monali, my brother Rahul I am grateful for bringing me so much joy and love.

Contents

Chapter 1

Résumé en Français

Introduction

La sclérose en plaques (SEP) est une maladie neuro-dégénérative caractérisée par une évolution hétérogéne entre les patients. L'origine et l'évolution de la maladie sont encore mal comprises, et de nombreuses études ont ainsi été conduites afin d' évaluer cette évolution et l'influence de la SEP sur les tissus du cerveau environnants. La caractéristique principale de la SEP est la démyél inisation, c'est-à-dire la destruction progressive de la myéline entourant les axones et participant au transport de l'influx nerveux. Dans la SEP, de nombreuses lésions sont présentes en supplément des tissus sains. L'imagerie par résonance magnétique (IRM) joue un rôle crucial dans la clinique de la SEP, grâce â la possibilité de caractériser la progression spatiale des lésions et un éventuel dommage microstructurel. Une meilleure compréhension de la maladie â travers la découverte de ses mécanismes permettra de mieux adapter les thérapies afin de mieux soulager le patient.

Bien que la segmentation manuelle des lésions de SEP par des experts soit considérée comme la vérité terrain (gold standard), l'évaluation objective de celles-ci devient difficile pour le radiologiste lorsque le nombre de modalités d'IRM disponibles grandit. De plus, il s'agit d'une tâche peu reproductible et son processus est fastidieux. Ainsi, de nombreuses études ont considéré la possibilité d'une segmentation automatique ou semi-automatique des lésions de SEP utilisant pour cela plusieurs modalités IRM. Le besoin de telles nouvelles méthodes d'analyse d'image est très important afin de pouvoir fournir des mesures quantitatives des lésions de SEP aux radiologistes.

Défis

De nombreux radiologistes plaident pour l'évaluation objective de la charge lésionnelle comme critère diagnostic et de suivi, ce qui reste à accomplir via les méthodes de la littérature. Bien que les techniques de segmentations visent à produire des masques exacts des lésions, elles manquent parfois totalement certaines lésions. Dans de telles situations, la détection de lésions SEP et leur localisation plutôt que leur segmentation précise peut aider à détecter toutes

les lésions. Bien que de nombreux algorithmes aient étudié la segmentation de lésions SEP, ils ne prennent pas en compte tous les aspects de l'anatomie du cerveau en présence de lésions. Voici certains de ces défis, classés selon leur ordre de considération dans la littérature.

- **Segmentation en présence de volumes partiels**

  Lors de l'acquisition de l'IRM, un voxel comprend une résolution finie, de sorte qu'il peut contenir un ensemble de différents tissus. Ce phénomène est connu sous le nom de volumes partiels. La bordure d'une lésion peut ainsi être floue à cause de ce problème inhérent à l'acquisition. Ce phénomène a été modélisé dans le contexte de la SEP [Dugas-Phocion 2004]. La segmentation de lésions peut devenir difficile en présence de volumes partiels, spécialement pour les méthodes utilisant un modèle de distribution d'intensités de chaque tissu.

- **Méthodes reposant sur le recalage**

  Des atlas anatomiques peuvent être utilisés pour certaines méthodes de segmentation. Ils sont construits à partir d'IRM de volontaires sains. Ces atlas sont ensuite comparés aux images du patient et il est ainsi possible de traiter les lésions comme des points aberrants. Cependant, construire de tels atlas est une tâche complexe. De plus, ce type de méthodes introduit également le problème de recalage dans la segmentation de lésions SEP. Cette étape de recalage est d'autant plus difficile en présence d'une atrophie sévère du cerveau, d'un large nombre de lésions.

- **Déséquilibre de classes (méthodes supervisées)**

  Les méthodes de segmentation basées sur un apprentissage supervisé demandent de larges bases de données afin d'être efficaces. Dans le cas de la SEP, les deux classes sont inégales: la classe de tissus normaux est toujours largement plus nombreuse que la classe de lésions ce qui cause des problèmes à la plupart des algorithmes produisant des résultats biaisés vers la classe majoritaire (tissus sains). Ainsi, ceux-ci ont une performance faible sur la classe d'intérêt (lésions SEP) [Chawla 2005]. Ce problème de déséquilibre est souvent associé à des couts asymétriques dans la mauvaise classification des éléments dans les différentes classes. De plus, la distribution des données test peut différer de celle de la base de données d'entrainement et les coûts d'une mauvaise classification être inconnus au moment de l'apprentissage. Bien que connu, ce problème clé est toujours un problème ouvert et est souvent rencontré, particulièrement pour des bases de données massives. Une solution potentielle est la considération d'approches mono-classe, réalisant leur apprentissage

uniquement sur les tissus sains. Ces approches proposent une alternative intéressante aux approches discriminatives traditionnelles, dans lesquelles l'apprentissage est réalisé uniquement à partir des lésions [Tax 2004].

- **Données multicentriques**

  Lors de larges études IRM, des données de différents scanners sont utilisées. Celles-ci peuvent comporter différents contrastes et intensités et ce même si le même protocole d'acquisition est utilisé. Des méthodes doivent alors être développées afin de prendre en compte cette variation, et ce sans biaiser les études cliniques.

- **Atteinte diffuse**

  La majorité de la littérature se concentre sur l'analyse des lésions focales en SEP. Cependant, dans certains cas, il est impossible de trouver une frontiére claire entre les lésions et la matière blanche avoisinante. Aucune méthode n'a à ce jour considéré explicitement ces problèmes d'atteinte diffuse de la matière blanche.

- **Segmentation longitudinale de lésions SEP**

  Pour le moment, un nombre restreint de méthodes est défini spécifiquement dans le but de détecter les nouvelles lésions apparaissant dans le temps. Un challenge de segmentation de lésions SEP orienté spécifiquement sur cet aspect longitudinal s'est tenu à la conférence ISBI 2015 (International Symposium on Biomedical Imaging). [1]

Les techniques et algorithmes issus du domaine de l'apprentissage (machine learning) sont un outil puissant permettant de résoudre des tâches de classification. Ces techniques permettent également de définir des outils automatiques de prédiction d'un phénomène reposant sur des observations connues. De plus, l'objectif du machine learning est non seulement de fournir des prédictions les plus exactes possible mais également de fournir une compréhension de la structure des données. Ces méthodes sont particulièrement adaptées au domaine de la vision par ordinateur. L'état de l'art dans ce domaine considère de larges volumes de données pour la détection d'objets [Dalal 2005, Felzenszwalb 2010, Uijlings 2013] et les techniques de machine learning y sont donc particulièrement adaptées. Ces méthodes emploient des techniques de recherche de données avancées afin de trouver parmi des millions d'occurrences négatives les occurrences importantes qui sont ensuite utilisées pour entrainer un classificateur. Une approche de fenêtre glissante traite la détection d'objet comme un problème de classification, explorant plusieurs

[1] http://iacl.ece.jhu.edu/MSChallenge

échelles à chaque position de l'image afin de détecter si l'objet est présent. Un tel détecteur aura ainsi souvent plusieurs détections proches à différentes échelles pour le même objet.

Inspiré de ces méthodes, nous proposons un cadre permettant l'apprentissage discriminatif basé sur des patchs extraits des images, permettant de propager des annotations riches et nombreuses issues de modalités multiples afin d'entrainer un détecteur de lésions SEP. Ce cadre est proche des méthodes de segmentation dans sa philosophie de recherche voxel à voxel. Cependant, des différences existent: la segmentation a pour objectif l'obtention de contours précis; nous sommes plutôt intéressés ici par le problème de détection et de localisation de ces lésion en accordant moins d'importance à leur définition précise. Une telle détection ne conduit alors pas à une segmentation précise mais plutôt un masque d'intérêt au sein duquel les lésions sont localisées. Cette localisation des lésions SEP est cruciale, permettant de fournir au clinicien un chiffre important pour le diagnostique (nombre de lésions). Pour cette localisation, nous reposons sur une approche de fenêtre glissante se reposant sur l'information du patch pour discriminer l'existence d'une lésion. Ce type d'approche fonctionne mais est très couteuse car chacun des millions de pixels de l'image du patient doit être testé.

L'état de l'art des méthodes de segmentation fait également face à des problèmes d'extension à de larges bases de données en termes de temps de calcul, de robustesse et de complexité. Des algorithmes efficaces sont ainsi nécessaires afin de prendre en compte des représentations d'images de haute dimension et de permettre la recherche dans de grandes masses d'images. De plus, l'obtention des données d'entrainement peut être très consommatrice de temps, celles-ci requérant d'être annotées par un radiologiste. Ainsi, les méthodes de l'état de l'art doivent également faire un compromis entre les données d'entrainement labellisées requises et la flexibilité de l'apprentissage. Dans ce contexte, l'objectif de cette thèse est également de fournir un cadre basé sur un ensemble bien analysé et limité d'algorithmes.

Organisation de la thèse

Cette thèse est organisée en deux parties. La première partie consiste en trois chapitres et présente le contexte et le raisonnement derrière les travaux présentés. En particulier, le chapitre 2 présente l'architecture et l'organisation générale de la thèse. Le chapitre 3 aborde le contexte clinique de la sclérose en plaques et le rôle de l'imagerie médicale (IRM) dans la prise en charge des patients et leur diagnostic. Le chapitre 4 se focalise quant à lui sur une étude large de la littérature en segmentation de lésions de SEP. Ce chapitre fournit

également un préambule aux expérimentations réalisées dans la thèse.

Dans une seconde partie, quatre chapitres couvrent les contributions réalisées dans cette thèse ainsi que les expérimentations réalisées et leurs résultats. Le chapitre 5 présente une technique de normalisation d'intensité en IRM permettant d'aider dans le suivi de l'évolution de la SEP pour un patient à différents temps successifs. Se reposant sur cette technique de normalisation, le chapitre 6 présente une technique de détection de lésions SEP via un cadre statistique de comparaison entre le patient et une population de sujets contrôles. Ensuite, le chapitre 7 considère la détection comme un problème d'apprentissage à une classe (celle des tissus sains), considérant les lésions comme des éléments aberrants de cette classe. Enfin, le chapitre 8 conclue la thèse par des perspectives sur les travaux accomplis.

Contributions

Chapitre 3 (Contexte). Ce chapitre présente une introduction au contexte clinique et la physiopathologie de la sclérose en plaques. Nous y décrivons aussi l'importance qu'a prise l'imagerie par résonance magnétique (IRM) dans son étude et son diagnostic. Ainsi, les différentes séquences IRM utilisées en routine clinique sont étudiées, et les différents aspects et types de lésions visibles dans ces images sont investigués.

Chapitre 4 (Revue des techniques de segmentation de lésions). Ce chapitre présente une revue de l'état de l'art en segmentation de lésions. Cette revue catégorise les différentes techniques en plusieurs sous groupes et présente leurs avantages et inconvénients. Ainsi, une large variété de méthodes manuelles, semi-automatiques et automatiques est étudiée. Dans tous les cas, l'étape automatique de segmentation est très dépendante de la qualité des images d'entrée. Dans la mesure où elles requièrent un degré divers d'interaction humaine, les méthodes manuelles / semi-automatiques ne sont en général pas applicables à de larges bases de patients du fait du temps nécessaire et de la fatigue de l'utilisateur. Le paradigme de la segmentation totalement automatique a donc pris une large part des recherches effectuées. Les avancées récentes dans ce domaine ont montré la faisabilité de l'apprentissage de modèles précis pour la détection de lésions. D'un point de bue général, la segmentation automatique peut être classifiée en deux catégories:(1) supervisée et (2) non supervisée. Les cadres de segmentation supervisée sont principalement basés sur un apprentissage de patchs d'images. Ces méthodes comprennent les support vector machine (SVM), les arbres et forêts de décisions ou aléatoires, et des modèles graphiques probabilistes comme les conditional random fields (CRF). De récentes avancées dans le domaine de la parcimonie ont conduit au

développement de détection de lésion reposant sur un apprentissage de dictionnaire. A l'inverse, la segmentation non supervisée considère principalement des modèles génératifs des intensités des tissus sains du cerveau en IRM (mixture finie de modèles) et décrivent les lésions comme des points aberrants de ces modèles. D'autres méthodes non supervisées incluent des étapes ou éléments supplémentaires comme les MRF ou les Graph cuts. Une autre approche populaire utilise des atlas ou des méthodes de classification floues. La majorité de toutes ces approches considèrent l'information de multiples modalités (T1-w, T2-w et FLAIR) pour effectuer la détection ou la segmentation des lésions. Un aperçu rapide de la détection de lésion prenant le produit de contraste Gadolinium est enfin présentée, ces méthodes reposant pour la plupart sur des CRF. Afin de valider ces approches, un large nombre de métriques ont été envisagées, et incluent le score de Dice, la sensibilité, la valeur prédictive positive, la distance de Hausdorff. Malgré cette large littérature, de nombreux défis restent à aborder et ceux-ci sont présentés afin d'exposer le raisonnement du reste de la thèse.

Chapitre 5 (Normalisation longitudinale d'intensités). L'IRM conventionnelle ne fournit pas de valeurs quantitatives telles que les unités Hounsfield en tomographie. Ceci peut causer des problèmes à des applications de suivi des intensités dans les images acquises au cours du temps, notamment en détection ou segmentation, dans la mesure où ces opérations de post-processing dépendent des différences relatives d'intensité. Dans ce chapitre, nous proposons une technique de normalisation d'intensité pour des données multimodales d'IRM, et ce en étant robuste à la présence de lésions variables. Les intensités des IRM de multiples modalités sont modélisées via une transformation paramétrique et une estimation robuste d'un modèle de mixture de Gaussiennes, en utilisant la γ-divergence, non affectée par la présence de lésions. Ainsi, les intensités des lésions ne seront pas affectées par la normalisation. L'évolution de lésions peut alors être suivie par une simple soustraction d'images et un seuillage automatique de Otsu. Cette méthode est comparée à diverses techniques de l'état de l'art [Nyul 2000, Hellier 2003] sur deux jeux de données comportant respectivement 18 et 40 patients, chacun avec au minimum 3 points temporels. La distance du χ^2 pour le matching d'histogrammes et des courbes ROC sont utilisées pour l'évaluation. L'aire sous la courbe ROC est bien meilleure pour la méthode proposée comparée aux autres approches évaluées.La même méthode est appliquée à la détection de lésions prenant le contraste Gadolinium (lésions actives) en considérant la normalisation d'images pré- et post-contraste.

Chapitre 6 (Détection robuste de lésions de sclérose en plaques sur des images multimodales normalisées). Nous présentons ici une nouvelle méthode de détection de lésions de SEP basée sur une analyse voxel à

voxel comparant les images d'un patient par rapport à un ensemble de sujets contrôle. Ce chapitre a pour objectif d'étudier les bénéfices de l'imagerie multimodale afin de détecter des différences significatives entre un patient et la base de contrôles. Cet algorithme comprend deux parties. Tout d'abord, une normalisation d'intensité telle que décrite dans le chapitre précédent est effectuée afin de minimiser les différences inter-sujet liées à la variabilité de l'acquisition et aux différentes machines IRM. La seconde partie repose sur la définition d'un cadre statistique pour la comparaison multimodale du patient et d'un atlas construits à partir des images des contrôles. La méthode proposée est évaluée sur deux jeux de données constitués respectivement de 16 et 40 patients. Les détections obtenues ont été comparées pour diverses combinaisons de modalités, par exemple T1-w, T2-w ou FLAIR individuellement, et une combinaison des trois modalités. Les détections ont été également comparées avec et sans normalisation. L'aire sous la courbe ROC est employée comme élément d'évaluation et a démontré une meilleure performance avec normalisation et en utilisant la séquence T2.

Chapitre 7 (Détection de lésions par apprentissage d'une classe). L'apprentissage pour la détection n'utilisant qu'une classe peut ÃŁtre vu comme un sous ensemble spécifique des problèmes de classification à deux classes, mais où seule une des deux classes est disponible pour la phase d'entrainement. Dans le cas de l'imagerie médicale, l'obtention de ces données d'apprentissage est une tâche particulièrement longue et fastidieuse et de telles méthodes à une classe sont donc d'un intérêt certain. Il est relativement simple d'obtenir dans notre cas les données de la classe de tissus sains via l'utilisation de données de volontaires sains. Ainsi, ce chapitre présente un algorithme de détection de lésions SEP reposant sur cette approche. Les contributions sont ici doubles: (1) la construction d'un algorithme automatique et probabiliste permettant de discriminer les lésions des tissus sains, se basant sur une représentation simple des intensités des images (utilisées pour l'entrainement d'un classificateur probabiliste au niveau du voxel); (2) la génération d'une carte de probabilité à partir du classificateur pour déterminer pour un patient donné la probabilité d'occurrence d'une lésion, carte ensuite seuillée par un seuillage automatique d'Otsu afin d'obtenir les détections. Ce cadre a été évalué sur deux jeux de données de 16 et 40 patients respectivement. Notre analyse des résultats basée sur l'aire sous la courbe ROC révèle une bonne capacité de détection des lésions, et notamment meilleure que d'autres approches telles que la méthode du minimum du déterminant de la covariance ou encore les SVM à une classe.

Chapitre 8 (Conclusion et perspectives). Ce dernier chapitre résume les contributions de cette th‘ese et ouvre des perspectives et travaux futurs reliés à chaque chapitre.

Conclusion

Cette thèse adresse principalement le problème de représentation de l'image appliqué à la détection de lésions de SEP. Les études longitudinales en SEP deviennent de plus en plus prépondérantes et l'IRM y joue un rôle crucial dans le diagnostic et l'adaptation des thérapeutiques au patient et à son évolution. Dans un futur proche, il sera crucial de définir des représentations et des algorithmes de détection adaptés afin d'analyser précisément ces images. Les méthodes statistiques et de machine learning y jouent un rôle prépondérant. L'analyse de données d'IRM longitudinales sera également de plus en plus importante pour suivre l'évolution de la maladie. Dans ces objectifs, nous avons proposé de nouvelles méthodes de représentation d'images et de machine learning étendant les méthodes de l'état de l'art afin de mieux exploiter la distribution spatiale des intensités de l'image ou en fournissant un test statistique de voisinages très proches des pixels. Ces méthodes ont été éprouvées via des expériences sur plusieurs bases de données et ont démontré une meilleure performance que les méthodes issues de l'état de l'art.

Chapter 2

Introduction

Multiple Sclerosis (MS) is a disease with heterogeneous evolution among the patients. In MS, White Matter (WM) lesions are also present in addition to healthy brain tissues. The origin and evolution of this disease are still not well understood, and numerous studies have been conducted to evaluate its evolution and its influence on neighboring brain structures. Quantitative analysis of longitudinal Magnetic Resonance Images (MRI) provides a spatial analysis of the brain tissues which may lead to the discovery of putative biomarkers of disease evolution. Better understanding of the disease will lead to a better discovery of pathogenic mechanisms, allowing suitable therapies to alleviate patient's sufferings.

Although manual lesion detection by experts is the Gold Standard, the objective evaluation of lesions becomes difficult for the radiologist when the number of MR sequences grows dramatically. Consequently, several studies investigated the automatic/semi-automatic segmentation of MS lesions using multi-channel MR images. Therefore, there is a strong demand for automated MS lesions detection algorithms to assist radiologists.

Over the past decades, scientists have addressed problems of prediction by deriving theoretical frameworks from empirical studies or have learned prior knowledge in order to model, analyze and understand the phenomenon under study. For example, medical practitioners know from past experience that persons with high blood sugar are generally at high risk of diabetes. For an increasing number of problems however, standard techniques limit the success of the study under consideration. For example, identification and correlation of the genetic risk factors for cancers, where knowledge is still very sparse, is nearly impractical for the cognitive abilities of human beings. This is essentially because of the very high complexity and intricacies of interactions that exist among DNA. Similarly, for very fine-grained near future financial market forecasts, a large number of variables need to be taken into account, which quickly goes beyond the capabilities of experts to put them all into a mathematical model. To improve the reasoning and knowledge behind such problems and further advance science, machines of high speed and capacity have been built and designed since the twentieth century to assist humans in their calculations. With the advent of technology

in terms of hardware, developments in theoretical computer science, artificial intelligence and statistics have made machines to become more than calculators. Recent advances have made them experts of their own kind, capable to learn from data and to unleash by themselves the predictive structure of problems. Methodologies and algorithms that have stemmed from the field of statistics, machine learning have indeed now become a powerful tool for the analysis of complex and large data, successfully assisting practitioners and scientists in numerous breakthroughs in various fields of computer vision. The MS lesions and other tissues are often complex, and difficult to analyze. In this thesis, the complexity of the MS lesions and other brain tissues are brought forth with analysis, illustration and visualization. The objective is to simplify the processing of large patient databases and assist the radiologist in decision making, which would increase the statistical power of clinical trials. This thesis has endeavored to develop MS lesion detection techniques within a principled framework, based on the development of appropriate medical vision analysis and machine learning.

The thesis is organized as follows:

Chapter 3: Background

MS is one of the main causes for developing physical and cognitive disabilities in young adults both in developed and developing world. MRI has emerged as a non-invasive imaging technique that offers the possibility of visualizing the brain. Recently, it ha been identified as a biomarker tool for MS and used extensively in diagnosis, follow-up and prognosis. In this chapter, we introduce the role of MRI in MS. Formally, MS and its main characteristics are studied.

Chapter 4: Multiple Sclerosis Lesions Segmentation

MS lesion segmentation suffers from many practical and theoretical problems. Despite these problems, MS lesion segmentation research has made notable strides within the last decades. Recent advances have shown the feasibility of learning accurate models for detecting MS lesions. Within the last few years, thanks in part to work developing standardized benchmark databases and MS lesion detection challenges, researchers have set their sights on more complex problems that involve detecting MS lesions within realistic settings. This chapter presents a comprehensive review of the current state-of-the art methods as well as their strong and weak points. Further, it builds the rationale for the thesis.

Chapter 5: Longitudinal Intensity Normalization in Multiple Sclerosis

MRI lacks a standard MR-sequence dependent intensity scale like the Hounsfield units in computed tomography. This may hamper the subsequent applications of the acquired images like detection, segmentation and registration. Since these post processing operations may depend on intensity space of

the acquired images. Numerous approaches to cope with this problem were proposed recently. In this chapter we propose a new intensity normalization framework for longitudinal MR images and compare with two state-of-the-art standardization methods regarding applicability and accuracy. As a part of performance metrics, the χ^2 distance for histogram and longitudinal lesion detection are employed.

Chapter 6: Robust Detection of Multiple Sclerosis Lesions from Intensity-Normalized Multi-Channel MRI

In this chapter, to characterize MS lesions, we propose a novel paradigm to detect white matter lesions based on a statistical framework. It aims at studying the benefits of using multi-channel MRI to detect statistically significant differences between each individual MS patient and a database of control subjects. This framework consists in two components. First, intensity standardization is conducted based on the technique developed in chapter 4. The second part studies the comparison of multi-channel MRI of MS patients with respect to an atlas built from the control subjects, thereby allowing us to look for differences in normal appearing white matter, in and around the lesions of each patient.

Chapter 7: Probabilistic One Class Learning for Multiple Sclerosis Lesion Detection

This chapter presents an automatic algorithm for the detection of multiple Sclerosis lesions (MS Lesions) from multi-sequence magnetic resonance imaging (MRI). We build a probabilistic classifier that can recognize MS Lesions as a novel class, trained only on Normal Appearing Brain Tissues (NABT). Patch based intensity information of MRI images is used to train a classifier at the voxel level. The classifier is in turn used to compute a probability characterizing the likelihood of each voxel to be a lesion.

Chapter 8: Summary and Perspective This chapter deals with perspectives on the problems addressed in thesis and possible future directions.

Chapter 3

Background

Contents

Multiple sclerosis (MS) is a chronic progressive disabling auto-immune disorder of the central nervous system with considerable social impact and economic consequences. It is a major cause of non-traumatic disability in young adults [Sadovnick 1993]. The socio-economic costs of MS are high. They are higher than those for stroke and Alzheimer's disease because of the long disease course, its higher prevalence and incidence among young adults, the subsequent early loss of productivity because of physical disability, fatigue and co-morbidity. Managing MS is an ongoing process, beginning with the very first symptoms and continuing throughout the disease course.

We first introduce the role of Magnetic Resonance Imaging (MRI) in MS. The first part of thsi chapter describes MS. Next, we describe how and why MRI has emerged as a para-clinical tool in MS.

3.1 Introduction

MS involves an immune mediated process in which an abnormal response of the body immune system is directed against the central nervous system (CNS), which is made up of the brain, spinal cord and optic nerves. The exact antigen or target that the immune cells are sensitized to attack are still unknown,

which is why MS is considered by many experts to be immune-mediated rather than autoimmune [Weiner 2004].

Myelin is a material that forms a layer, the myelin sheath, usually around only the axon of a neuron. The formation of the myelin sheath is called myelination. Within the CNS, the immune system attacks myelin the fatty substance that surrounds and insulates the nerve fibers as well as the nerve fibers themselves. It is essential for the proper functioning of the nervous system. It is an outgrowth of a type of glial cell.

CNS myelin is produced by special cells called oligodendrocytes. It plays an instrumental role in proper functioning of CNS. The main purpose of a myelin layer (or sheath) is to increase the speed at which electrical impulses propagate along the myelinated fiber. Myelin decreases capacitance and increases electrical resistance across the cell membrane. Thus, myelination helps prevent the electrical current from leaving the axon. It has been suggested that myelin permits larger body size by allowing fast communication among distant body parts. Demyelination is the loss of the myelin sheath insulating the nerves, and is the hallmark of MS. When any part of the myelin sheath or nerve fiber is damaged or destroyed, nerve impulses traveling to and from the brain and spinal cord are distorted or interrupted, producing a wide variety of symptoms. Damaged myelin forms scar tissue (sclerosis), which gives the disease its name. The expression of pathological feature in the form of MS lesion is essentially demyelination phenomenon. Both the myelin sheath and the oligodendrocyte itself are impaired within lesions. Consequently, it leads to attacks by cells of the immune system that react with myelin-related sites, such as myelin basic protein. Immune attack consists of two types of attacks upon: (1) T cells directed at myelin and oligodendrocytes in cellular immunity thus leading to phagocytosis by macrophages; (2) humoral immunity, with the secretion of anti-myelin antibodies from B cells and subsequent fixation of complement of the myelin sheath along with possibility of making it vulnerable to phagocytosis and the oligodendrocyte by macrophages [Brück 2005]. Figure 3.1 depicts the demyelinated neuron to be the primary consequence of MS. The disease is thought to be triggered in a genetically susceptible individual by a combination of one or more environmental factors [Ebers 2008].

3.1.1 Causes

The cause of MS is not fully understood; however, it is believed to occur as a result of some combination of environmental factors such as infectious agents and genetics [DH 2005]. MS affects more than 2.3 million people worldwide. While the disease is not contagious or directly inherited, epidemiologists have identified factors in the distribution of MS around the world

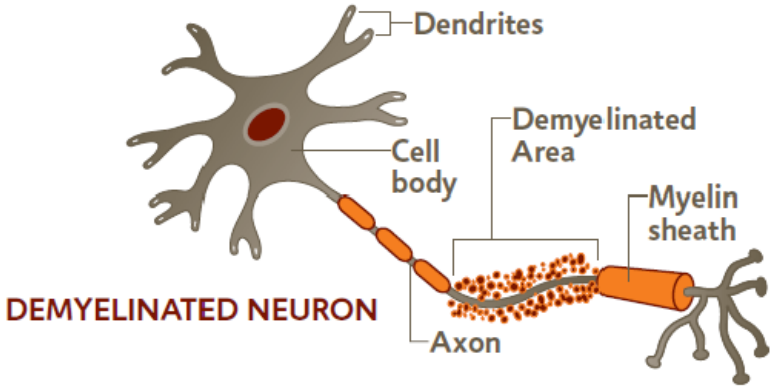


Figure 3.1: Image showing the demyelinated axons. Courtesy:http://www.nationalmssociety.org/

that may eventually help determine what causes the disease. These factors include gender, genetics, age, geography and ethnic background. MS is not considered a hereditary disease; however, a number of genetic variations have been shown to increase the risk [Dyment 2004]. The probability is higher in relatives of an affected person, with a greater risk among those more closely related. Specific genes that have been linked with MS include differences in the human leukocyte antigen (HLA) system a group of genes on chromosome 6 that serves as the major histocompatibility complex (MHC) [Compston 2008]. The abnormal deviation of gene expression in the MHC and HLA region are related to has been known identified as one of the main factor for susceptibility of the disease. Females are more often affected than males by a ratio of 2:1 [Mumford 1992], and age at clinical onset is typically between 20 and 40 years of age. The risk factors include the alleles associated with major histocompatibility complex molecules, infectious agents such as the EpsteinBarr virus and Chlamydia pneumoniae, lack of exposure to sunlight and vitamin D, and smoking [Levin 2005, Lincoln 2005, Pekmezovic 2006]. Different populations and ethnic groups have a different prevalence of MS. In France alone, there exists 94.7 MS patient per 100000 people [Fromont 2010]. The spatial distribution of MS patients density around the world is shown in Figure 3.2. MS is more common in people who live farther from the equator. MS is more common in regions with northern European populations and the geographic variation may simply reflect the global distribution of these high-risk populations.

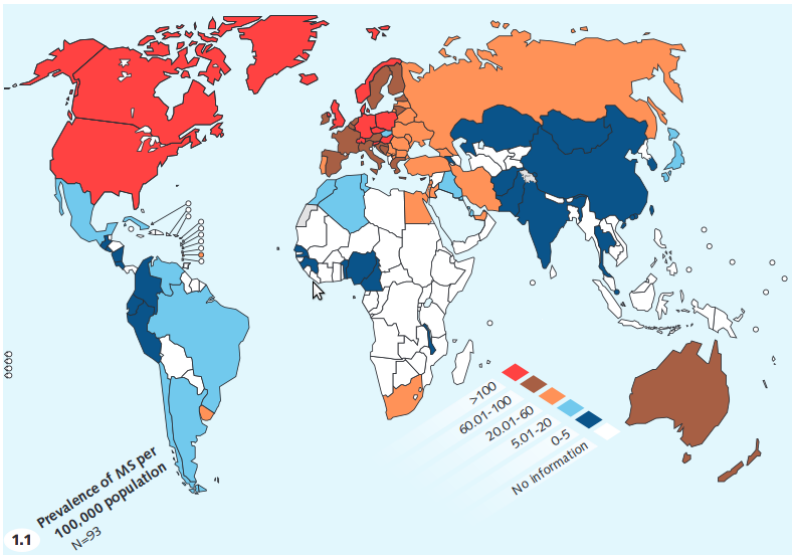


Figure 3.2: High regional prevalence of MS across the world. Courtesy : [Pietrangelo 2015]

3.1.2 Disease Course and Clinical Subtypes

A survey of international MS experts proposed the guidelines to describe types of the disease [Lublin 1996]. Patients usually suffer a first neurological event suggestive of MS known as Clinically Isolated Syndrome (CIS). It lasts for at least a day, with symptoms and signs indicating either a single lesion (monofocal) or more than one lesion (multifocal) within the central nervous system. Four types of MS have been established: relapsing-remitting MS (RRMS), primary-progressive MS (PPMS), secondary-progressive MS (SPMS), and progressive-relapsing MS (PRMS). Each of these disease courses might be mild, moderate or severe. Around 85% of patients initially suffer from a relapsing remitting disease type. RRMS consists of clearly defined disease relapses with full or partial recovery and no further progression of disease between relapses. Inflammation and lesion formation are likely to be the precursor to relapses. Of those patients with RRMS, most will go on to develop a progressive form of the disease within an average of 20 years [Vukusic 2003]. This is called SPMS and these patients may have occasional superimposed relapses, minor remissions and plateaus during the progressive phase. PPMS describes the 15% of people who have a progressive form of the disease from onset with gradual but almost continuous worsening of disability and only occasional plateaus and temporary minor improvements in function. New inflammatory lesions are seen less in the progressive stages of the disease. A progressive relapsing disease course has also been described and is seen in a

minority of patients. It is characterized by progressive disease from onset, but with clear acute relapses and continued progression between relapses. The progression of MS subtypes is shown in Figure 3.3.

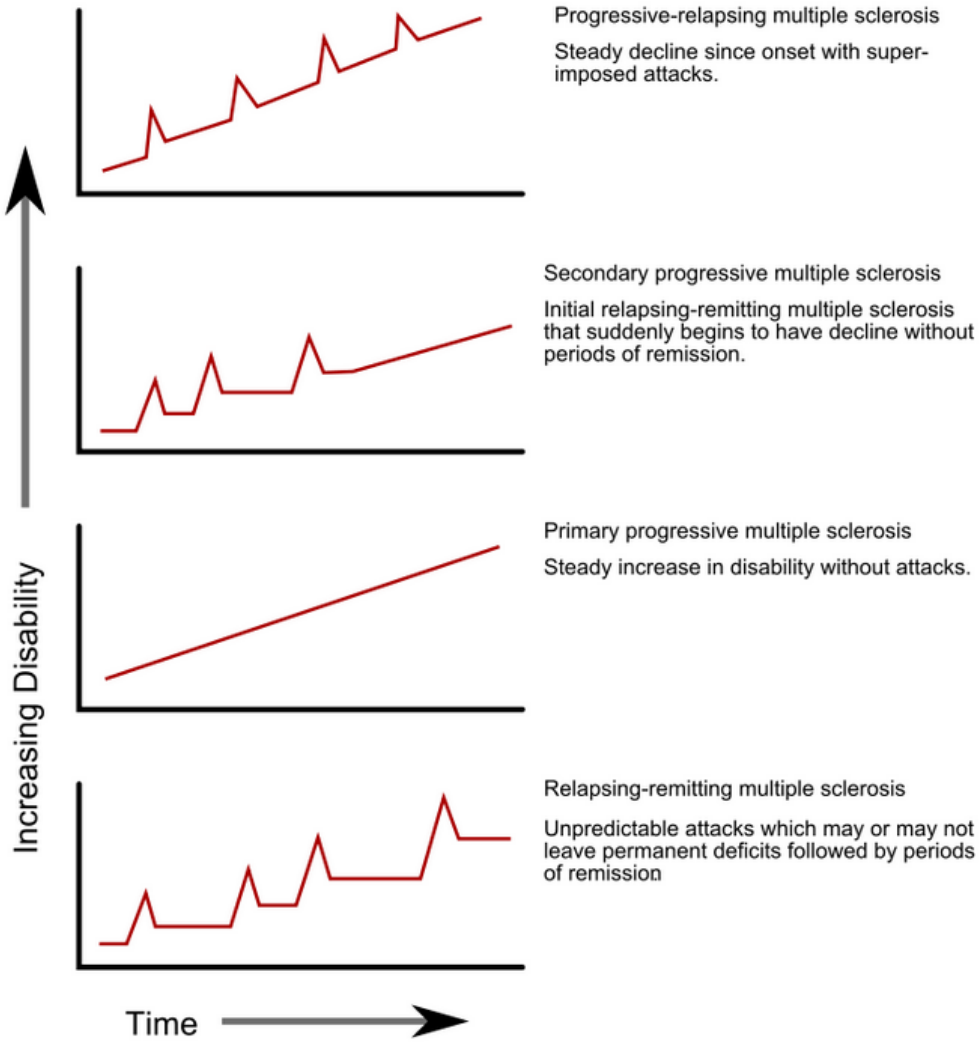


Figure 3.3: The progression of MS subtypes. Courtesy: http://en.wikipedia.org/wiki/Multiple_sclerosis

3.2 Role of MRI in MS

In order to acquire MRI, the patient is positioned inside a MRI machine which uses a strong magnetic field around the area to be imaged. In most medical applications, protons (hydrogen atoms) in tissues containing water molecules are used to create a signal that is processed to form an image of the tissues.

With the advent of technology, sophisticated techniques added to MRI made it possible to detect MS lesions earlier and with more precision than ever. It has emerged as a key principal tool in the diagnosis of MS and is increasingly used in studies seeking to monitor disease progression. It is a non-invasive technique and does not utilize ionizing radiation like computerized tomography, making it more practical for repeated examination of people with a condition that may persist for long years. Technological advances of MR in recent years have dramatically improved our understanding of MS disease. MRI scanning gave researchers faster and more sophisticated ways of testing drugs to treat MS. The benefits of a new drug can be seen on MRI scans before they can be seen in patients themselves [Ge 2006].

3.2.1 MRI Sequences

Different MRI sequences provide different information on the anatomy. The following sequences are complementary and routinely used in clinical practice:

1. A T1-w brain MRI scan, enhanced with gadolinium (injected intravenously for enhancement of scan), provides prognosis of disease activity by highlighting areas of active inflammation. Gadolinium (Gd) is a contrast agent that marks blood activities. Due to the large size of the Gd molecule, it cannot normally penetrate the blood-brain barrier. The blood-brain barrier is a highly selective permeability barrier that separates the circulating blood from the brain extracellular fluid in the central nervous system. However, in case of active inflammation, the blood brain barrier is broken and Gd can penetrate and highlight the inflamed areas. These areas of inflammation appear as active lesions, meaning that they are new or getting bigger or active. T1-weighted images without contrast agent depict dark areas (black-holes) that are suggestive areas of permanent nerve damage.

2. Fluid-attenuated inversion-recovery (FLAIR) images, in which the white matter lesions appear as bright spots, reflecting different levels of myelin loss and inflammatory activity. The sensitivity of the sequence is high but its specificity is low, as other lesions (e.g. vascular lesions) can mimic MS Lesions. Besides these, some part of healthy brain tissues like also exhibit the similar intesnity profile as MS lesions.

3. Dual-echo T2- and proton density-weighted images, in which the white matter lesions appear hyper-intense similarly to FLAIR images. They can be particularly useful in the posterior fossa, where FLAIR images have limited sensitivity.

4. Magnetization transfer (MT-MRI) is a scan which provides information about tissue damage associated to the disease. The sequences mentioned above are sensitive for detecting lesions and track their evolution over time. Lesions do not adhere to single-phenomenon of deviation from normal brain tissue. They are indicators of various areas of inflammation, demyelination, ischemia, edema, cell loss and gliosis. Sequences decscribed as above are unable to differentiate among these different pathologies. Conventional imaging also poorly characterizes the degree of injury in demyelinated lesions. In addition, conventional imaging does not identify all of the pathology in MS: there are widespread abnormalities in the white matter which appears normal on T2- and T1-weighted images. This problem is alleviated by MT because of its higher specificity than conventional T2-weighted scans [Filippi 2007, Fox 2011].

5. Diffusion-weighted imaging (DWI) MR scans provide a contrast that enables the diffusional motion of water molecules to be quantitatively measured. DWI provides information about of the brain micro structures. It is helpful to quantify apparent mean diffusion coefficient (MD) and Fractional Anisotropy (FA). The values for MD and FA are larger and smaller for lesion areas than for the normal appearing white matter [Filippi 2001].

3.2.2 Diagnosis

The diagnostic criteria for MS in conjunction with MRI observations with clinical and other para-clinical tools were introduced in 2001. The diagnosis of MS requires elimination of more likely diagnoses and demonstration of dissemination of lesions in space and time. The McDonald criteria shown in Figure 3.4 for MS were recommended in 2001 [McDonald 2001] by an international panel and revised in 2005 [Polman 2005] and 2010 [Polman 2011].

The McDonald criteria take into account the clinical presentation and MRI. When a patient experiences two or more episodes with clinical evidence of two or more neurological deficits, there is no need for additional requirements to make the diagnosis of MS, because there is dissemination in space and time. In all other cases, which are less than two episodes or less than two clinical significant lesions, there arises a need for MRI to fulfill the diagnostic criteria by demonstrating dissemination in space, in time or both. The McDonald criteria are quite specific pertaining to the fact that use of MRI to diagnose MS is only utilized when patient is screened for MS. Typical types of lesions which are suggestive of MS are shown in Figure 3.5. An involvement of the temporal lobe is shown by the red arrow, the green arrow indicates juxtacortical lesions

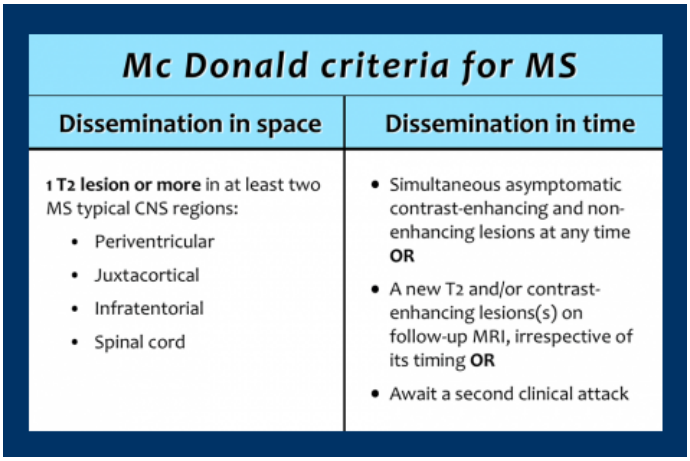


Figure 3.4: McDonald criteria for MS. Courtesy:http://www.radiologyassistant.nl/

touching the cortex, involvement of the corpus callosum is depicted as blue arrow and periventricular lesions touching the ventricles. The lesions in the deep white matter which are non specific to MS are shown by yellow arrow.

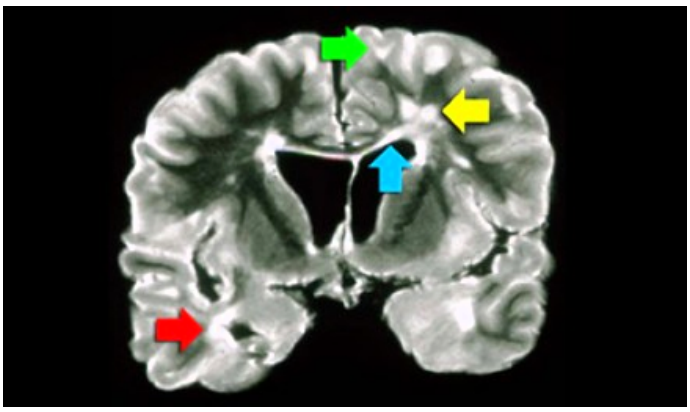


Figure 3.5: Coronal PD image of a brain specimen with MS involvement. Courtesy: http://www.radiologyassistant.nl/

3.2.3 MS Lesions (MSL)

Several MR sequences are necessary in order to detect the MS lesions. These lesions are classified into three subtypes of MS lesions depending upon the peculiar intensity characteristics they possess on respective sequences. They are Active/Gd-Enhancing lesions, black-holes and T2-w as shown in Figure 3.7.

- T2w lesions: These lesions exhibit a hyper intensity profile compared to normal-appearing white matter on T2w, PDw and FLAIR sequences. They may be iso- or hypo-intense in T1w images. T2w lesions are not clinically specific and can result from inflammation, edema, demyelination, or axonal loss. New contrast-enhanced lesions with hyper intensity also corroborate for the same location with a hyperintense lesion on T2-weighted images. These new T2 hyperintense lesions tend to reduce in size over time and their intensity decreases because of tissue repair [Meier 2007b, Meier 2007a].

- Gd Enhancing lesions: Longitudinal and cross-sectional MRI studies have demonstrated that formation of new MSL can strongly be linked with a focal area of contrast enhancement on T1-weighted images obtained after Gd injected intravenously. Typically, it can be observed in patients with RR or SP MS [Lassmann 2008]. This enhancement correlates with altered blood brain barrier permeability in the setting of acute perivascular inflammation and enables differentiation between acute, active lesions and chronic, inactive ones. In Figure 3.6, FLAIR image shows multiple focal demyelinating lesions that are hyperintense relative to the normal appearing brain tissue. After contrast administration, some of the lesions are hyperintense on T1-weighted images, indicating increased permeability of the bloodbrain barrier, a feature that distinguishes acute from chronic demyelinating lesions. The Gd enhancement varies in size and shape, and usually lasts from a few days to weeks, with an average duration of 3 weeks (97 % of lesions enhance during less than 2 months) [Cotton 2003]. Lesions which are new and increase in size are classified as active lesions.

- Black Holes: A T1-weighted MRI scan shows black holes which are suggestive areas of permanent axonal damage. These are hypo-intense lesions because of their dark intensity profile. To be a candidate for black hole, a T1-w lesion should not enhance with gadolinium and should generally be persistent for at least several months.

3.3 Conclusion

The origin and evolution of MS are still not well understood, and numerous studies have been conducted to evaluate its evolution and its influence on neighboring brain structures. Nowadays, a strong emphasis is put on early detection to slow down the disability and disease.
Quantitative analysis of MRI of different patients provides a spatial analysis

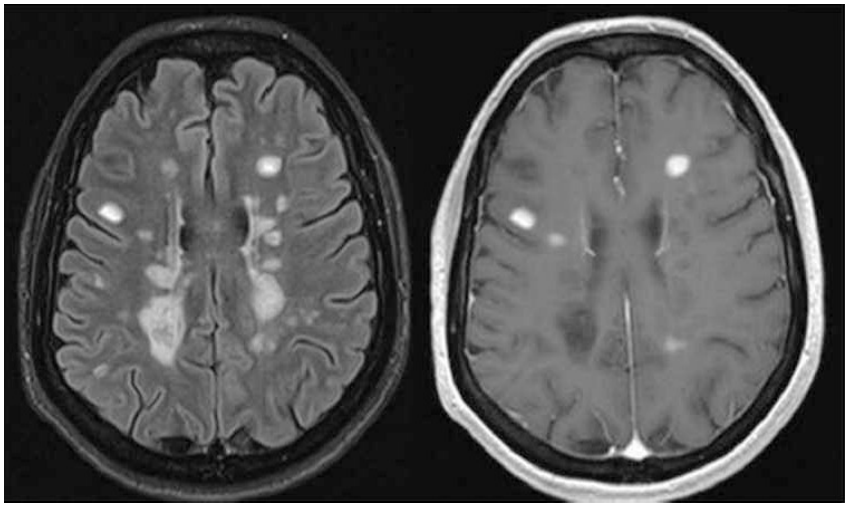


Figure 3.6: Conventional magnetic resonance imaging in multiple sclerosis. FLAIR (left) and gadolinium-enhanced T1-weighted (right) sequences. Courtesy: [Rovira 2013]

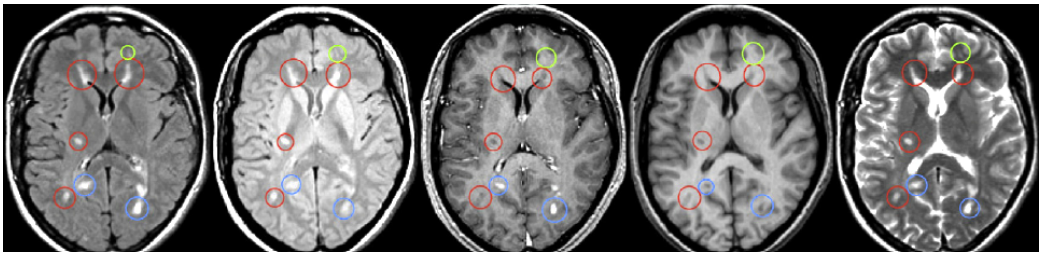


Figure 3.7: Example of MS lesions on MRI. From left to right: FLAIR, PDw, Gd-enhanced T1w, T1w, and T2w images. Several types of MS lesions can be observed: Enhancing lesions (blue), lesions visible only on T2w (green), black holes (red). In Gd-enhanced T1w and FLAIR images, multiple bright regions are observed that may be mislabeled as lesions. courtesy: [Garcia-Lorenzo 2013]

of the brain tissues, which may lead to the discovery of putative biomarkers of disease evolution. Nowadays, clinical trials use the total lesion load in conventional images. In the next chapter, we will focus on the state of the art methods of MS lesion segmentation and detection on conventional MRI.

Chapter 4

MS Lesions Segmentation

Contents

4.1 Introduction

MRI is playing an increasing role in the scientific investigation and clinical management of MS. Conventional MRI sequences are highly sensitive for detecting brain pathologies, like MS lesions and can provide quantitative assessment of inflammatory activity and lesion load. Quantitative MRI provides assistance for a large variety of applications, e.g. to predict brain lesion load and monitoring, longitudinal studies of cognitive aging [Bakshi 2008], or even the analysis of the fetal brain development [Rousseau 2013]. MRI achieves a great tissue contrast enabling the distinction between brain tissues; namely gray matter (GM), white matter (WM), and the cerebrospinal fluid (CSF).

Detecting and localizing MS lesions in MRI is a hard task and generally requires an expert neurologist or radiologist. The detection process is also time consuming and includes some subjectivity in interpreting the images. It

requires multi-sequence intensity fusion, deep anatomical knowledge and solid spatial awareness. MS lesions do not exhibit peculiar shapes and geometries. They possess nodular and oval like, ring to hole like shapes. Consequently, the MS Lesions detections performed by different experts can vary in the number and size of MS Lesions identified. Consequently, the MS Lesions detection performed by different experts can vary in the number and size of MS Lesions identified. As statistics reveal that eye fatigue is commonly encountered problem with radiologists as they visually need to inspect copious imaging data.[1] This becomes more prevalent when the imaging modalities grow dramatically. Due to the volume overload and constrained clinical information available as part of imaging studies, there may be room for diagnosis errors. Radiologists are a scarce resource in many countries. Therefore, it is of paramount importance to reduce the burden of data to be investigated by radiologists. To alleviate this problem, many computer assisted methods have been proposed for diagnostic interpretation of medical imaging datasets guided by clinical knowledge. These methods have the advantage to be consistent and repeatable, although they do not always achieve results as good as manual expert annotations.

In spite of these clear challenges, MS lesion segmentation research has made notable strides within the last decades. The techniques take different approaches to the problem of MS Lesions segmentation and consist of comprehensive frameworks made of several steps, including pre- and post-processing. Recent advances have shown the feasibility of learning accurate models for detecting MS lesions. Adapting various methodologies from different streams of science, researchers are making efforts for detecting MS lesions within realistic settings. As a part of the effort, various standardized benchmark databases and MS lesion detection challenges have been developed. While a numerous studies have been done, they all must make a few common choices: how will the MR images be represented? using that representation, how is a model learned? Given a new MR image, how is detection carried out? This chapter reviews state-of-the-art of strategies for MS lesions detection/segmentation methods with the aim of pointing out their strengths and weaknesses which in turn explains the rationale behind the proposed methods for this thesis. Further, it concludes with a discussion of the techniques as well as the perspectives on future improvements.

There are three main types of segmentation approaches depending on the user intervention: manual, semi-automatic, and automatic.

[1] http://researcher.watson.ibm.com/researcher/view_group.php?id=4384

4.2 Manual Segmentation

The first method to delineate MSL is manual segmentation. An expert rater examines different MR sequences to identify the lesion voxels. Unfortunately, the manual process is time consuming and somewhat subjective. Different experts (inter-rater variability) and even the same expert (intra-rater variability) may therefore provide different segmentations for the same data. Even so, manual annotations are considered the best results available and usually serve as the baseline for evaluating other methods. The expert segmentations can be considered as a silver standard since they are not perfect representations of the ground truth but provide the best estimates available. Automatic methods provide some assistance to MSL segmentation. Where experts can have difficulty in infusing multi modal MRI information, well designed frameworks can efficiently blend this data. As a result, it is interesting to pursue the development of semi-automatic and automatic lesion segmentation methods.

4.3 Semi-Automatic Segmentation

In order to reduce the inter- and intra- rater variability in segmentation of MSL, several semi-automatic methods have been developed. Semi automatic techniques need some human input as the prior knowledge for additional automatic processing steps. This knowledge could be an input in the form of focused region of interests (fROI) or a coarse-grained selection of object of interest. Though semi-automatic methods can relieve some of the work from radiologists, they do require some human interaction. A method based on prior knowledge with fuzzy logic is presented in [Horsfield 2007]. The prior knowledge resides in the form of probabilistic feature distribution and feature size maps, in a standard anatomical space. The fuzzy affinity between pixels is modified to capture this information. Here, the user is required to identify each lesion with a mouse click, to provide a set of seed pixels. The algorithm then grows the features from the seeds to define the lesions as a set of objects with fuzzy connectedness above a preset threshold. Ruben et al. [Cárdenes 2003] proposed the technique for interactive segmentation. It consists of three steps. First, a KNN classifier is applied to classify brain tissues in CSF, GM, WM and MS lesions on template based on user input. The second step concerns the detection of MS lesion which is done by computing a fast distance transformation in conjunction with intensity information on template. Last, a connected component technique is used to refine the voxels detected as MS lesions. Another approach proposed by [Derraz 2010] was segmentation based upon Active Contour Model and statistic prior knowledge of MS lesions

in fROI within MRI. In particular, the user selects coarse fROI that encloses potential MS lesions and a sufficient background of the healthy White Matter tissues (WM). Texture features corresponding to Normal Appearing Brain Tissues and MS lesions were incorporated to achieve final segmentation. Graph Cuts (GC) algorithm is a method for finding the maximum a posteriori (MAP) estimate of a binary image [Boykov 2006]. The method treats the image like a flow graph with two nodes, the source and the sink. The source represents the object class in the image, in this case the lesions. The sink represents the background: the NABT. The other nodes of the graph are the image voxels. A network of weighted and directed edges connects the nodes in the graph. The GC makes use of regional and voxel-neighborhood information to differentiate between the two classes. The MAP estimate corresponds to the maximum flow through the node network. The result is two sets of strongly connected nodes that correspond to the MSL and NABT [Biediger 2014]. Some authors [Lecoeur 2009] proposed to use GC with spectral gradient and multi-sequence MRI for lesions segmentation. All these methods need seed points defined by the user.

In any case, the automated step of the framework is highly dependent on the quality of the input. Since they require some level of user intervention, the semi-automatic methods may not cater the needs of large patient studies because they are still time consuming and tedious for the user.

4.4 Automatic Segmentation

Automatic methods require no user intervention. The comprehensive surveys of [Lladó 2012b, Garcia-Lorenzo 2013] provide the different types of segmentation frameworks. In general, there are three main types of fully automated segmentation schemes: data guided methods, learning based methods, and statistical methods. The data dependent methods use thresholding and region growing to segment the lesions in an image, like the watershed and graph cut methods. The learning based methods require a training set and some feature extraction. These methods learn the characteristics of lesions and then classify based on discriminative learning approaches. The statistical methods involve estimations of probability density functions of intensity of voxels. These methods are based on inference methods with some neighborhood or classification examples and include probabilistic graphical models and support vector machines. All have pros and cons in their use and the results they provide. Figure 4.1 shows the broad range of methods based on supervised and unsupervised techniques. It helps to take a glimpse on the huge literature of automatic MSL segmentation.

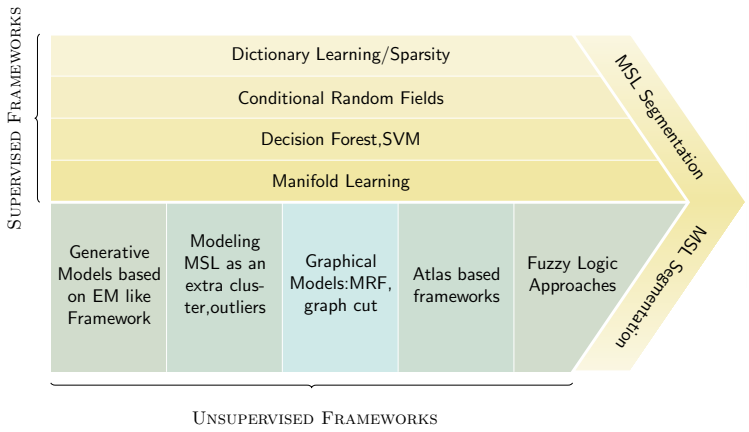


Figure 4.1: The various approaches for MSL detection based on their characteristics

4.4.1 Multi-Sequence Information

By and large, a typical MRI under consideration can be a representation in one or combination of the four possible sequences, namely, PD-w, T1-w, T2-w, and FLAIR. The objective is to find the various MS lesions through the use of these sequences. Prior works in the context of MS lesion detection deal with either single or multi-sequence approaches i.e. the use of a single MRI sequence or combination of several MRI sequences respectively. Single-sequence approaches are mainly used to segment the brain tissues.

For instance, T1-w sequences are widely used for this purpose, since they show the best contrast between the three main brain tissues: WM, GM and CSF. Another example of the single-sequence approach is the segmentation of MS lesions using just the FLAIR sequence [Khayati 2008]. The multi-sequence approaches, on the other hand, use at least two sequences. One of the benefits of using more than one of the different MRI sequences is that it increases the intensity feature space, producing a better discrimination between brain tissues. Garcia et al. [Garcia-Lorenzo 2011] propose a typical example of this approach which uses T1-w, T2-w, and FLAIR to detect MS lesions. There is also a method to make use of the initial single sequence approach to be used as a basis for further analysis in multi-sequence context to obtain the final lesion map. For example, the T1-w sequence could be used to influence a multi-sequence approach using T2-w and PD-w.

4.4.2 Unsupervised approaches for MS lesions segmentation

Generative methods remain the popular choice for MS lesions segmentation. It consists of tissue classification with an expectation maximization (EM) [McLachlan 2008] algorithm. The approaches based on EM typically modify the EM algorithm to factor it into their classification methodologies.

These EM derived algorithms are typically modified to be robust against lesion affected regions. The outcome is then parsed in order to detect outliers which, in this case, coincide with MS lesions. In their seminal work, Van Leemput et al. [Leemput 2001] developed a framework for segmentation of MS lesions based on multi-sequence information. It relies on classification based upon the intensity information of tissues using a stochastic model for normal brain sequences. Furthermore, it models MS lesions as outliers. It removes MR field inhomogeneities and incorporates contextual information in the classification of lesions using a Markov random field (MRF). Another similar work proposed by Garcia et al. [AïtAli 2005, Garcia-Lorenzo 2011] incorporates a tissue classification methodology based on a model of intensities of the normal appearing brain tissues. In order to estimate the model parameters, a trimmed likelihood estimator [Neykov 2007] is initialized with a hierarchical random approach in order to be robust to MS lesions and other outliers present in MR sequences. An iterative scheme of recursive EM was then used to compute this estimator. It is a robust algorithm using 3D+t MR data to segment MS lesions over time in a standardized clinical protocol. In the last step, refinement of the segmentation was done using both the Mahalanobis distance of intensity of WM voxels and prior information coming from clinical knowledge on lesion appearance across sequences. Another approach considers the EM with a partial volume model among tissue classes in conjunction with a Mahalanobis distance thresholding which detects MS lesions [Dugas-Phocion 2004]. A post-processing morphological operation was performed to refine the segmentation from regions of interest in order to improve the classification performance [Souplet 2008]. Freifeld et al. [Freifeld 2009] proposed a Constrained Gaussian Mixture Model (CGMM) technique based on a mixture of multiple spatially oriented Gaussians per tissue. The intensity of a tissue remains unchanged over the entire set of Gaussians for that tissue. MS lesions are modeled explicitly as an extra class by GMM in addition to healthy tissue classes. MS lesions were detected by estimation of parameters for outlier class followed by the refinement of lesion contours based upon the probability-based curve evolution technique. Rather than estimating the distribution of lesions, [Harmouche 2006] proposed to cast lesions as a separate class. It is based on unsupervised Bayesian framework. It models the different intensity distributions for different tissues of

the brain. MS lesion detection was performed using posterior probabilities along with entropy based features. Khayati et al. [Khayati 2008] classified MS lesions voxels based upon adaptive mixtures method (AMM) and a MRF model from a FLAIR sequence. The intensity of each lesion voxel is modeled as a linear combination of intensities related to the normal and pathological tissues. Applying an optimal threshold, the voxels with new intensities are primarily classified into two stages: chronic and acute MS lesions. Finally, the acute lesions are classified into two new stages based upon their activities, early- and recent acute. Schmidt et al. [Schmidt 2012] developed MS lesion detection using three-dimensional (3D) gradient echo (GRE) T1-w and FLAIR sequences. It initially classifies the three tissue classes of CSF, WM and GM from the T1-weighted image. In the subsequent stage, the FLAIR intensity distribution of each tissue class is taken into account to detect outliers, which are called lesion beliefs. The neighboring voxels in lesion belief maps are analyzed and assigned to lesions. This is done for all voxels that are associated with the MS lesions. A fuzzy C-Means algorithm was also investigated [Aymerich 2010a, Aymerich 2010b]. These approaches pursue the grouping of voxels into a number of clusters, which maximize inter-cluster variability while minimizing intra-cluster variability. Rather than a crisp or hard classification, the fuzzy approach establishes the degree to which a pixel belongs to a given cluster. In this way, a voxel can belong to more than one cluster to varying degrees.

4.4.3 Supervised Approaches

Machine learning plays an essential role in the medical imaging field, including computer-aided diagnosis, image segmentation, image registration, image fusion, image-guided therapy, image annotation, and image database retrieval. The objects such as lesions and organs may not be represented accurately by a simple models; thus, medical computer vision requires learning from training examples. One of the most popular uses of machine learning is classification of objects such as tissues into certain classes (e.g., pathological or non-pathological) based on input features obtained from segmented object candidates.

Support vector machines (SVM) [Vapnik 1995] is a popular and widely used supervised learning algorithm. It has also been used in the context of MS lesion detection [Fiot 2008, Abdullah 2011]. The method extracts intensity values as features from examples of lesion and non-lesion voxels. It then attempts to divide the two classes by dividing the hyper-plane of features in a discriminating fashion. While there are many possible dividing planes, the method seeks the plane with the widest margin between classes. SVM in conjunction with

various kernels facilitates the non-linear classification by projecting the features into Reproducible Kernel Hilbert Space (RKHS) [Shawe-Taylor 2004]. One problem with the SVM approach in MS lesion detection is the imbalance between class representations. In general, the number of voxels that represent normal brain tissue far exceeds the number of voxels that represent MS lesions. This can lead to the over fitting of the classifier.

Ensemble learning is the process by which multiple models, such as classifiers or experts, are strategically generated and combined to solve a classification problem. Two ensemble learning approaches were studied in the context of MS lesions detection. In a first study, Geremia et al. [Geremia 2010] proposed to build a discriminative random decision forest framework to provide a voxel-wise probabilistic classification of the image. The method uses multi-sequence data (T1, T2, FLAIR), a prior knowledge of tissue classes and long-range spatial context to distinguish lesions from healthy tissues. The authors also utilize a symmetry feature, which takes into account the fact that MS lesions tend to develop in an asymmetric way compared to healthy brain which remains approximately symmetric with respect to the mid-sagittal plane. In a second study, [Wels 2008] developed a framework based on the probabilistic boosting trees technique. It incorporates the context of a voxel under consideration and its transformation into feature space of an over-complete set of Haar-like features. This information establishes the class specific characteristics. A discriminative model for voxel classification was developed based upon boosting within a tree structure. It consists of selection and combination of most discriminative features which are established recursively in cascade. Consequently it yields posterior probabilities for voxels in learning phase. The final segmentation was obtained after refining the preliminary result by stochastic relaxation and a standard level set approach.

Further, a multi-scale segmentation can be combined with discriminative classification to take into account regional properties [Akselrod-Ballin 2006]. It relies on a combination of segmentation by Soft Weighted Aggregation (SWA), a rich feature vocabulary describing the segments, and a decision tree-based classification of the segments. Then, successively selecting and combining the most discriminative features during ensemble, the overall procedure was able to learn in terms of posterior probabilities. Beyond the information introduced via the spatial prior atlases, these methods are limited in their ability to take advantage of long-range spatial context in the classification task.

4.4.4 Gd-Enhancing Lesion Detection

All approaches mentioned above work on the conventional MRI sequences without contrast agent. However, it must be noted that in the context of active

MS lesions, the initial T1-w sequences can be enhanced using Gadolinium (Gd). Hence, the initial data to be analyzed would be different in a Gd-enhanced T1-w sequence. He et al. [He 2002] put forth the preliminary research incorporating the detection of MS lesions in a Gd-enhanced sequence. Using this as a base, the authors were able to overcome some of the deficiencies of the regular T1-w sequences. It must be noted that the authors use a standard methodology based on adaptive local segmentation that would work even with the regular MRI sequences. A Gd-enhanced sequence would consequently also affect the non-lesion parts of the brain. The authors identify the non-lesion voxels from the vasculature and extrameningeal tissues by exploiting their topological relationship to the brain mask. In these cases, there is hence a need to quantify the dynamics of lesion which is computed by using fuzzy connectivity. Karimaghaloo et al. [Karimaghaloo 2012] proposed a supervised-learning approach that uses Gd-enhanced sequences. This technique learns information pertaining to MS lesions by using Conditional Random Field (CRF). The novelty of the approach relies on applying the CRF two times in different contexts and has been termed as a Hierarchical CRF (HCRF) approach. The contexts that form the basis for training the CRF are based on voxel intensities and lesion-specific features. Karimaghaloo et al. [Karimaghaloo 2013] then extended this work on HCRF [Karimaghaloo 2012] using temporal information for longitudinal lesion segmentation. In this case the authors change the second stage of the HCRF mechanism to refine the candidate regions with rotation invariant texture features of the sequence. This methodology has been termed as Adaptive Multi-level CRF (AMCRF).

4.4.5 Miscellaneous Approaches

Recently, signal modeling using sparse representations (SR) has gained tremendous attention and is an area of active research. SR allows coding data as sparse linear combinations of the elements of an over-complete dictionary. It learn a class specific dictionaries for healthy brain tissues and lesions that promote their sparse representation. The lesion patches are well adapted to their own class dictionary, as opposed to the other. Weiss et al. [Weiss 2013] proposed an unsupervised approach for MS lesion segmentation, in which a dictionary learned using healthy brain tissue and lesion patches is used as basis for classification. On the other hand, [Deshpande 2014] achieved the detection using a supervised technique based on reconstruction error derived from sparse decomposition of test patches to learned dictionaries. Further [Deshpande 2015] introduced a framework for classification of MS lesions using adaptive dictionary learning. The MS lesions classification is based on multiple MR modalities including T1-w MPRAGE, T2-w, PD-w and FLAIR.

The method is based on sparse coding and dictionary learning. The authors propose that learning dictionaries with adaptive sizes for different classes can achieve better classification result. Their experimental results show that learning more dictionaries for each anatomical structure in the brain and adapting the dictionary sizes for different classes can improve the classification. Few authors evaluated manifold learning to distinguish MS lesions from NABT [Kadoury 2012, Kayhan 2008]. But all of these methods have huge computational complexity and do not scale to large data easily.

4.5 Performance Metrics

The evaluation comprises many evaluation measures, each of them highlighting different aspects of the segmentation quality. We describe four principal evaluation measures that are employed to evaluate the quality of a segmentation compared to a reference ground-truth: the dice coefficient (DC) denotes the volume overlap, the average symmetric surface distance (ASSD) the surface fit, the Hausdorff distance (HD) the maximum error and precision & recall the volume overlap.

- Dice Coefficient: DC measures the similarity between two datasets. Considering two sets of volume voxels A and B as shown in Figure 4.2, the DC value is given as:

$$DC = \frac{2|A \cap B|}{|A| + |B|} \tag{4.1}$$

  where $|\cdot|$ denotes the cardinality. A value of 0 indicates no overlap, a value of 1 perfect similarity. During interpretation, it should be kept in mind that the DC is known to yield higher values for larger volumes i.e. a DC of 0.9 for lung segmentation is considered average, as is a DC of 0.7 for MS lesions.

- Precision and Recall: The precision (also called positive predictive value) and recall (also known as sensitivity) of two sets are defined as:

$$precision = \frac{TP}{TP + FP}$$

  and

$$recall = \frac{TP}{TP + FN} \tag{4.2}$$

  where TP (true positive) denotes the overlapping points, FP (false positives) that are absent in the ground truth but present in algorithm/framework generated segmentation. FN (false negatives) are

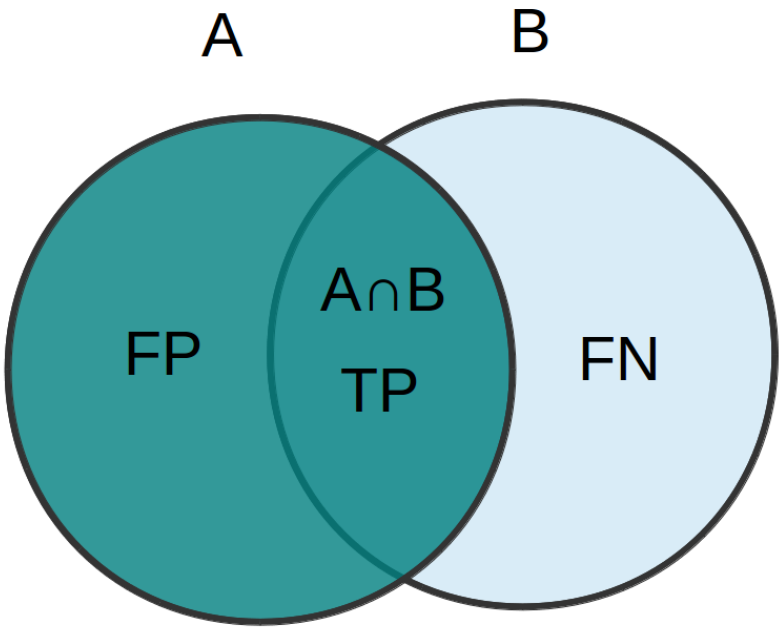


Figure 4.2: The two sets A, B and their intersection, where B is the reference segmentation.

points which are part of ground but absent in algorithm/framework generated segmentation. Both measures take values in the range of $[0, 1]$. A relatively high precision compared to the recall reveals under-segmentation and vice-versa, as depicted in the following Figure 4.3.

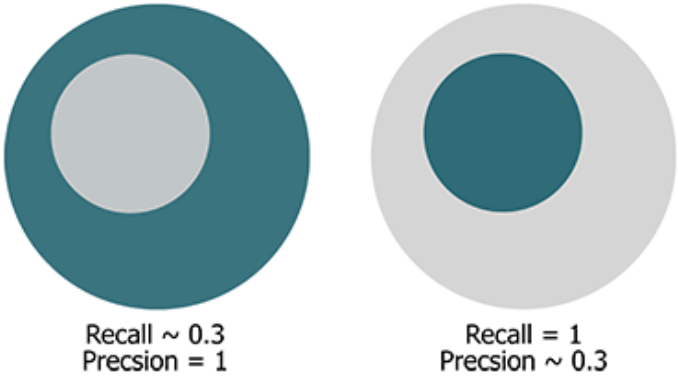


Figure 4.3: Gray circle: Segmented points. Green circle: Reference points.

- Average symmetric surface distance (ASSD): the ASSD denotes the

average distance between the volumes surface points averaged over both directions. Considering two sets of surface points A and B, the average surface distance (ASD) is given as:

$$ASD(A,B) = \frac{\sum_{a \in A} min_{b \in B} d(a,b)}{|A|} \tag{4.3}$$

with $d(a,b)$ being the Euclidean distance between the points a and b. Since $ASD(A,B) \neq ASD(B,A)$, the ASSD is given by

$$ASSD(A,B) = \frac{ASD(A,B) + ASD(B,A)}{2} \tag{4.4}$$

It is given in mm, the lower the better, and works equally well for large and small objects. However, it is sensitive to surface extraction from binary images.

- Hausdorff distance (HD): the HD denotes the maximum distance between two volumes of surface points and hence is sensitive to outliers, especially when multiple objects are considered. It is defined as:

$$HD(A,B) = \max\{\max_{a \in A} \min_{b \in B} d(a,b), \max_{b \in B} \min_{a \in A} d(b,a)\} \tag{4.5}$$

  Similar to the ASSD, the HD is given in mm and a lower value denotes a better segmentation.

- True Positive Rate : This is measured by dividing the number of lesions in the segmentation that overlap with a lesion in the reference segmentation with the number of overall lesions in the reference segmentation. This evaluates whether all lesions have been detected that are also in the reference segmentation. It is given in %, although it is possible to have a perfect score of 100% and have additional lesions as compared to the reference segmentation. A caveat for this measurement is further that if correctly detected lesions are fused as compared to the reference segmentation, then this is considered as a partial error.

- False Positive Rate: This is measured by dividing the number of lesions in the segmentation that do not overlap with any lesion in the reference segmentation with the number of overall lesions in the segmentation. This rate represents whether any lesions are detected that are not in the reference. A method that oversegments lesions would have a low value for this method, whereas a very conservative method would have high values. It is given in %, an empty image would always score a perfect score of 0%.

4.5.1 Publicly Available Resources

The inability to compare evaluation results due to the use of different data sets and different evaluation measures has been a major drawback to evaluate MS lesion segmentation methods. Very few methods are available as open source methods. Consequently, reproducing such methods is a very hard task. Majority of the segmentation algorithms are accompanied by heuristics rules. When heuristics fail, they fail catastrophically. Parameter tuning, change of heuristics are not easily generalized across the datasets. Some methods only work on specific datasets and their generalization on other data sets is a cumbersome task. Ideally, methods should be applied to a common database and compared to a ground truth. Moreover, for many researchers it is difficult to obtain a sizable amount of training and test scans with reference segmentation. Not only is the number of available methods increasing, more and more methods are, or claim to be, generic and applicable to multiple segmentation tasks, usually after applying some suitable modifications or tweaks. Even for experienced researchers in the field it is difficult to choose the appropriate technique for a particular problem. The competition/challenge serves purposes other than as a comparative study of a range of algorithms on a common database. It also provides a snapshot of currently popular methods for medical image analysis. This is however very difficult due to the lack of common public databases of real images along with their ground truth and the fact that only few methods are publicly available. Furthermore, the MS Lesion Segmentation Challenge [Styner 2008] provided a common framework for evaluating MS lesion segmentation algorithms, but data is old and image quality is poor to get adequate results. The MS lesions segmentation/detection field has continued to develop new approaches in recent years, and another challenge has been organized by the adding longitudinal dimension in the data which provides a new element for the researchers to consider in their algorithm development [2]. In that competition, participating teams were asked to evaluate their automatic lesion segmentation algorithms to MR neuroimaging data acquired at multiple time points from MS patients.

4.6 Challenges

Many radiologists advocated the objective evaluation of total lesion load which yet to be achieved with current segmentation techniques. Though segmentation techniques yield exact lesion masks, they fail to delineate all lesions successfully. In such situations, MS lesions detection and localization is setting trend with

[2] http://iacl.ece.jhu.edu/MSChallenge

its usefulness in identifying MS lesions instances considerably. Though many MS lesion segmentation frameworks are developed, they do not address some of the aspects of brain anatomy in presence of MS lesion. Following are the challenges which have been addressed from few to none:

- Segmentation in presence of partial volumes (PV): Due to the finite spatial resolution of imaging devices, a single voxel in a medical image may be composed of a mixture of tissue types, an effect known as partial volume effect (PVE). The border of lesions is fuzzy in part due to acquisition parameters of the MR scan. It was modeled explicitly in the context of MS [Dugas-Phocion 2004]. Lesion segmentation becomes quite difficult in presence of PVE; especially with methods which model the tissue distribution.

- Registration based methods: Anatomical atlases may be used for some methods. They are built from MR scans of healthy volunteers. These atlases are compared with patient images and thus it is possible to treat the lesions as outliers in the tissue. However, building atlases is a hard task. In addition, they also introduce the registration problem into the MS lesion segmentation. Note that this registration step is even more difficult when dealing with cases with severe atrophy, large numbers of lesions and inter-subject registration etc.

- Class Imbalance in supervised methods: Supervised learning methods require large training datasets. Compared to NABT data, MSL is always a sparse class leading to class imbalance problems. The majority of discriminative learning algorithms suffers from this problem. They produce suboptimal models which are biased towards the majority class (NABT) and have low performance on the minority class (MSL) [Chawla 2005]. The problem of imbalanced data is often associated with asymmetric costs of misclassifying elements of different classes. Additionally, the distribution of the test data may differ from that of the learning sample and the true misclassification costs may be unknown at learning time. Although much awareness of the issues related to data imbalance has been raised, many of the key problems still remain open and are in fact often encountered, especially when applied to massive datasets. Yet another school of thought is a recognition based approach in the form of a one-class (NABT) learner. The one-class learners provide an interesting alternative to the traditional discriminative approach, in which the classifier is learned on the target class alone [Tax 2004].

- Multi-center datasets: Images from different scanners have different contrasts or intensities, even when the same protocol is employed. Methods

should be designed specifically to deal with this variation without biasing the posterior clinical studies.

- Diffuse disease: The majority of the literature focuses on focal lesions but in some cases it is impossible to find the border between lesions and the neighboring diffuse white matter. No method has attempted to address these issues with the diffused white matter explicitly.

- Longitudinal MS Lesion segmentation: So far, very few methods are specifically tailored to detect newly appearing MS lesions for longitudinal MR scans. A longitudinal MS lesion segmentation challenge has been organized which took place at International Symposium on Biomedical Imaging (ISBI) 2015 conference this year.[3]

Techniques and algorithms that have stemmed from the field of machine learning have now become a powerful tool for the analysis of complex and large data, successfully assisting scientists in numerous breakthroughs of various fields of science and technology. In particular, machine learning provides algorithms that are able to solve classification or regression tasks, hence bringing automated procedures for the prediction of a phenomenon based on past observations. However, the goal of machine learning is not only to produce algorithms making accurate predictions, it is also to provide insights on the predictive structure of the data. One of the beneficiaries of machine learning field is computer vision. The current state-of-the-art methods from main stream computer vision in object detection and their derivatives [Dalal 2005, Felzenszwalb 2010, Uijlings 2013] are particularly well suited for handling large amounts of training data. They employ data mining to iteratively sift through millions of negatives and find the hard ones which are then used to train a discriminative classifier. The sliding window approach treats object detection as classification, and checks at every position and scale within the image whether the object is present. To run a multi-scale search, the input image is resampled into a pyramid. The window of interest is then slid through each level, and the classifier outputs are stored. A detector will usually have positive responses at multiple windows nearby the true detection.

Inspired from this, we propose an inclusive framework of the discriminative patch based learning methodology, which allows us to propagate rich annotations from exemplars onto detection windows, with discriminative training, which allows us to learn powerful exemplar-based classifiers from vast amounts of positive and negative data. Note that our framework is closely related to MS lesion segmentation with classification at voxel level. There exists subtle differences, though: segmentation is a symmetric setting for obtaining exact or

[3] http://iacl.ece.jhu.edu/MSChallenge

near exact, crisp contours of the lesion whereas we are interested in addressing the implicit problem of detecting and localizing MS lesions instances with discriminative learning based upon the patch information. Such detection may not yield exact lesion delineations, but a rough segmentation mask or bounding box around the lesion. The localization of MS Lesions is crucial, but for some applications a rough localization using a bounding box is enough. This is the task considered for MS Lesions detection. Most of the existing methods cast detection as a classification problem. Every possible window in the image is considered, and given to a classifier that decides if the window contains an MS Lesion or not. This type of approach is successful but extremely costly as such a sequential system has to classify thousands if not millions of windows for a single input image.

Aside from these issues relating to robustness, current state of the art segmentation algorithms also face notable challenges in computational complexity and scalability. Highly efficient algorithms are necessary to accommodate rich high-dimensional image representations, to search large image databases. In addition, scalability concerns also arise when designing a recognition system's training data: while expert annotations by radiologists tend to be most informative, they are also most expensive to obtain. Thus, methods today must consider the trade-off between the extent of expert annotations that an algorithm requires, and the advantages given to the learning process. Unfortunately, the current state-of-the-art in machine learning often makes it difficult for non-experts to understand and interpret the results of an algorithm. While considerable efforts have been put to improve their prediction accuracy, it is still not clearly understood what makes machine learning algorithms truly work, and under what assumptions. Likewise, few of them actually provide clear and insightful explanations about the results they generate. In this context, the goal of this thesis is to provide frameworks based on comprehensive and self contained analysis of a set of algorithms.

Chapter 5

Longitudinal Intensity Normalization in Multiple Sclerosis Patients

Contents

In recent years, there have been many Multiple Sclerosis (MS) studies using longitudinal MR images to study and characterize the MS lesion patterns. The intensity of similar anatomical tissues in MR images is often different

because of the variability of the acquisition process and different scanners. This chapter proposes a novel methodology for a longitudinal lesion analysis based on intensity standardization to minimize the inter-scan intensity difference. The intensity normalization maps parameters obtained using a robust Gaussian Mixture Model (GMM) estimation not affected by the presence of MS lesions. Experimental results demonstrate that our technique accurately performs the task of intensity standardization. In this chapter, we show consequently how the same technique can improve the results of longitudinal MS lesion detection.

5.1 Introduction

Quantitative analysis of longitudinal Magnetic Resonance Images (MRI) of subjects taken at different time points provides a time varying analysis of the brain tissues which may lead to the discovery of new biomarkers of disease evolution. In MS, White Matter (WM) lesions are also present in addition to healthy brain tissues. Lesions can remain stationary, change of volume, or disappear in later time points depending upon the state of MS. MRI suffers from two drawbacks: (1) intensity variations due to magnetic field in-homogeneities; (2) scanner-related intensity artifacts. Therefore, it is very difficult to create standardized intensity range for particular sequence. Algorithms dealing with the problem of the correction of signal intensity inhomogeneities (bias field correction) usually deal with intra-volume signal intensity distortions and do not account for native cause of lack of intensity standardization. Hence, MRI acquired from different patients may have different intensity ranges related to a specific anatomical tissue even if they are acquired on the same MRI scanner. Consequently, this has a direct impact on the accuracy and precision of following image processing, analysis, segmentation and registration methods relying on intensity similarity. Furthermore, standard presets cannot be used to display MR images or to visualize certain tissue classes and/or pathologies. These settings need to be tuned for each patient. Due to such protocol variations in the scanners, following the evolution of tissue intensities in a patient (e.g. changing appearance of lesions) makes quantitative evaluation of lesions difficult. In order to alleviate this problem, intensity normalization is necessary.

Histogram matching is a widely used technique in intensity standardization. In their seminal work, Nyul et al. [Nyul 2000] proposed landmark based methods. Essentially the method is based on a multi-segment linear transformation model. A standard intensity space (or common intensity space) is defined by an intensity value range. During the training phase, the intensity values at certain cut-off percentiles of each image are computed and a single-segment

linear mapping from them to the standard intensity space range limits is created. Then the image intensity values at a number of landmark percentiles are extracted and passed to the linear mapping to be transfered roughly to the standard intensity space. The mean of all these mapped landmark intensities form the learned model. When presented with an image to transform, these images intensity values are extracted at the cut-off percentile as well as at the landmark percentile positions. This results in a number of segments. Using these, corresponding standard intensity space range values and learned mean landmark values, a multi-segment linear transformation model is created for the image. It is then applied to the target images intensity values to map them to the standard intensity space. The images intensity values that lie outside of the cut-off percentiles i.e, outliers, are treated separately. Note that this does not mean the transformed images intensity values always reside inside the standard intensity space range, but are fitted as close as possible.

Jäger et al. [Jäger 2006] showed that a normalization can be achieved by finding a deformation of the joint histograms of two sets of images pertaining to certain divergence metrics. Each of these histograms should be at least two dimensional and contain the intensity information of multi-sequence MRI. If the probability density functions are considered as images, the intensity normalization boils down to a registration problem. The resulting non-linear correction function is used to adjust the image intensities of the longitudinal MRI. Furthermore, Jäger et al. [Jäger 2007] extended this approach to a whole body MRI scan. In this approach, the image volume is split up into K sub volumes for which intensity standardization is performed independently. In order to include the influence of small local structures the other $K-1$ sub volumes are used as regularizer.

Cox et al. [Cox 1995] studied the intensity standardization problem as dynamic histogram warping closely related to histogram specification task often studied in early computer vision. Dynamic programming can be used to find an optimal alignment between two images constrained by a monotonic and separable cost function. Let h_m^A and h_n^B represent the histogram of image A and B respectively. Let H_m^A and H_n^B represents the cumulative distribution function such that $H_m^A = \sum_{i=1}^{m} h_i^A$ and $H_n^B = \sum_{i=1}^{n} h_i^B$. The cost function of matching intensity I_m^A of image A with respect to intensity I_n^B of image B is simply $|h_m^A - h_n^B|$. Since, the histogram is nothing but the frequency of intensity values, for one-to-one mapping the cost function should be $|h_m^A - (h_n^B + h_{n-1}^B)|$ and for one-to-k mapping $|h_m^A - \sum_{i=1}^{k-1} h_{n-1}^B|$. The fact that the cost function of matching h_{m+1}^A to h_n^B depends on whether or not h_m^A was matched to h_n^B, may cause problem to dynamic programming. However, the upper bound on size of increase or decrease of histogram is always finite. In general the cost of k-to-l mapping

is $d_{k,l} = \left|\sum_{i=1}^{k-1} h^A_{m-i} - \sum_{j=1}^{j-1} h^B_{n-j}\right| = \left|(H^A_m - H^A_{m-k}) - (H^B_n - H^B_{n-l})\right|$, where m and n represent the maximum allowable shrinking of respective histogram. This can be computed by dynamic programming.

An algorithm proposed by Wang et al. [Wang 1998] expands or shrinks a windowed part of the input image histogram with a multiplicative factor, found by minimizing the bin-count difference between the source and moving images histograms. The window is used to include only voxels of interest and exclude the background. This makes the technique linear in the intensity range of interest. Other techniques use parametric models, such as the technique proposed by Hellier [Hellier 2003]. This intensity normalization method is applicable only to brain images. This is done by matching intensities of head specific anatomical tissue classes. It consists of two stages. In the first step the histograms from a source and a target data set, i.e. two 3D images, are approximated by a Gaussian Mixture Model with five classes i.e. background, WM, GM, CSF and a mixture of fat and muscle. Each tissue class k is modeled by a Gaussian probability density function which has a mean μ_k and variance σ_k that is approximated using the Expectation-Maximization (EM) algorithm. In a second step, a polynomial correction function f^p of order p is used to interpolate the correction of the intensities smoothly:$f^p(x) = \sum_{i=0}^{p} \Theta_i x^i$. By minimizing the the cost function $\sum_{k=1}^{n}(f^p(\mu_k) - v_k)^2$ the coefficients Θ_i are obtained, where μ_k and v_k are the means of the source and target image respectively. Weisenfeld et.al. [Weisenfeld 2004] have proposed to estimate a multiplicative correction field that alters the intensity statistics of an image or set of images to best match those of a model. In that paper, the Kullback-Leibler divergence between the source and moving images is minimized iteratively to estimate the parameters of a model, thus histograms are equalized. All these methods may be affected by the presence of white matter lesions.

Recently, Robitaille et al. [Robitaille 2012] developed an automatic intensity standrdization technique STandardization of Intensities (STI) based on Nyul's method. It consists of (1) using both histogram and tissue-specific intensity information; (2) providing a nonlinear intensity transformation between images. In that technique, a target image is aligned with respect to the source image using global nonlinear registration which establishes spatial alignment between tissues in the source and moving images. Due to the spatial spatial correspondence, a joint intensity histogram of the frequency distribution of intensity correspondences can be computed. From the most frequent tissue-specific correspondences, an intensity transfer function can be computed to transform the nonlinearly registered input image onto the source, which is then applied to the linearly registered target image to compute the desired standardized image, in the standard intensity space. Roy et

al.[Roy 2013a, Roy 2013c] proposed an intensity standardization framework based patch matching technique that takes patches from the atlas (source) image and finds its best matching patches in target image. Let $X = x_i$ and $Y = y_j$ be the collection of target and source patches. To match the target intensities to the source, $X \rightarrow Y$ matching is done. It makes two assumptions: (1) WM peaks of both the source and the target image are nearly same; (2) each target patch x_i is a realization of a Gaussian random vector whose mean is one of the atlas patches i.e. $x_i \sim \mathcal{N}(y_j, \Sigma_j)$ for some j. Then both the best matching source patch for each target patch and the covariances of all source patches are defined by maximum likelihood and found using an expectation maximization algorithm. The normalized image is produced by replacing the center pixel of each observed patch by the corresponding value of the matching source patch. Both of these methods assume that MR image does not contain any pathology. In the same year, Roy et al. [Roy 2013b] introduced longitudinal intensity normalization in the presence of multiple sclerosis lesions. The framework consists of two steps: (1) the longitudinal transformation of intensities of the normal tissues of different time-points for T1-w MPRAGE sequence using a first order Auto Regressive model; (2) a prior knowledge for the lesions at each time-point is incorporated from an atlas based topology preserving lesion segmentation method, called Lesion-TOADS [Shiee 2010], which takes both the T1w-MPRAGE and the FLAIR scans of a subject and provides fuzzy lesion memberships at every voxel.

In this chapter, propose a longitudinal intensity normalization algorithm for multichannel MRI in the presence of MS lesions, which provides consistent and reliable longitudinal detections. The tissue intensities from multichannel MRI are modeled with parametric transform using a robust GMM estimation based on γ-divergence, thereby keeping the lesions unaffected. The proposed technique is built on ideas similar to Hellier [Hellier 2003] but taking into account the presence of pathological tissues in the intensity transformation function. It provides a technique that (1) uses tissue-specific intensity information by modeling them using a robust GMM; (2) provides a consistent intensity normalization between longitudinal images. Subsequently, we demonstrate its crucial role for further lesion analysis.

This chapter is organized as follows. The modeling and parameter estimation of multi-sequence MRI with γ divergence followed by intensity normalization are reviewed in Section 5.2. The details of experiments and their results on longitudinal MS patients are discussed in Section 5.3.

5.2 Methodology

5.2.1 γ-Divergence

The aim of this section is to give a general introduction to the γ-divergence and the γ-loss function [Eguchi 2010]. Suppose a random sample is generated from a population distribution with density function z. Let $\{f(\cdot,\theta)\}$ be a family of density functions indexed by parameter θ. The γ cross entropy between z and $f(\cdot,\theta)$ is defined as

$$C_\gamma\left(z, f(\cdot,\theta)\right) = -\kappa_\gamma(\theta)\int z(x) f(x,\theta)^\gamma dx \tag{5.1}$$

with power index $\gamma > 0$, where $\kappa_\gamma(\theta)$ is the normalizing constant defined as

$$\kappa_\gamma(\theta) = \left(\int f(x,\theta)^{1+\gamma} dx\right)^{\frac{-\gamma}{1+\gamma}} \tag{5.2}$$

The Boltzmann-Shannon cross entropy between z and $f(\cdot,\theta)$ is defined as

$$-\int z(x) log f(x,\theta) d\theta \tag{5.3}$$

The γ-cross entropy and the Boltzmann-Shannon entropy have the following relationship since $\kappa_\gamma(\theta)$ converges to 1 if γ tends to 0.

$$\lim_{\gamma\to 0}\frac{C_\gamma\left(z, f(\cdot,\theta)\right)+1}{\gamma} = \int z(x)\lim_{\gamma\to 0}\frac{f(x,\theta)^\gamma - 1}{\gamma} dx = -\int z(x) log f(x,\theta) dx \tag{5.4}$$

Hence the Boltzmann-Shannon cross entropy can be seen as the 0-cross entropy, and the γ-cross entropy can be regarded as an extension of the Boltzmann-Shannon cross entropy. γ entropy of z is defined as

$$H_\gamma(z) = C_\gamma(z, z) \tag{5.5}$$

And the γ-divergence between z and $f(\cdot,\theta)$ is defined as

$$D_\gamma(z, f(\cdot,\theta)) = C_\gamma(z, f(\cdot,\theta)) - H_\gamma(z) \tag{5.6}$$

Note that the γ-divergence $D_\gamma(z, f(\cdot,\theta))$ is nonnegative and $D_\gamma(z, f(\cdot,\theta))$ is equal to 0 if and only if θ satisfies that $z(x) = f(x,\theta)$ for almost every x. From these properties $D_\gamma(z, f(\cdot,\theta))$ can be seen as a distance between z and $f(\cdot,\theta)$ although it does not satisfy the symmetry condition. The objective is to find the closest distribution to z in model $f(\cdot,\theta)$ with respect to γ-divergence. In order to accomplish this task, we only need to find the global minimum point of $D_\gamma(z, f(\cdot,\theta))$ with respect to θ which is equal to one of $C_\gamma(z, f(\cdot,\theta))$.

5.2.2 γ-Loss Function

The γ-loss function is defined by an estimator of γ-cross entropy. Let $u_1, u_2, ..., u_n$ be a random sample generated from population distribution with density function z and $f(\cdot, \theta)$ be the model under consideration. The γ-loss function for $f(\cdot, \theta)$ associated to the γ-divergence is given by

$$L_\gamma(\theta) = -\kappa_\gamma(\theta)\frac{1}{n}\sum_{i=1}^{n} f(u_i, \theta)^\gamma \tag{5.7}$$

For any distribution function G, the γ-cross entropy between G and $f(\cdot, \theta)$ is defined as

$$C_\gamma(G, f(\cdot, \theta)) = -\kappa_\gamma(\theta)\int f(u, \theta)^\gamma dG(u) \tag{5.8}$$

Note that $L_\gamma(\theta)$ equals to $C_\gamma(\widehat{G}, f(\cdot, \theta))$ with empirical distribution $\widehat{G}$ and $L_\gamma(\theta)$ almost converges to $C_\gamma(z, f(\cdot, \theta))$. The γ-loss function and the log likelihood function satisfy the following relation

$$\lim_{\gamma \to 0} \frac{L_\gamma(\theta) + 1}{\gamma} = -\frac{1}{n}\sum_{i=1}^{n} log f(u_i, \theta) \tag{5.9}$$

Hence maximum likelihood estimation (MLE) can be regarded as 0-estimator and γ-estimator can be seen as extension of MLE.

5.2.3 Brain Tissue Intensity Modeling

Given two MR images of a single MS patient at time instant t_1 and t_2, we seek to estimate a correction factor such that corresponding anatomical tissues adopt the same intensity profile. We model the image intensities of a healthy brain with a 3-class GMM, where each Gaussian represents one of the brain tissues White Matter (WM), Gray Matter (GM) and Cerebrospinal fluid (CSF). We consider the m MR sequences as a multidimensional image with n voxels. Each voxel i is represented as $\mathbf{x}_i = [x_{i1}...x_{im}]$. The probability of intensity $\mathbf{x}_i$ is calculated as follows:

$$f(\mathbf{x}_i|\theta) = \sum_{k=1}^{3} \pi_k \mathcal{N}(\mu_k, \Sigma_k) \tag{5.10}$$

where the mean μ_k and covariance Σ_k define the parameters $\mathcal{N}(\mu_k, \Sigma_k)$ of each Gaussian of the model along with their mixing proportions π_k merged into parameter θ. If the proportions were known, θ could be estimated through the Maximum Likelihood Estimator (MLE):

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmax}}\ L(\theta) = \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^{n} f(\mathbf{x}_i|\theta) \tag{5.11}$$

Where $\mathbf{x}_i$ are considered as i.i.d. samples. However, as π_k are unknown, an Expectation Maximization (EM) algorithm [Dempster 1977] is used to estimate the parameters.

5.2.4 γ-loss Function for the Normal Distribution

The parameter estimation with classic MLE for GMM can deviate from its true estimation in presence of outliers. In MS patients, such outliers may be of crucial importance as they may denote appearing or disappearing lesions. Notsu et al. [Notsu 2014] proposed a modification of the MLE in order to make it more robust to outliers. The basic idea is to maximize equation (5.11) in the form of γ-divergence. We consider the γ-loss function for the Normal distribution with mean vector μ and covariance matrix Σ.

$$L_\gamma(\mu, \Sigma) = \left|\Sigma^{-\frac{\gamma}{2(1+\gamma)}}\right| \sum_{i=1}^{n} \exp\Big(-\frac{\gamma}{2}(\mathbf{x}_i - \mu)^T \Sigma^{-1}(\mathbf{x}_i - \mu)\Big) \tag{5.12}$$

Where $|.|$ indicates the determinant. The bounded influence function of an estimator is an indicator of its robustness to outliers. The influence function for GMM with γ-loss function is bounded whereas the one for regular GMM is unbounded. As γ grows larger, bounds become tighter. For a sufficiently large γ, ($\gamma \geq 0.1$), the estimating equation has little impact from outliers in the data set. Equation (5.12) can be casted to yield an EM style algorithm as follows:

Expectation Step. In the case of a GMM, the latent variables are the point-to-cluster assignments $k_i, i = 1, ..., n$, one for each of n data points. The auxiliary distribution $q(k_i|\mathbf{x}_i) = q_{ik}$ is a matrix with $n \times K$ entries. Each row of q_i can be thought of as a vector of soft assignments of the data points $\mathbf{x}_i$ to each of the Gaussian modes.

$$q_{ik} = \frac{\pi_k \exp\Big(-\frac{\gamma}{2}(\mathbf{x}_i - \mu_k)^T \Sigma_k^{-1}(\mathbf{x}_i - \mu_k)\Big)}{\sum_{l=1}^{K} \pi_l \exp\Big(-\frac{\gamma}{2}(\mathbf{x}_i - \mu_l)^T \Sigma_l^{-1}(\mathbf{x}_i - \mu_l)\Big)} \tag{5.13}$$

Maximization Step. The maximization step estimates the parameters of the Gaussian mixture components and the mixing proportions π_k, given the auxiliary distribution on the point-to-cluster assignments computed in the expectation step. The mean μ_k of a Gaussian mode is obtained as the mean of the data points assigned to it (accounting for the strength of the soft assignments). The other quantities are obtained in a similar manner, yielding

$$\mu_k = \frac{\sum_{i=1}^{n} q_{ik}\mathbf{x}_i}{\sum_{i=1}^{n} q_{ik}} \tag{5.14}$$

$$\Sigma_k = (1+\gamma)\frac{\sum_{i=1}^{n} q_{ik}(\mathbf{x}_i - \mu_k)(\mathbf{x}_i - \mu_k)^T}{\sum_{i=1}^{n} q_{ik}} \tag{5.15}$$

$$\pi_k = \frac{\sum_{i=1}^{n} q_{ik}}{\sum_{i=1}^{n}\sum_{l=1}^{K} q_{il}} \tag{5.16}$$

5.2.5 Selection of Parameter γ

The estimation of power index γ plays a critical role in our approach, since γ affects the estimated parameters in the presence of outliers. Notsu et al. [Notsu 2014] suggested the selection of γ as a model selection problem based on Akaike information criterion (AIC) [Akaike 1974]. It is simply the sum of the negative log likelihood of the observed data under the current model plus the number of its fitted parameters. The first term of this function reflects the quality of fitting, whilst the second penalizes overly complex models. AIC is based on the Kullback-Leibler divergence - a (non-symmetric) measure of difference between two probability distributions. AIC is a measure of the relative quality of a statistical model for a given set of data. That is, given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models. Hence, AIC provides a means for model selection.

Let K be the number of clusters, p_k be the total numbers of parameters of a model and (μ_k, Σ_k), $k = 1, .., K$ be the means and the covariance matrices of the clusters respectively. From equation(5.10), the AIC is defined as follows:

$$\mathbf{AIC}_\gamma = -2\sum_{i=1}^{n} \log f_\gamma(\mathbf{x}_i|\theta) + 2\left\{K\frac{p(p+3)}{2} + K - 1\right\} \tag{5.17}$$

The value of γ which minimizes AIC is used as the optimal γ. For various values of γ, equation (5.17) is evaluated in a cross validation manner and the γ which results in minimum value is chosen for the experiment. Mathematically, it can be put as

$$\widehat{\gamma} = \arg\min_{\gamma} \mathbf{AIC}_\gamma \tag{5.18}$$

5.2.6 Intensity Correction

We obtain the means and covariances of tissues for the source and target images using the procedure mentioned above. We chose a linear correction function such that $g(\mathbf{x}) = \Sigma_i \beta_{i=0}^{i=1} \mathbf{x}^i$. The coefficients β_i are estimated to minimize the following cost function: $\Sigma_{k=1}^{K}(g(\mu_{source,k}) - \mu_{target,k})^2$. This function can be solved by linear regression. Using the results of the linear regression, the intensity profiles of the two images are normalized by mapping the intensity of the source image to the target image. The resulting correction function is smooth and interpolates the intensity correction. Figure 5.1 depicts the intensity correction method.

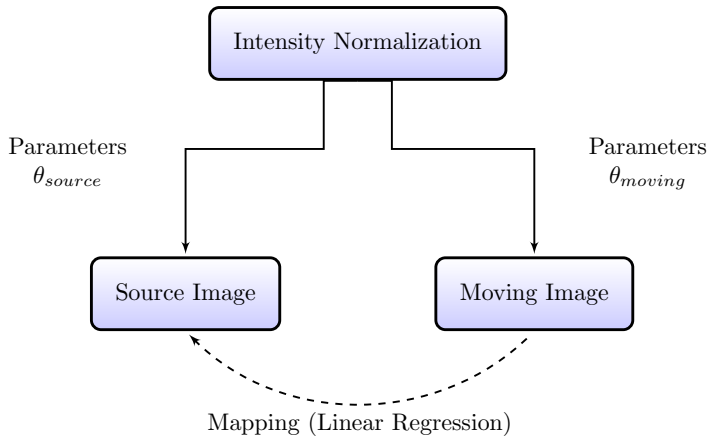


Figure 5.1: Schematic description of intensity normalization.

5.2.7 MR Serial Change Detection

Longitudinal studies for MS patients are frequently performed on patients with diseases of the brain, to track and observe changes. MRI provides comprehensive and valuable information to study brain evolution over a period of time. There exist many studies which discuss the MS lesion detection problem, very few of them target the lesion evolution studies. The comprehensive review of the state-of-the-art methods for automated detection of multiple sclerosis lesions in serial brain MRI is done in [Lladó 2012a]. Analyzing literature, broadly, three categorizes of (semi)automatic methods exist: intensity based point-to-point comparison, deformation field based and temporal analysis based methods. Their brief description is shown in Figure 5.2. The most routine approach for the detection of change on imaging studies is manual inspection. However, this approach suffers from a number of problems. One of the most important of these is the number of images that the radiologist has to

investigate. With growing number of MR sequences at each acquisition time, data grows dramatically. Curati et al. [Curati 1996] studied the serial MRI

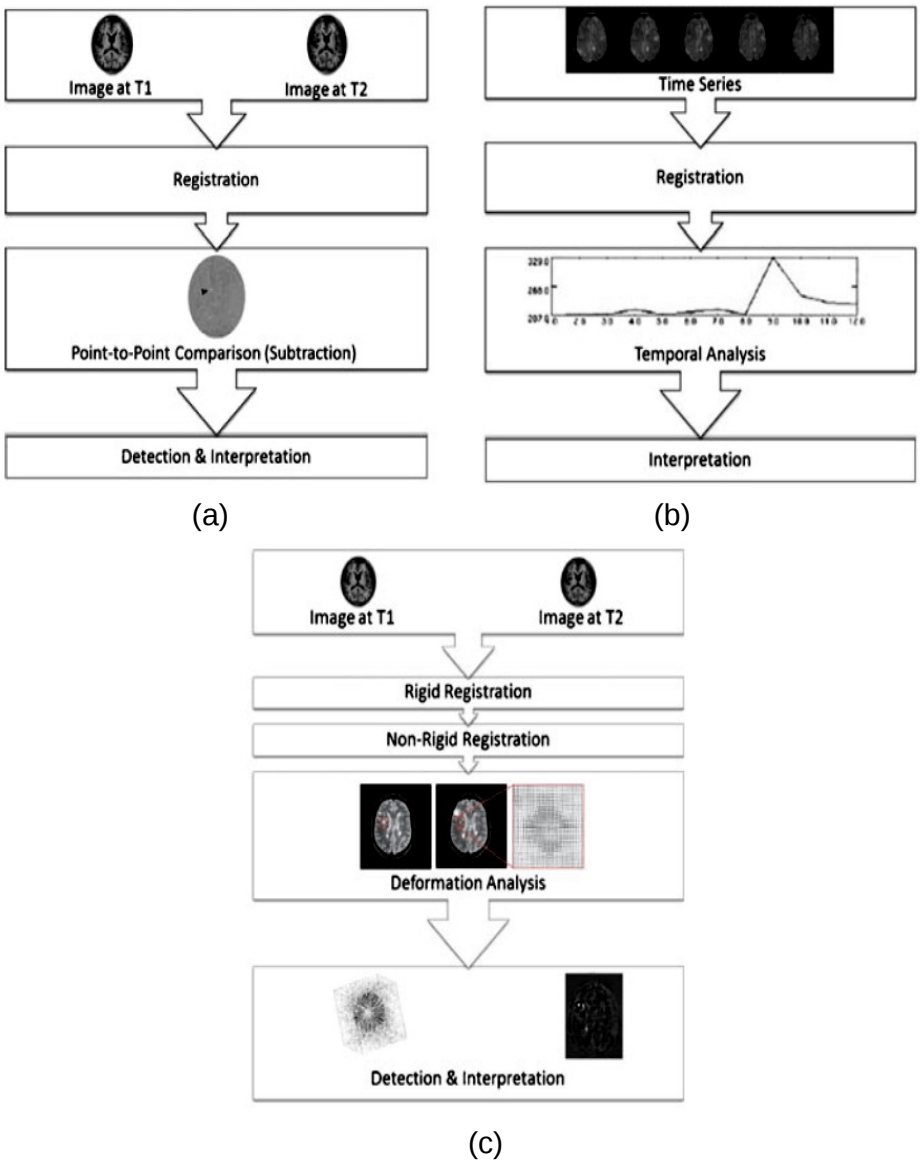


Figure 5.2: Schematic description of change detection approaches: (a) Intensity-based techniques; (b) temporal analysis; (c) deformable approaches. Courtesy: [Lladó 2012a].

study. In that method, the baseline and repeat scan images were registered

to sub-voxel accuracy using rigid registration and subtraction images were obtained from these. Normal contrast enhancement was demonstrated better with registered volume and subtraction images than with conventional images. Abnormal enhancement was seen better in meningeal disease, multiple sclerosis and tumors.

Evolution can be expressed in many ways, to different degrees spread across various MR sequences. The radiologist is required to look at all of this data in order to perform diagnosis, which often is quite difficult [Moraal 2009]. Moraal et al. [Moraal 2010] extended this approach on 2D and 3D subtraction methods and conclude that 3D subtraction techniques, after image registration, provided greater inter-observer agreement. Furthermore, they evaluated several image sequences (3D DIR, 3D FLAIR, 3D T2-w, 3D T1w-MPRAGE) and concluded that even small active lesions could be detected using the 3D T1w-MPRAGE sequence because of its good contrast among tissues. However, the manual detection of change is not only time-consuming; it is also constrained by intra-observer and inter-observer variability. On the other hand, it is a well proven fact that automatic systems may outperform the human expert. Bosc et al. [Bosc 2003] showed that those small and subtle changes in lesion evolution missed by the expert, are correctly detected by the automated change detection algorithm. Change detection based methods are intensity-based approaches based on subtracting two successive images in order to find intensity differences due to evolving lesions. Tan et al. [Tan 2002] observed that after the subtraction of two consecutive temporal images, unchanged areas (normal tissue) appear as gray areas, while changed areas are due to the appearance or disappearance of lesions. Hence, the positive activity (new or enlarging lesions) appears as a bright area while the negative activity (resolving or shrinking lesions) appears as a dark area against the gray background.

Thirion et al. [Thirion 1999] advocated MS lesions is observed as phenomenon of the combination of two different factors. First is the tissue transformation (image intensity changes without deformation) and second is the expansion or contraction effects reflecting a change of mass within the tissue. Therefore, using only approaches based on intensity changes between serial scans to evaluate the evolution of lesions may not give satisfactory results, since the surrounding tissue deformation due to the presence of the lesion is not taken into account. In order to account for the mass effect of the lesions, deformation-based approaches should be employed. In deformation field-based approaches, a nonlinear registration is performed between successive scans, and the structural changes are determined based on the local deformation of voxels. A semiautomatic approach using vector displacement fields obtained by a nonrigid registration of two successive scans to track MS lesions [Thirion 1999]. The method makes use of both the divergence and the norm of the displacement

vector fields to capture deformation and intensity transformation. The high values of the norm are an indicator of large areas undergoing large deformation, while high divergence indicate evolving lesions.The sign of the divergence operator suggests whether the lesion is shrinking or expanding.

In their seminal work, Meier at al. [Meier 2003] assessed the longitudinal MRI intensity profiles in MS patients. Multiple serial MRI scans were concatenated into a spatio-temporal volume for direct quantitative evaluation of the temporal intensity profiles. The dynamics of temporal intensity signal threw the light on MS pathogenesis and exploits the possibility for MRI biomarkers of disease activity and progression. The framework consists of five steps: (1) spatial alignment: rigid and non-rigid registration of follow-up scans to baseline scan; (2) intensity normalization: to account for similarity of intensity distributions among corresponding tissues; (3) artifact removal: Partial Volume Effect removal; (4) information fusion: point-to-point comparison using subtraction; (5) parametric analysis: analysis of time series of intensities and formation of feature maps.

Note, however, the all these approaches are based on the differences between the successive scans, lesions which remain stationary cannot be detected. We propose a novel MR serial change detection approach based on intensity approach which provides consistent and reliable longitudinal detections. The tissue intensities from multichannel MRI are modeled with a parametric transform using a robust GMM estimation based on γ divergence, thereby keeping the lesions unaffected. The framework consists of (1) tissue-specific intensity normalization of target image with respect to source image; (2) point-to-point comparison subtraction; (3) Otsu thresholding to get final lesion change detection. The block diagram of the approach is shown in Figure 5.3.

5.2.8 Detection of active lesions in T1-w Gadolinium (Gd) Images

To keep track of the active lesions, contrast agent (Gd) is administered and subsequently T1-w images are acquired. Detection of Gd-enhancing lesion is challenging task because lesions and other brain parts like vessels and fat tissues exhibit the very similar intensity profiles. To alleviate this problem, for detecting lesions in T1w-Gd images, T2-w lesions masks are used to cancel out the effect of Gd-absorbing tissues in lesion detection. The rest of the process is the same as in Section 5.2.7 except that the two time points are replaced by pre-abd post-contrast images.

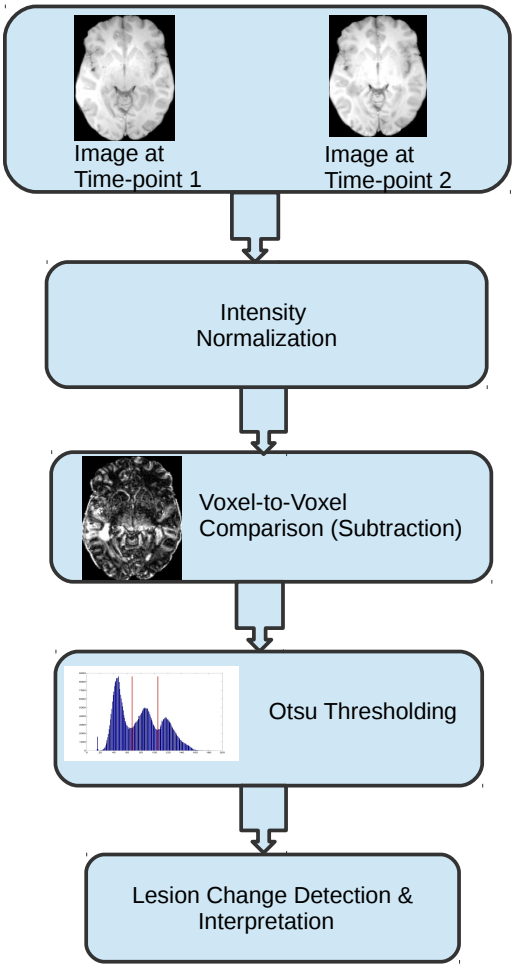


Figure 5.3: Block diagram for pipeline of the proposed architecture.

5.3 Experiments and Results

5.3.1 Dataset and Preprocessing

Two datasets were chosen to evaluate intensity normalization. The first dataset consists of 18 MS patients having 4 time-points each, approximately separated by a period of three to six months. The second dataset consists of 40 patients having 3 time points, approximately separated by a period of three months. Whole-brain MR images were acquired on MS patients in both datasets. T1-w MPRAGE, T2-w and FLAIR modalities were chosen for the experiment. Expert annotations of lesions were carried out by an expert radiologist on all MS patients. The volume size for T1-w MPRAGE and FLAIR is $256 \times 256 \times 160$

and voxel size is 1×1×1 mm^3. For T2-w, the volume size is 256×256×44 and voxel size is 1×1×3 mm^3. All imaging experiments for dataset 1 were performed on a 3T Siemens Verio (VB17) scanner with a 32-channel head coil. Dataset 2 consists of multi site data in which imaging experiments were performed on two 3T Siemens Verio (VB17) scanners and a 3T Philips Achieva scanner, both of them having 32-channel head coil.

For Gd related experiments, the acquired volumes were: a T1-w volume without contrast agent (pre-contrast) and T1-w Gd-enhanced (post-contrast). For dataset 1, the volume size of images is 256×256×44 and voxel size is 1×1×3 mm^3. For dataset 2, the volume size of images is 240×240×44 and voxel size is 1×1×3 mm^3. For dataset 1, radiologist performed the manual delineation of the enhanced lesions on T1 whereas for dataset 2 expert annotations of lesions for T1 Gd images are unavailable.

MR images from each patient are de-noised [Coupe 2008], bias field corrected [Tustison 2010] and registered with respect to T1-w MPRAGEvolume [Ourselin 2000, Commowick 2012a]. All the images are processed to extract intra-cranial region using **BET** (Brain Extraction Tool) [Smith 2002]. The block diagram for the pre-processing pipeline is shown in Figure 5.4.

We show the effect of longitudinal intensity normalization followed by detection on both normal tissues and lesions for MS subjects. The first time point is considered as the reference point to which the subsequent time points (moving ones) are aligned using intensity normalization. First, the parameters of reference and moving images are estimated using γ-divergence GMM estimator as described in Section 5.2.4. Secondly, the voxels of the moving image are aligned with respect to the reference image using the procedure in Section 5.2.6. Each patient and each time point $t = 2, ..., t_n$, are rigidly registered to the T1-w MPRAGE of the first time instance. The obtained difference image is processed further to obtain a soft detection by using heuristic thresholding iteratively (1) by Otsu's threshold [Otsu 1979]; (2) erosion of image by one voxel. The detections from this difference image are compared with difference image of ground truth at corresponding time points.

5.3.2 Intialization of GMM

Starting with some initial mixture model, the EM algorithm alternates between computing a lower bound of the log-likelihood and improving the current model with respect to this lower bound. The algorithm converges to a certain stationary point of the likelihood function. Unfortunately, the likelihood function is generally non-convex, possessing many stationary points, including small local maxima, and even worse, local minima and saddle points. Moreover, the convergence of the EM algorithm to either type of point highly depends

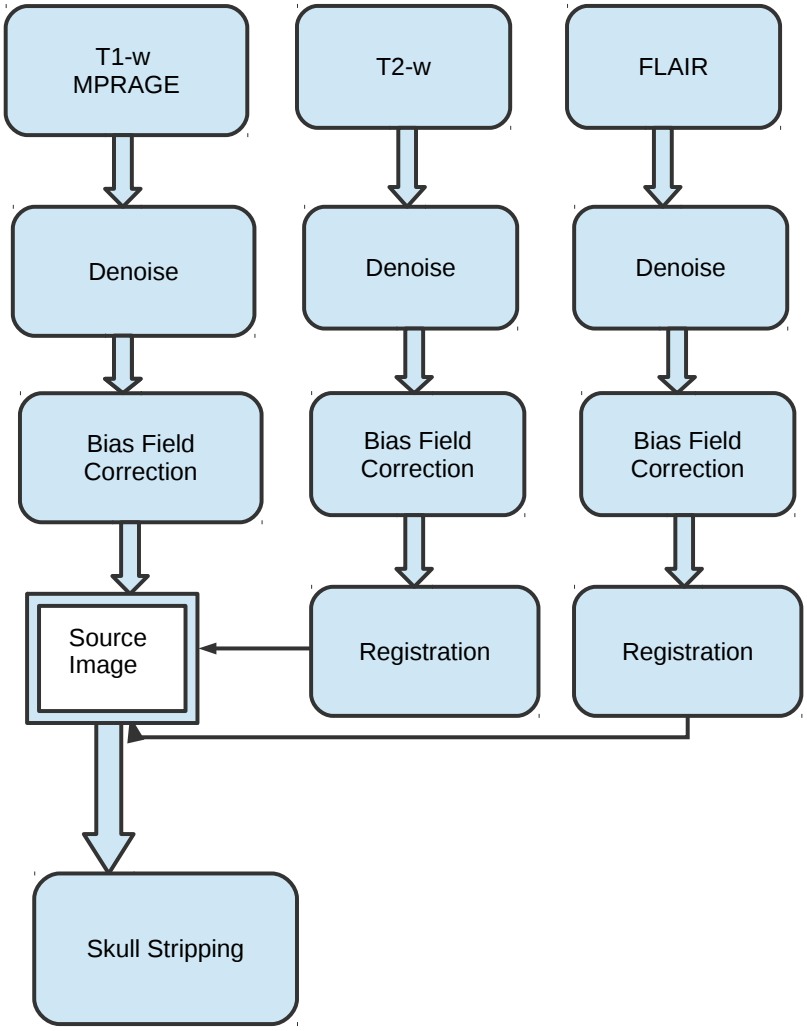


Figure 5.4: Block diagram pre-processing pipeline.

on the chosen initial mode. We obtain initial values for these parameters by running KMeans [MacQueen 1967] first.

5.3.3 Intensity Correction Evaluation

To evaluate the quality of intensity normalization, we compare the histograms of reference, moving and intensity normalized moving image using chi-squared distance given by $\chi^2_{x,y} = \frac{1}{2}\sum \frac{(x_i - y_i)^2}{x_i + y_i}$. Lower values of this distance indicate better alignment of intensities. Tables 5.1 and 5.2 report the chi-squared distance for datasets 1 and 2 for various imaging sequences. Different methods are compared against the proposed one. For first dataset, we report the mean

χ^2 distance for our method as 0.18(±0.045), 0.28(±0.037) and 0.32(±0.038) for T1-w MPRAGE, T2-w and FLAIR respectively. For the second dataset, we report the mean χ^2 distance for our method as 0.16(±0.062), 0.26(±0.027) and 0.30(±0.048) for T1-w MPRAGE, T2-w and FLAIR respectively. These performance metrics being χ^2 distance, it is evident that our proposed framework performs better than the rest.

For T1w-Gd dataset 1 and dataset 2, results are reported in Table 5.3. We report the mean χ^2 distance for our method as 0.12(±0.062), 0.16(±0.03) for dataset 1 and 2 respectively.

Table 5.1: Chi-squared distance analysis for histogram matching for dataset 1.

	Before Normalization	After Normalization		
Modality		Proposed	Hellier	Nyul
T1-w	0.56 (±0.03)	**0.18 (±0.045)**	0.35 (±0.029)	0.3 (±0.019)
T2-w	0.62 (±0.029)	**0.28 (±0.037)**	0.414 (±0.03)	0.315 (±0.042)
FLAIR	0.56 (±0.027)	**0.32 (±0.038)**	0.45 (±0.051)	0.39 (±0.045)

Table 5.2: Chi-squared distance analysis for histogram matching for dataset 2.

	Before Normalization	After Normalization		
Modality		Proposed	Hellier	Nyul
T1-w	0.3 (±0.08)	**0.16 (±0.062)**	0.25 (±0.039)	0.22 (±0.016)
T2-w	0.45 (±0.039)	**0.26 (±0.027)**	0.312 (±0.04)	0.325 (±0.034)
FLAIR	0.58 (±0.031)	**0.30 (±0.048)**	0.38 (±0.041)	0.41 (±0.055)

	Before Normalization	Method	After Normalization
dataset 1	0.34(±0.06)	Nyul	0.29(±0.01)
		Proposed	0.12(±0.06)
		Hellier	0.22(±0.03)
dataset 2	0.38(±0.04)	Nyul	0.31(±0.01)
		Proposed	0.16(±0.03)
		Hellier	0.27(±0.04)

Table 5.3: Chi-squared distance analysis for histogram matching for T1-w Gd dataset 1 and 2.

Figures 5.5 and 5.6 show the intensity correction results for T1-w MPRAGE, T2-w and FLAIR images. Three time points and their corresponding MR modalities of a subject are shown before and after normalization. Each row represents the imaging modality and each column depicts the first time point,

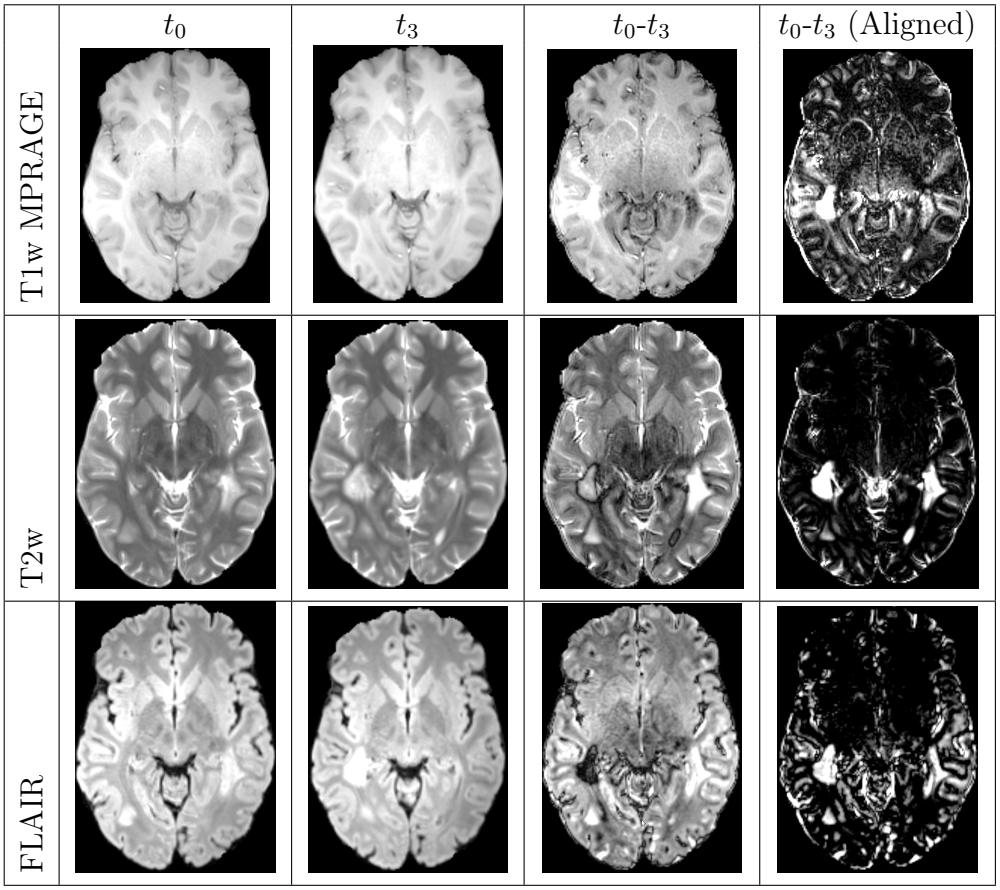


Figure 5.5: Intensity correction for one of the patient from dataset 1.

second time point, the absolute difference image without and with intensity normalization respectively. This figure demonstrates visually the ability of our approach to normalize intensities. As seen from the difference image of the first and second time points, intensity alignment reduces significantly the difference in intensities without affecting the lesion appearance. It will be easier to automatically detect evolving lesions on the images in the last column.

5.3.4 Longitudinal Lesion Detection

To show the quantitative improvement for identification of lesions, we report in Table 5.4 and 5.5 the precision (Positive Predicted Value) and recall (Sensitivity) of lesion detection averaged across the patients for various overlap thresholds. The lesion is said to be detected if $\frac{R_c \cap R_{GT}}{R_{GT}} \geq \varphi$ where R_c, R_{GT} and φ are respectively the candidate region in the image, the ground truth and a threshold. For datasets 1 and 2, Table 5.4 reports values of precision and recall for various thresholds. As from the figures, our approach outperforms other methods. For

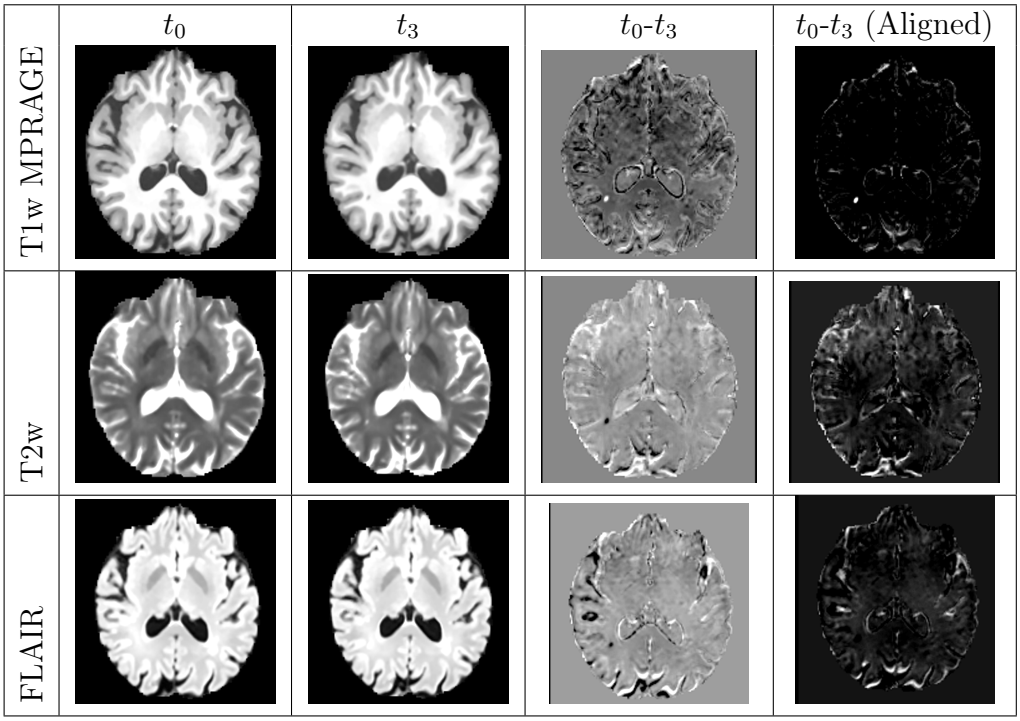


Figure 5.6: Intensity correction for one of the patients from dataset 2.

the dataset 1, we have a very high recall of 0.90 at $\varphi = 0.2$ and 0.82 even at $\varphi = 0.4$ where as for the dataset 2, high recall of 0.94 at $\varphi = 0.2$ and 0.78 even at $\varphi = 0.4$ is reported.

For evaluating the quality of MS lesion detection proposals, we follow the overlap criterion as described above. We compare the proposed method against Nyul and Hellier method using 1− Precision Vs Recall curves. 1−Precision is also known as False Discovery Rate (FDR) . We report Area Under Curve (AUC) as described for various overlap thresholds (φ) ranging from 0.1 to 0.8. Figures 5.7 and 5.8 demonstrate the AUC for dataset 1 and dataset 2 respectively. For proposed method, we report the AUC as 0.78 for dataset 1 and 0.74 for dataset 2. Both these plots clearly shows that our method as a better choice compared to other baseline approaches.

Table 5.4: Performance analysis for lesion detection on dataset 1.

	$\varphi = 0.1$		$\varphi = 0.2$		$\varphi = 0.4$	
Method	Precision	Recall	Precision	Recall	Precision	Recall
Nyul	0.47±0.5	0.52±0.4	0.63±0.01	0.60±0.02	0.58±0.03	0.54±0.03
Proposed	**0.61±0.04**	**0.62±0.05**	**0.73±0.04**	**0.90±0.05**	**0.63±0.03**	**0.82±0.01**
Hellier	0.59±0.04	0.65±0.02	0.65±0.02	0.74±0.03	0.62±0.03	0.59±0.05

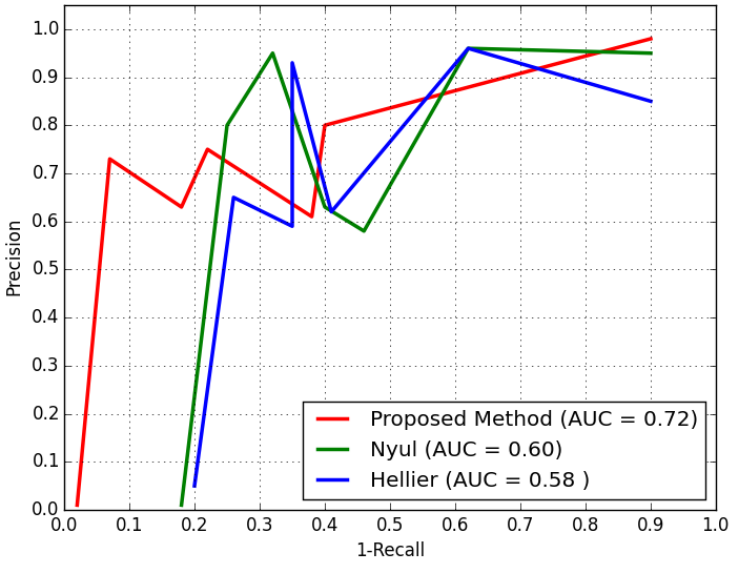


Figure 5.7: Precision vs 1-Recall curve for dataset 1.

Table 5.5: Performance analysis for lesion detection on dataset 2.

	$\varphi = 0.1$		$\varphi = 0.2$		$\varphi = 0.4$	
Method	Precision	Recall	Precision	Recall	Precision	Recall
Nyul	0.50±0.04	0.61±0.04	0.62±0.04	0.72±0.08	0.51±0.01	0.55±0.03
Proposed	**0.58±0.05**	**0.62±0.05**	**0.76±0.06**	**0.94±0.07**	**0.56±0.04**	**0.78±0.03**
Hellier	0.41±0.08	0.56±0.03	0.69±0.03	0.78±0.02	0.56±0.03	0.62±0.08

Table 5.6: Performance analysis for lesion detection on Gd-dataset 1.

	$\varphi = 0.1$		$\varphi = 0.2$		$\varphi = 0.4$	
Method	Precision	Recall	Precision	Recall	Precision	Recall
Nyul	0.42±0.04	0.53±0.6	0.060±0.04	0.64±0.04	0.62±0.02	0.55±0.04
Proposed	**0.59±0.03**	**0.68±0.03**	**0.77±0.05**	**0.92±0.06**	**0.65±0.04**	**0.79±0.04**
Hellier	0.56±0.05	0.62±0.04	0.63±0.01	0.75±0.02	0.61±0.04	0.57±0.03

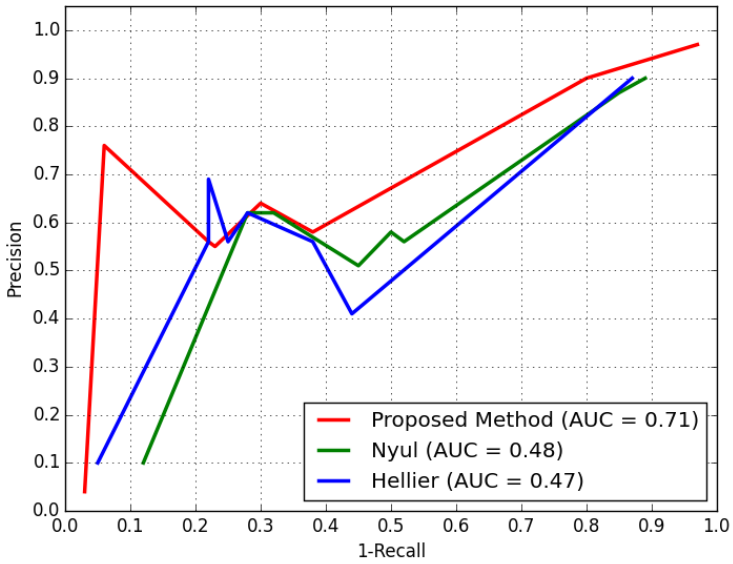


Figure 5.8: Precision vs 1-Recall curve for dataset 2.

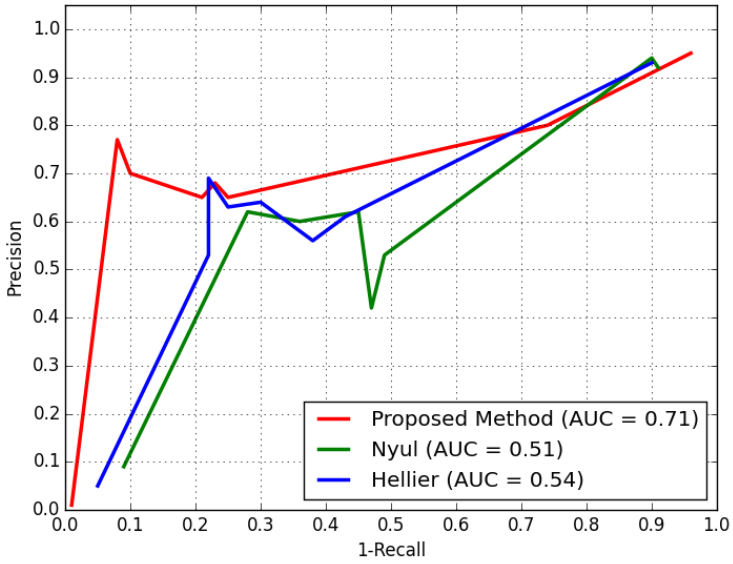


Figure 5.9: Precision vs 1-Recall curve for Gd dataset 1.

Figure 5.10 depicts the detected lesions for a representative image. The green label shows new lesions at t_3, orange shows stationary lesions which are also a part of t_1; blue shows false positive detections. We are able to accurately detect appearing and disappearing lesions thanks to the proposed method.

5.3.5 Active Gd-Enhanced Lesions Detection

Table 5.6 reports values of precision and recall for various thresholds for Gd-dataset 1. For this dataset, a high recall is achieved at 0.92 at $\varphi = 0.2$ and 0.79 even at $\varphi = 0.4$. Figure 5.11 shows the active lesion in T1-w Gd image and its detection. The 1−Precision vs Recall plot is constructed as described in Section 5.3.4. We report the AUC for proposed method as 0.73. It is evident from the plot that our proposed method outperforms other methods even for Gd dataset.

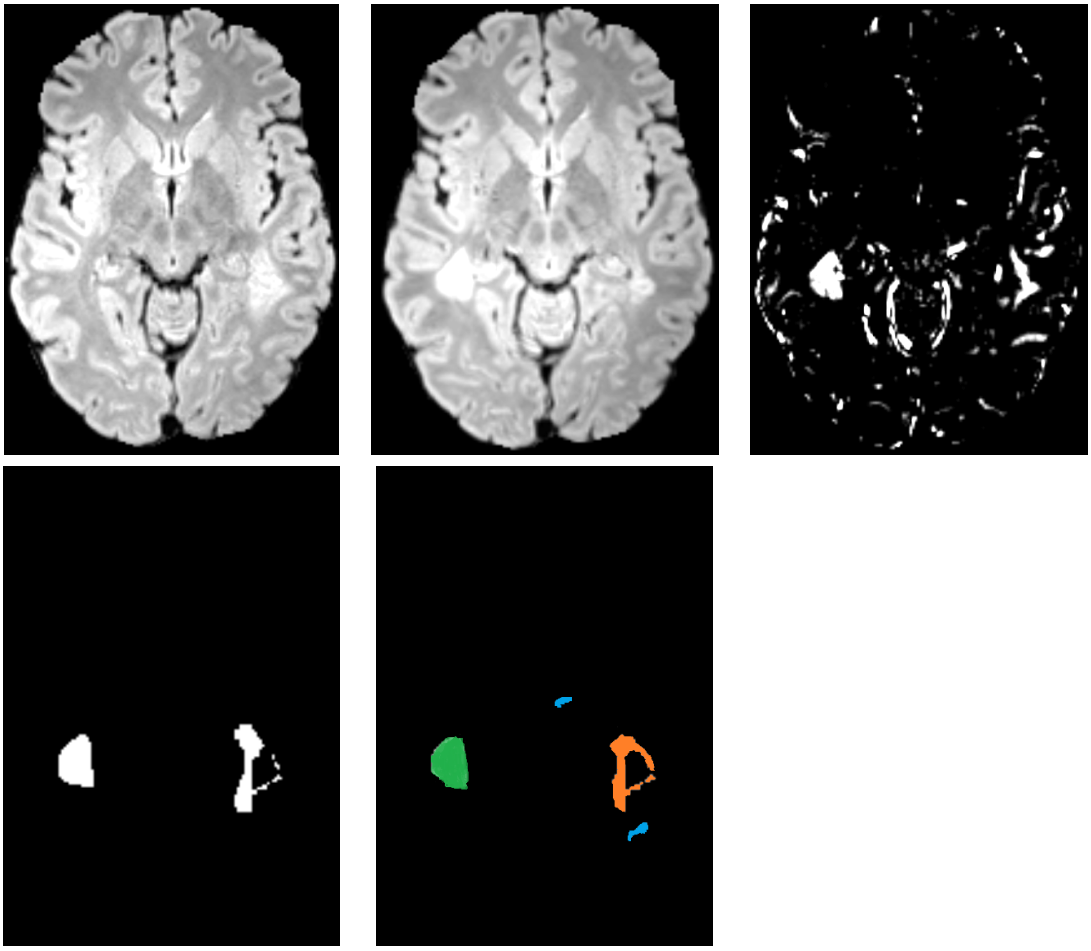


Figure 5.10: Lesion detection examples. For top and bottom, from left to right: Slice of FLAIR for t_0, t_3, $|t_0 - t_{3(Normalized)}|$, ground truth and lesions detected by our algorithm.

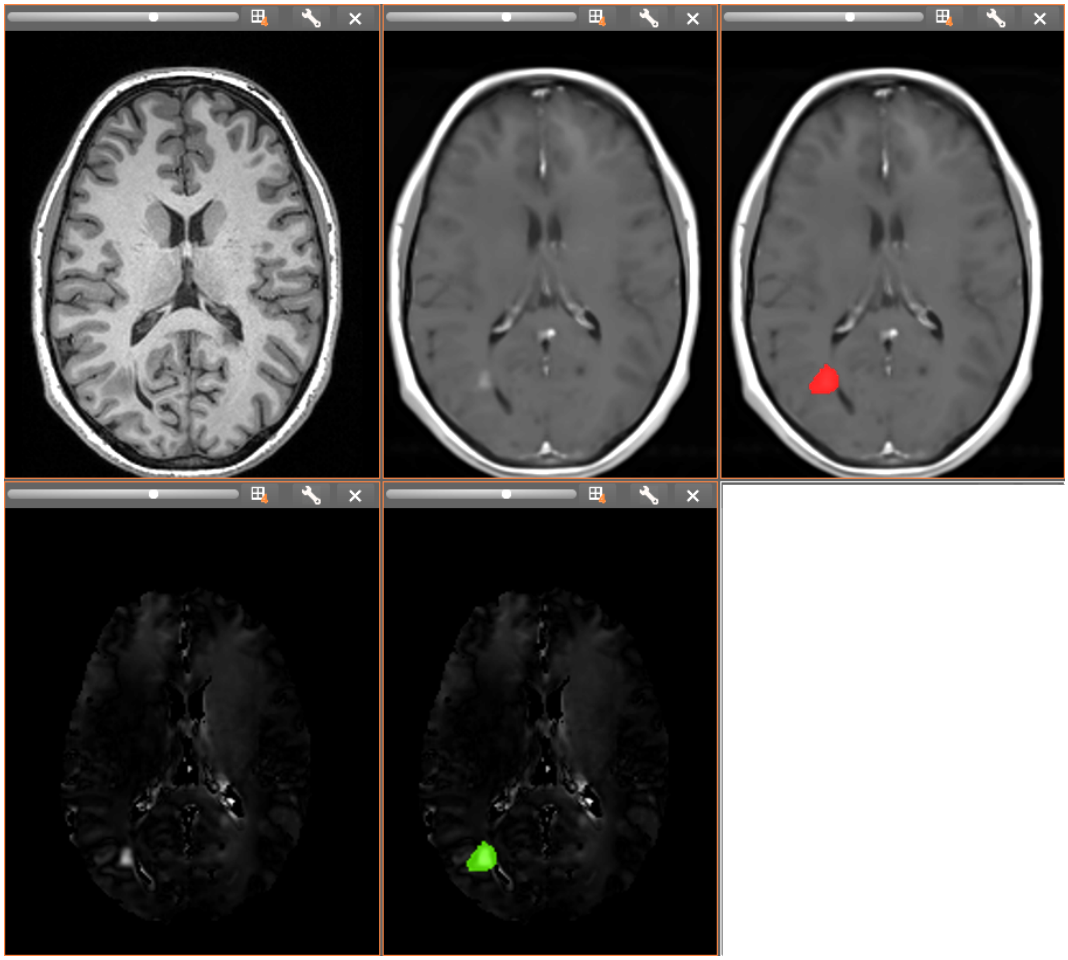


Figure 5.11: Top row: from left to right is the slice of T1-w SE(pre-contrast), T1-w Gd(post-contrast), T1-w Gd with its ground truth (red) respectively. Bottom row: left to right is difference image and detected active lesion (green) respectively.

5.3.6 Computational Complexity

The computational complexity of GMM using EM-algorithm is $O(klmn^3)$ where k is the number of clusters, l is a number of iterations, m is a number of samples and n is the number of dimensions in a data [Bishop 2006]. The C++ implementation of our intensity normalization framework takes 5 minutes on a single core machine (3.8 GB RAM, Intel Core i7 @2.40GHz, with 8 cores) per sequence. Here it should be noted that it is very fast compared to other techniques for robust GMM e.g. [Neykov 2007] with computational complexity of $O((klmn^3)^2)$.

5.4 Discussion and Conclusion

The method proposed in this chapter builds on the robust GMM estimation with γ divergence and addresses fundamental limitations of most of the state-of-the art algorithms i.e. intensity normalization without harming the lesion information. We have proposed an efficient change detection algorithm to track MS lesion evolution, based on subtraction images after intensity normalization accompanied by Otsu thresholding, which in turn allow to adapt the spatial information to the lesions unaffected. Furthermore, we have experimentally showed that our representation, compared to Nyul and Hellier method, is more suitable for longitudinal MS lesion analysis because of its ability to preserve the intensity variations caused by pathological changes.

The lesion information is very important for longitudinal studies. Our method is able to capture such information and adds performance to the MR serial change detection representation. However, the proposed method model is built only on difference of images. Thus, the static lesions do not get detected. Furthermore, we proposed a semi-automatic framework for the detection of Gd-Enhancing lesions which can be a very important tool for the assessment of active lesions.

Making each patient image adapt to the intensity space of an unbiased atlas is an interesting idea which could be then pursued to detect the lesion voxels based upon voxel-wise comparisons using statistical methods. However, since such a method can be very naive to lesion information, we propose in the next chapter a new method in the spirit of per image per patient intensity adaptation with respect to an atlas to detect the MS lesions.

Chapter 6

Robust Detection of Multiple Sclerosis Lesions from Intensity-Normalized Multi-Channel MRI

Contents

6.1 Introduction

To characterize MS lesions, we propose a novel paradigm based on a statistical patient to population comparison framework. It aims at studying the benefits of using multi-channel MRI to detect statistically significant differences between each individual MS patient and a database of control subjects. This framework consists in two components. First, intensity standardization is conducted to minimize the inter-subject intensity difference arising from the variability of the acquisition process and different scanners. The intensity normalization maps parameters obtained using a robust Gaussian Mixture Model (GMM) estimation not affected by the presence of MS lesions as described in chapter 5. The second part studies the comparison of multi-channel MRI of MS patients

with respect to an atlas built from the control subjects, thereby allowing us to look for differences in normal appearing white matter, in and around the lesions of each patient.

This chapter deals with a detection method based on the comparison between a set of controls and one MS patient. In their pioneering work of usage of Diffusion Tensor Imaging (DTI) in MS, Filippi et al. [Filippi 2001] evaluated the differences in fractional anisotropy (FA) between MS patients and controls on normal appearing white matter regions. Commowick et al. [Commowick 2008] proposed the use of the whole diffusion tensor information (DTI) to detect statistically significant differences between MS patients and controls. We propose a framework that relies on the multi-channel MRI information to detect statistically significant differences between MS patients and controls. The proposed technique is built on ideas similar to Commowick et al. [Commowick 2008] but but using conventional multi-channel MRI. Such an approach provides consistent and reliable lesion detections.

The proposed framework consists of two major parts. First, intensity standardization technique as described in Chapter 5 is incorporated for minimizing the inter subject intensity differences which may occur due to protocol variations across the scanners. Since in MS, White Matter (WM) lesions are also present in addition to healthy brain tissues, the intensities from multi-channel MRI are modeled with a parametric transform using a robust GMM estimation based on γ divergence. Such a formulation allows for keeping the lesions unaffected. It provides a technique that (1) uses tissue-specific intensity information by modeling them using a robust GMM; (2) provides a consistent intensity normalization between inter-subject images. The second part concerns the computation and use of an unbiased atlas of multi-channel MRI from a database of controls. Based on this atlas, we show how to compute statistical differences between a patient and the atlas using a Mahalanobis distance derived from the multi-channel MRI. Subsequently, we demonstrate its crucial role for further lesion analysis.

6.2 Methodology

6.2.1 Multiple Sclerosis Lesions (MSL) Detection

Our framework relies on the voxel-wise comparison of multi-channel MRI of an MS patient against a set of controls. Here, we assume that all images (patients and controls) are aligned into a common space where each voxel of each image describes the exact same spatial position. Commowick et al. [Commowick 2008, Commowick 2015] introduced a methodology for the

comparison of diffusion tensors between a patient and a group of controls.

We aim at comparing, for each voxel, the patient vector of intensities $\mathbf{x}_p$ and intensity vectors from control subjects $\mathbf{x}_q, q = 1, ..., M$. The control population $\mathbf{X} = \mathbf{x}_q, q = 1, ..., M$ is assumed to follow a multivariate Normal distribution $\mathcal{N}(\bar{\mathbf{X}}, \Sigma_{\mathbf{X}})$, where $\bar{\mathbf{X}}, \Sigma_{\mathbf{X}}$ denote respectively the average and covariance matrix of population $\mathbf{X}$. Following this assumption, we compute the difference statistic between $\mathbf{x}_p$ and $\mathcal{N}(\bar{\mathbf{X}}, \Sigma_{\mathbf{X}})$ as a Mahalanobis distance as

$$d^2(\mathbf{x}_p) = (\mathbf{x}_p - \bar{\mathbf{X}})^T \Sigma_{\mathbf{X}}^{-1} (\mathbf{x}_p - \bar{\mathbf{X}}) \tag{6.1}$$

where d^2 will vary between 0 and $+\infty$, getting smaller as the patient vector gets likely to belong the multivariate Normal distribution of the controls. A p-value can then be computed from this distance as the statistic $T = \frac{M(M-h)}{h(M^2-1)} d^2$ follows a Fisher distribution with parameters h and $M - h$: $T \sim F(h, M - h)$ where h is the vector dimension (1 for scalar images). The difference test p-value is therefore written as:

$$p(\mathbf{x}_p) = 1 - F_{h,M-h}\left(\frac{M(M-h)}{h(M^2-1)} d^2(\mathbf{x}_p)\right) \tag{6.2}$$

where $F_{h,M-h}$ is the cumulative distribution function of a Fisher distribution with parameters h and $M - h$.

6.2.2 Multiple Comparisons Correction

Finally, as we are computing voxel-wise comparisons, a final step is to correct for multiple comparisons utilizing FDR correction as introduced by Benjamini et al. [Benjamini 1995].

It is a set of tools of controlling the false discovery rate (FDR) in a set of statistical tests. FDR is the mean proportion of statistically significant test results that are really false positives. When we perform multiple statistical tests, the probability of false positive results rapidly increases. FDR control, like Bonferroni correction, reduces the probability of false positive results by using a more conservative alpha level for each test. FDR attempts to ensure that the great majority of statistically significant results are accurate but generally lets in a small proportion of false positives in the process.

6.3 Data Processing

Two datasets were chosen to evaluate intensity normalization. The first dataset consists of 18 MS patients while the second dataset consists of 40 patients.

Whole-brain MR images were acquired on MS patients in both datasets. T1-w MPRAGE, T2-w and FLAIR modalities were chosen for the experiment. Expert annotations of lesions were carried out by an expert radiologist on all MS patients. The volume size for T1-w MPRAGE and FLAIR is 256×256×160 and voxel size is 1×1×1 mm^3. For T2-w, the volume size is 256×256×44 and voxel size is 1×1×3 mm^3. All imaging experiments for dataset 1 were performed on a 3T Siemens Verio (VB17) scanner with a 32-channel head coil. Dataset 2 consists of multi site data in which imaging experiments were performed on two 3T Siemens Verio (VB17) scanners and a 3T Philips Achieva scanner, both of them having 32-channel head coil.

MR images from each patient are de-noised [Coupe 2008], bias field corrected [Tustison 2010] and registered with respect to T1w-MPRAGE volume [Ourselin 2000, Commowick 2012a]. All the images are processed to extract intra-cranial region [Smith 2002]. We built a geometrically unbiased atlas for each sequence from dataset of controls [Guimond 2000]. For each MS patient, nonlinear registration of the T1-w MPRAGE image onto the atlas was done [Commowick 2012b]. Patient to atlas registration consists in two steps. First, a global affine transformation was computed between the T1 images. Then, nonlinear registration was performed to locally align the anatomies of the atlas and the patient. The final transformations were then applied to T2-w and FLAIR images.

We built a composite vector image for each patient and control from T1-w MPRAGE, T2-w and FLAIR. The atlas is considered as the reference image to which all controls and patient images are aligned using intensity normalization as described in Chapter 5. Finally, we applied our framework (Section 6.2.1) to compare the differences in images and report the results in the next section.

We performed experiments of lesion detection, to enhance the interest of our approach with respect to (1) intensity normalization and (2) to combine several modalities for lesion detection. To this end, we compared detections using each individual sequence i.e. T1-w MPRAGE, T2-w or FLAIR individually with and without intensity normalization; and then on composite vector images formed from all sequences with and without intensity normalization. Figure 6.1 explains the architecture of the proposed framework.

6.4 Results

6.4.1 Quantitative Results

To show the quantitative improvement for identification of lesions, we report in Tables 6.1, 6.2, 6.3 and 6.4 show the precision (Positive Predicted Value)

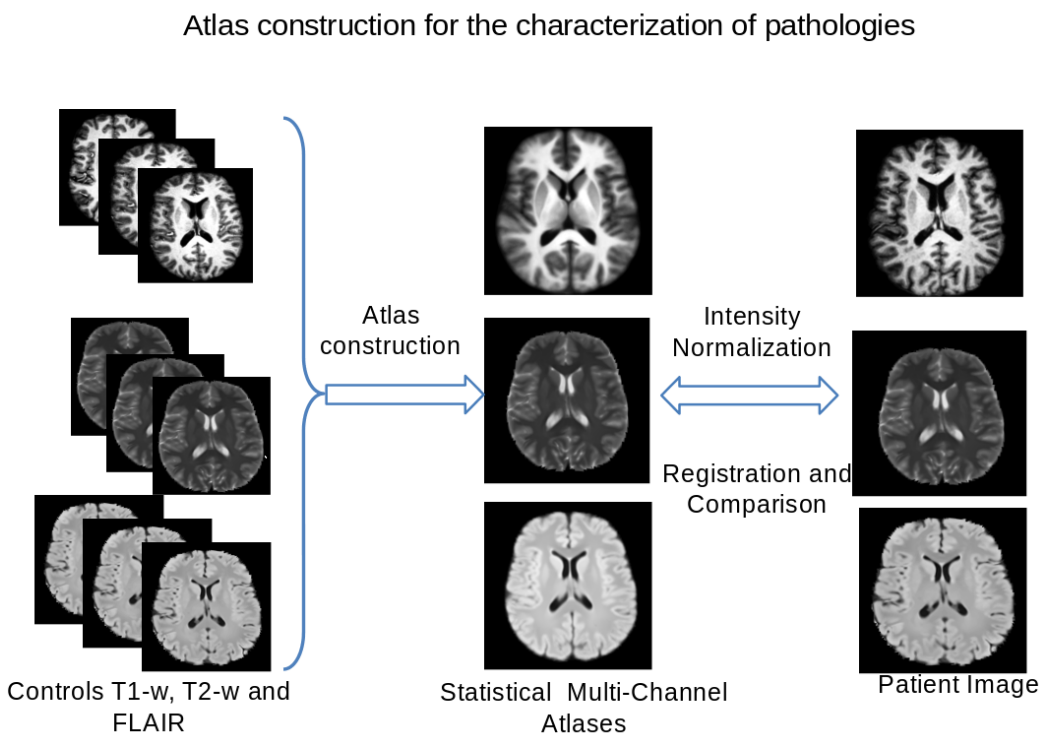


Figure 6.1: Block diagram of proposed framework.

and recall (Sensitivity) of lesion detection averaged across the patients for various overlap thresholds. The lesion is said to be detected if $\frac{R_c \cap R_{GT}}{R_{GT}} \geq \varphi$ where R_c, R_{GT} and φ are respectively the candidate region in the image, the ground truth and a threshold.

For dataset 1, we report in Tables 6.1 and 6.2 the precision and recall score average across 16 patients with and without intensity normalization respectively. For dataset 2, we report in Tables 6.3 and 6.4 the precision and recall score average across 40 patients with and without intensity normalization respectively. For dataset 1 , we report the best mean precision and recall score which is of T2-w as 0.75($\pm$0.064) and 0.92($\pm$0.041) respectively. For dataset 2, the best mean precision and recall score reported are for T2-w as 0.78($\pm$0.044) and 0.94($\pm$0.038) respectively are identified. From all the tables, it is clear that intensity normalization plays crucial role in our framework. The lesion detection of T2-w achieves good performance because some lesions may not be found on T1-w MPRAGE and FLAIR. For dataset 1, we report the mean precision and recall score of 0.61 ($\pm$0.02) and 0.64 ($\pm$0.03) on vector images without intensity normalization and 0.72 ($\pm$0.057) and 0.89 ($\pm$0.087) with intensity normalization. For dataset 2, we report mean precision and recall score of 0.63 ($\pm$0.012) and 0.66 ($\pm$0.04) on vector images without

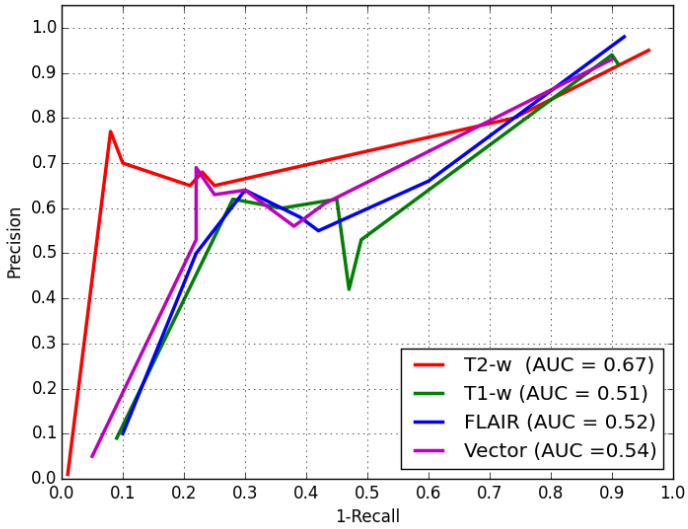


Figure 6.2: Precision vs 1-Recall curve for dataset 1 without intensity normalization.

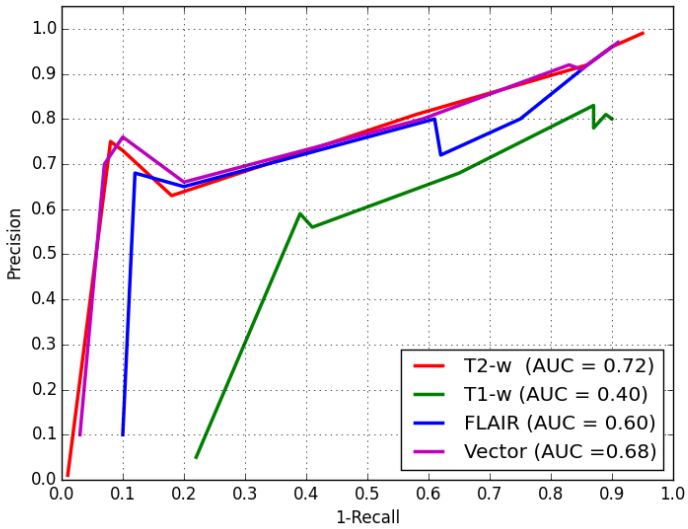


Figure 6.3: Precision vs 1-Recall curve for dataset 1 with intensity normalization.

intensity normalization and 0.76 ($\pm$0.047) and 0.92 ($\pm$0.075) with intensity normalization. Clearly, these results show a clear improvement on the best detection score from individual modalities. For evaluating the quality of MS lesion detection proposals, we follow the overlap criterion as described

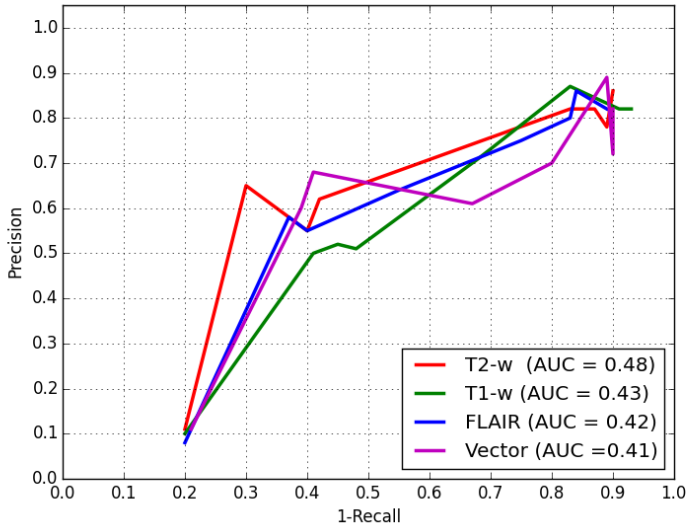


Figure 6.4: Precision vs 1-Recall curve for dataset 2 without intensity normalization.

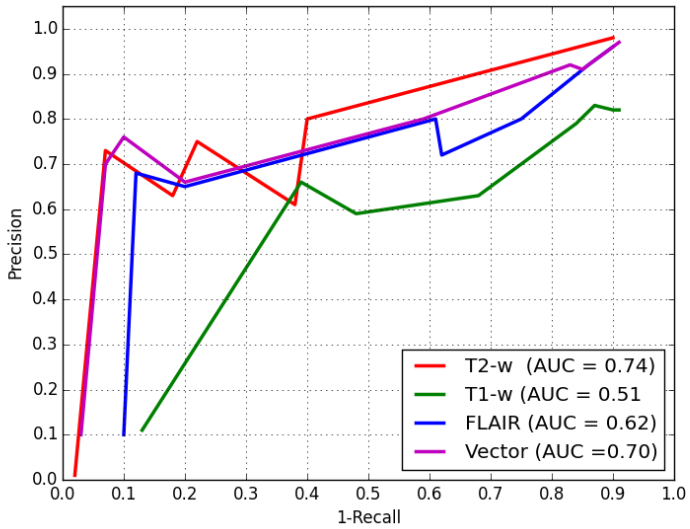


Figure 6.5: Precision vs 1-Recall curve for dataset 2 with intensity normalization.

above. We compare the different image sequences using Receiver Operating Characteristic (ROC) for Precision Vs $1-$Recall curves. We report Area Under Curve (AUC) as described for various overlap thresholds (φ) ranging from 0.1 to 0.8. Figures 6.2 and 6.3 demonstrate the ROC-AUC for T2-w of 0.48 and

0.74 for dataset 1 with and without intensity normalization. Figures 6.4 and 6.5 demonstrate the ROC-AUC for T2-w of 0.76 and 0.40 dataset 2 with and without intensity normalization. The plots 6.3 and 6.5 clearly show that T2-w as a better choice compared to other imaging sequences. The primary reason of better performance of T2-w sequence may be due to the fact that lesions are more pronounced on T2-w sequence compared to other sequences.

Table 6.1: The performance analysis of various image sequences for lesion detection on dataset 1 without intensity normalization. T1-w MPRAGE is denoted as T1-w MP.

	$\varphi = 0.1$		$\varphi = 0.2$		$\varphi = 0.4$	
Method	Precision	Recall	Precision	Recall	Precision	Recall
T1-w MP	0.71±0.04	0.32±0.04	0.55±0.03	0.58±0.04	0.54±0.02	0.52±0.04
T2-w	**0.66±0.05**	**0.40±0.03**	**0.62±0.05**	**0.67±0.04**	**0.58±0.05**	**0.61±0.03**
FLAIR	0.64±0.03	0.45±0.05	0.58±0.03	0.63±0.03	0.55±0.04	0.60±0.06
Vector	0.69±0.04	0.35±0.03	0.66±0.02	0.64±0.03	0.62±0.03	0.58±0.04

Table 6.2: The performance analysis of various image sequences for lesion detection on dataset 1 with intensity normalization. T1-w MPRAGE is denoted as T1-w MP.

	$\varphi = 0.1$		$\varphi = 0.2$		$\varphi = 0.4$	
Method	Precision	Recall	Precision	Recall	Precision	Recall
T1-w MP	0.68±0.03	0.35±0.04	0.59±0.04	0.61±0.03	0.56±0.03	0.59±0.04
T2-w	**0.81±0.04**	**0.42±0.05**	**0.75±0.06**	**0.92±0.04**	**0.73±0.05**	**0.90±0.04**
FLAIR	0.72±0.03	0.38±0.02	0.68±0.08	0.88±0.04	0.65±0.03	0.80±0.05
Vector	0.82±0.03	0.45±0.04	0.72±0.04	0.89±0.04	0.68±0.05	0.84±0.03

Table 6.3: The performance analysis of various image sequences for lesion detection on dataset 2 without intensity normalization. T1-w MPRAGE is denoted as T1-w MP.

	$\varphi = 0.1$		$\varphi = 0.2$		$\varphi = 0.4$	
Method	Precision	Recall	Precision	Recall	Precision	Recall
T1-w MP	0.51±0.05	0.52±0.06	0.52±0.04	0.55±0.05	0.50±0.01	0.59±0.03
T2-w	**0.55±0.05**	**0.68±0.05**	**0.65±0.06**	**0.70±0.07**	**0.62±0.04**	**0.58±0.03**
FLAIR	0.51±0.08	0.54±0.03	0.56±0.03	0.64±0.02	0.55±0.03	0.60±0.06
Vector	0.58±0.04	0.62±0.03	0.64±0.03	0.68±0.02	0.60±0.03	0.63±0.04

Table 6.4: The performance analysis of various image sequences for lesion detection on dataset 2 with intensity normalization. T1-w MPRAGE is denoted as T1-w MP.

	$\varphi = 0.1$		$\varphi = 0.2$		$\varphi = 0.4$	
Method	Precision	Recall	Precision	Recall	Precision	Recall
T1-w MP	0.75±0.03	0.20±0.04	0.63±0.04	0.65±0.03	0.60±0.04	0.58±0.03
T2-w	**0.82±0.03**	**0.50±0.03**	**0.72±0.03**	**0.94±0.05**	**0.70±0.04**	**0.86±0.03**
FLAIR	0.7±0.03	0.3±0.04	0.70±0.04	0.86±0.03	0.68±0.03	0.80±0.04
Vector	0.85±0.03	0.35±0.03	0.76±0.03	0.92±0.03	0.74±0.03	0.88±0.05

6.4.2 Qualitative Results

Figure 6.6 and 6.8 show the lesion detection results for T2-w where as 6.7 and 6.9 show the lesion detection results for composite vector image. These figures demonstrate visually the ability of our approach to detect lesions. As seen from the figures, there is considerable improvement of lesion detection, thanks to intensity normalization.

6.5 Conclusion

A novel paradigm for detection of MS lesions has been proposed. We devised a two layers strategy, first intensity standardization of images with respect to an atlas; and then computation of difference scores of a specific patient to a population of controls. Such comparisons allowed the accurate detection of lesions or other diffuse pathology-related changes. The efficacy of our method was evaluated though detection with and without intensity correction. Compared to un-normalized images, detection with intensity correction is a better choice for MS lesion analysis thanks to its ability to preserve the intensity variations caused by pathological changes while normalizing healthy tissue intensities. The resulting system is both efficient and accurate. This performance suggests that it can provide valuable assistance in detecting MS lesions in clinical routine with high reliability. The proposed method, however requires controlled environment of preprocessing techniques, specifically construction of atlases. In next chapter, we will study the lesion detection framework in adaptive settings.

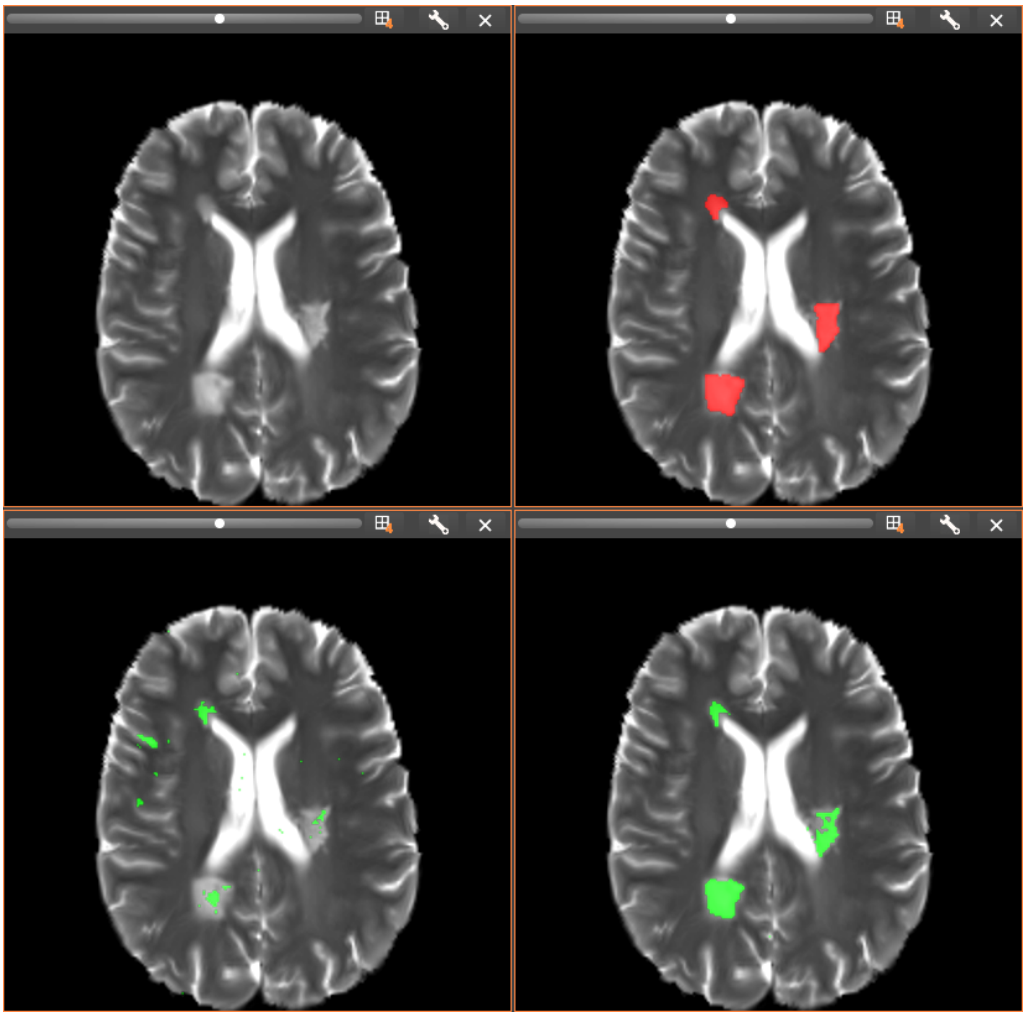


Figure 6.6: Top row : from left to right, a slice of MRI from T2-w Sequence from dataset 1, corresponding ground truth (red). Bottom row: from left to right, MSL detection (green) obtained by without and with intensity normalization.

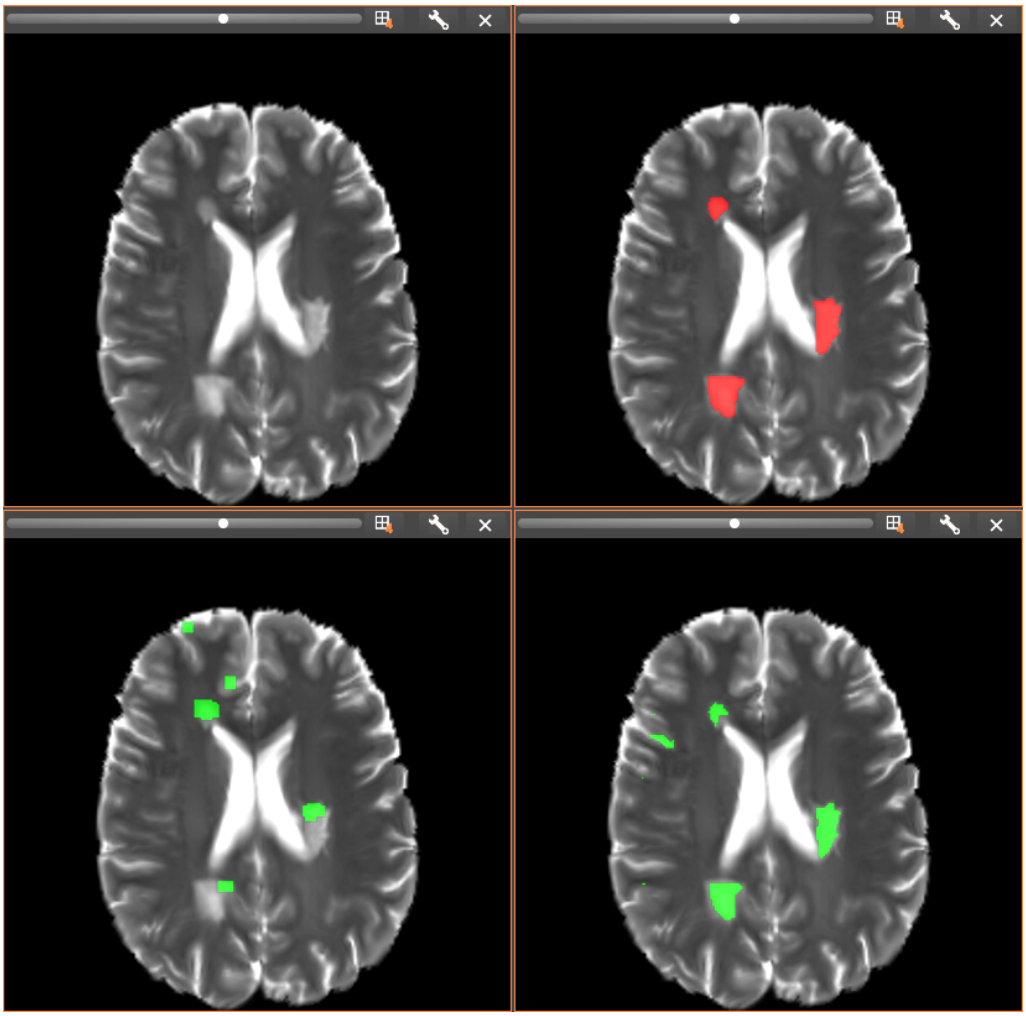


Figure 6.7: Top row : from left to right, a slice of MRI from composite vector image from dataset 1, corresponding ground truth (red). Bottom row: from left to right, MSL detection (green) obtained by without and with intensity normalization.

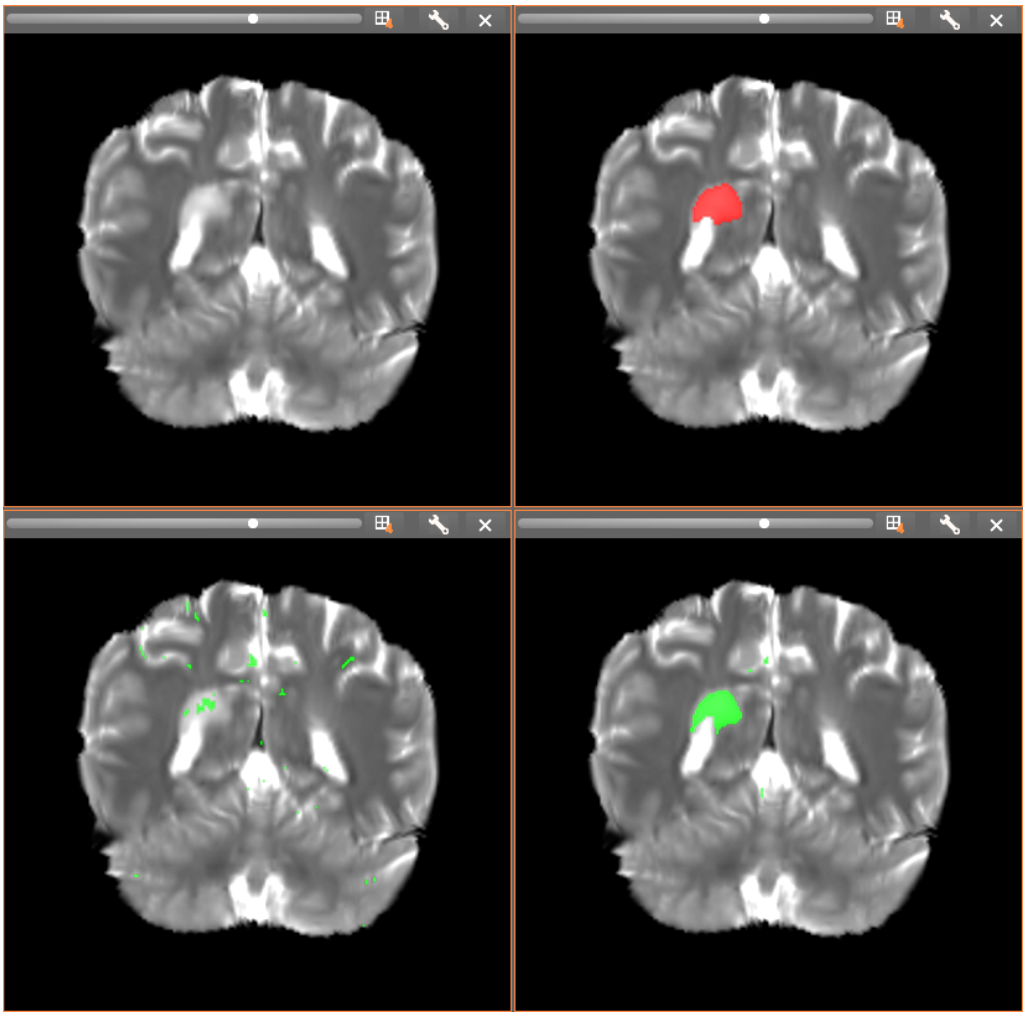


Figure 6.8: Top row : from left to right, a slice of MRI from T2-w Sequence from dataset 2, corresponding ground truth (red). Bottom row: from left to right, MSL detection (green) obtained by without and with intensity normalization.

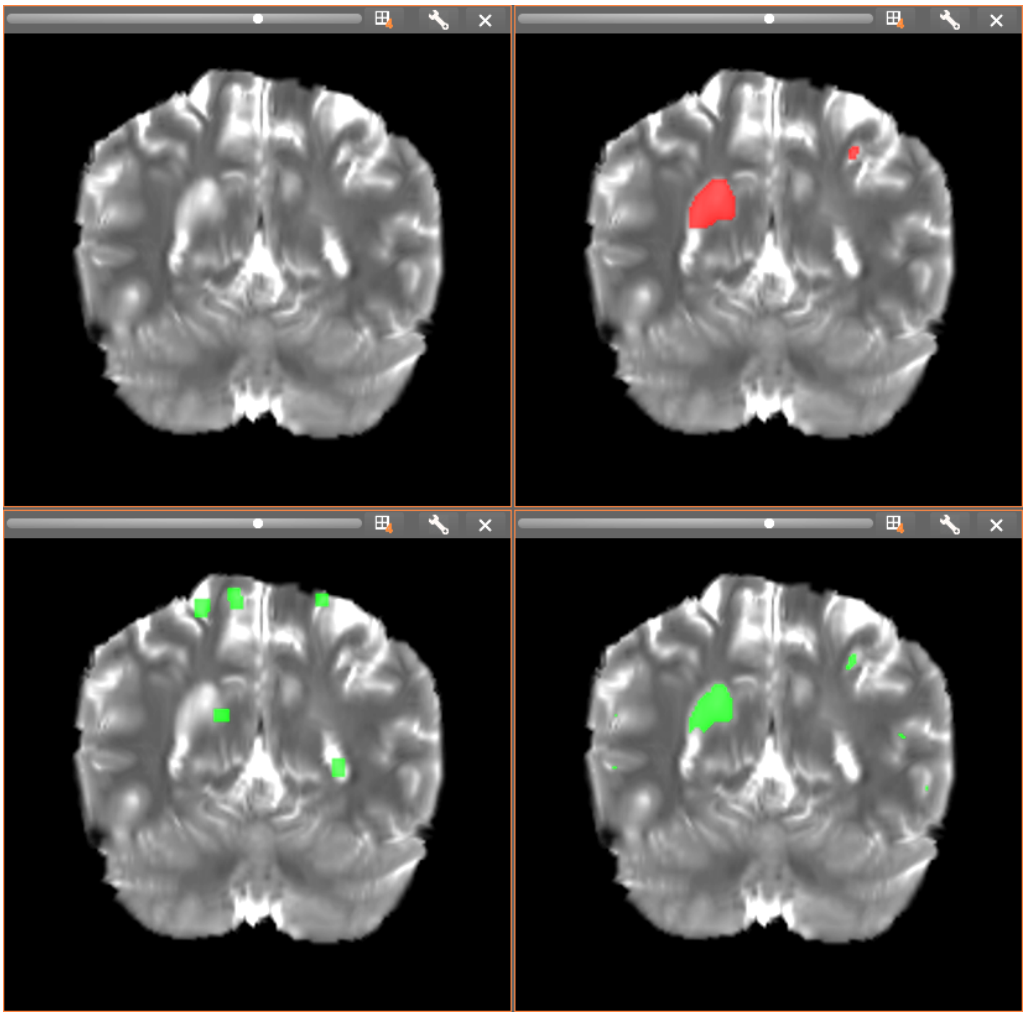


Figure 6.9: Top row : from left to right, a slice of MRI from composite vector image from dataset 2, corresponding ground truth (red). Bottom row: from left to right, MSL detection (green) obtained by without and with intensity normalization.

Chapter 7

Probabilistic One Class Learning for Multiple Sclerosis Lesions Detection

Contents

7.1 Introduction

The aim of any classification algorithm can be stated as a problem of assigning an object represented by a feature vector to a category of objects i.e. class. Using a set of training dataset examples, a classifier learns to label previously unseen objects from the test set. One class classification can be viewed as a special type of two-class classification problem, where data only from a single class is available for training the classifier. One class classifiers are applied when the data from other classes is extremely hard or impossible to collect. In the case of medical computer vision, obtaining ground truth data for big datasets is a cumbersome task. It is relatively easy to form training data for healthy volunteers. On the other side, collecting example data for the pathology can be rather expansive, since images need to be manually delineated by radiologists. Even if pathology examples are used in training, there is no way to guarantee that all types of pathology are included and thus recognized in a traditional binary classification problem. To cope with this problem, one-class classification problems are particularly adapted.. By just providing the non-pathological training data, one-class classifier learns a representative model of this data. If test sample is too different from the learned model, it is labeled as out-of-class with some degree of confidence.

This chapter presents an automatic algorithm for the detection of multiple sclerosis lesions (MS Lesions) from multi-sequence magnetic resonance imaging (MRI). We propose to use simple intensity features extracted from multi-parametric MRI for lesion detection. The contributions of our work are two-fold: (1) to build an automatic, probabilistic framework to discriminate NABTs and MS Lesions based on simple image representation i.e. bag of words features which is used to train a classifier at the voxel level; (2) generate probability map from the classifier and used to guide detection for a lesion voxel based on Otsu thresholding. Note that our framework is closely related to binary classification between two classes (NABT and MS Lesions) of observations. There exist subtle differences, though: binary classification is a symmetric setting for discrimination between two sets whereas we are interested in addressing the asymmetric problem of finding novel instances in one set relative to another. We focus on the form of the problem in which training samples (NABT) without anomalies (MS Lesions) are provided, and we calculate a lesion score (anomaly scores) for test data. The proposed framework is evaluated for two datasets consisting of 16 and 40 patients respectively. Our analysis reveals that our approach is well suited for MS Lesions detection and outperforms other benchmark approaches.

The chapter is organized as follows. In section 7.3, the framework for MS Lesions detection is presented. Section 7.4 gives the details of classifier,

development of probability score and detection based on probability score. Results and conclusion of the proposed methods are presented in sections 7.6 and 7.7.

7.2 State-of-the-art Methods for One Class Learning

In machine learning, one-class classification, also known as unary classification, attempts to identify objects of a specific class amongst all objects, by learning from a training set containing only the objects of that class. In this section, we discuss the major benchmark methods for one class learners.

7.2.1 Review of Binary SVM

Let us first take a look at the traditional two-class linear support vector machine (SVM) [Vapnik 1995]. Consider a data set $D = \left\{(\mathbf{x}_i, y_i)|\mathbf{x}_i \in \Re^d, y_i \in (-1, 1)\right\}_{i=1}^{N}$ where y_i is either -1 or $+1$, indicating the class to which the point $\mathbf{x}_i$ belongs.

We want to find the maximum-margin hyperplane that divides the points having $y_i = 1$ from those having $y_i = -1$. Any hyperplane can be written as the set of points $\mathbf{x}$ satisfying

$$\mathbf{w} \cdot \mathbf{x} - b = 0 \tag{7.1}$$

where $\cdot$ denotes the dot product and $\mathbf{w}$ a normal vector to the hyperplane. The parameter $\frac{b}{\|\mathbf{w}\|}$ determines the offset of the hyperplane from the origin along the normal vector $\mathbf{w}$. If the training data are linearly separable, we can select two hyperplanes in a way that they separate the data and there are no points between them, and then try to maximize their distance. The region bounded by them is called the margin. These hyperplanes can be described by the equations

$$\mathbf{w} \cdot \mathbf{x} - b = 1 \tag{7.2}$$

and

$$\mathbf{w} \cdot \mathbf{x} - b = -1 \tag{7.3}$$

By using geometry, the distance is found out between these two hyperplanes as $\frac{2}{\|\mathbf{w}\|}$, so the objective is to minimize $\|\mathbf{w}\|$. As we also have to prevent data points from falling into the margin, we add the following constraint:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1, \quad \text{for all } 1 \leq i \leq n. \tag{7.4}$$

The quadratic form of the optimization is given as follows

$$\arg\min_{(\mathbf{w},b)} \frac{1}{2}\|\mathbf{w}\|^2 \quad \text{subject to} \quad (\text{for any} \quad i = 1, \ldots, n) \qquad y_i(\mathbf{w} \cdot \mathbf{x_i} - b) \geq 1. \tag{7.5}$$

By introducing Lagrange multipliers $\boldsymbol{\alpha}$, the previous constrained problem can be expressed as

$$\arg\min_{\mathbf{w},b} \left\{ \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i [y_i(\mathbf{w} \cdot \mathbf{x_i} - b) - 1] \right\} \tag{7.6}$$

Now we briefly discuss about non-linear SVM [Shawe-Taylor 2004]. A very nice property of SVMs is that it can create a non-linear decision boundary by projecting the data through a non-linear function ϕ to a space with a higher dimension. This means that data points which can't be separated by a straight line in their original space I are projected to a feature space F where there can be a straight hyperplane that separates the data points of one class from an other. When that hyperplane is projected back to the input space I, it will have the form of a non-linear curve. The hyperplane is represented with the equation $w^T x + b = 0$, $w \in F$ and $b \in \Re$.

$$\arg\min_{w,b,\xi_i} \frac{||w||^2}{2} + C\sum_{i=1}^{n} \xi_i, \quad \text{subjected to} \quad y_i(w^T\phi(x_i) + b) \geqslant 1 - \xi_i \quad \forall i = 1, ..., n \tag{7.7}$$

with constraints $\xi_i \geqslant 0 \quad \forall i = 1, ..., n$. This minimization problem is solved using Lagrange multipliers with quadratic programming. The decision function (classification) rule for a data point x then becomes: $f(x) = sgn(\sum_{i=1}^{n} \alpha_i y_i K(x, x_i) + b)$. The function $K(x, x_i) = \phi(x)^T\phi(x_i)$ is known as the kernel function. Here α_i are the Lagrange multipliers; every $\alpha_i > 0$ is weighted in the decision function and thus supports the machine; therefore it is called Support Vector Machine. Since support vectors are considered to be sparse, there will be relatively few Lagrange multipliers with a non-zero value.

7.2.2 Kernel Function

The outcome of the decision function only relies on the dot-product of the vectors in the feature space F i.e. all the pairwise distances for the vectors, it is not necessary to perform an explicit projection to that space. As long as a function K has the same results, it can be used instead. This is known as the kernel trick and it is what gives SVM such a great power with non-linear separable data points; the feature space F can be of a very large dimension and thus the hyperplane separating the data can be very complex. Popular

choices for the kernel function are linear, polynomial and Gaussian radial basis function (RBF); amongst which RBF is used widely because of its single parameter σ to be optimized.

$$K(x, x^{'}) = exp\left(-\frac{\left\|x - x^{'}\right\|^2}{2\sigma^2}\right) \tag{7.8}$$

With this set of formulas and concepts, we are able to classify a set of data points into two classes with a non-linear decision function.

7.2.3 One class SVM

One class SVM was introduced in [Schölkopf 2001]. It tries to approximate the function which captures regions in the input space where the probability density of the data resides. Thus the function returns -1 in a small region (capturing the training data points) and $+1$ elsewhere. The quadratic programming minimization function is slightly different from the original stated above, but the similarity is still clear:

$$\arg\min_{w,b,\xi_i} \frac{\|w\|^2}{2} + \frac{1}{\nu n}\sum_{i=1}^{n}\xi_i - \rho \tag{7.9}$$

subjected to $(w^T\phi(x_i)) \geqslant \rho - \xi_i \; for \; all \; i = 1, ..., n$ and $\xi_i \geqslant 0 \; for \; all \; i = 1, ..., n$. In two-class SVM, the parameters C decides the smoothness. In this formula, it is the parameter ν that characterizes the solution. It does two things:(1) it sets an upper bound on the fraction of outliers (training examples regarded out-of-class); (2) it is a lower bound on the number of training examples used as Support Vector. Therefore, this algorithm is often referred to as ν-SVM. A standard Gaussian kernel is used for projecting feature vector into high dimensional space $\phi(x)$.

7.2.4 Minimum Covariance Determinant (MCD)

Rousseeuw et al. [Rousseeuw 1999] introduced the Minimum Covariance Determinant (MCD) estimator which is a robust estimator of a dataset covariance. The idea is to find a given proportion observations which are not outliers and compute their empirical covariance matrix. This empirical covariance matrix is then rescaled to compensate the performed selection of observations. Having computed the Minimum Covariance Determinant estimator, one can give weights to observations according to their Mahalanobis distance, leading the a reweighted estimate of the covariance matrix of the data set.

7.3 Methodology

7.3.1 Framework

We introduce a novel framework to identify MS lesions, as illustrated in Figure 7.1. It is based on two main stages. The first stage consists of two parts: (1) extraction of feature vectors from multi-channel MRI and their dimensionality reduction using Principle Component Analysis (PCA); (2) training a classifier as described in current and next sections. The second stage consists of testing patches from the patient image by applying the learned model. The feature designing from the patient patches is the same as mentioned above. Testing is performed by doing full search over an image by placing a patch at every voxel.

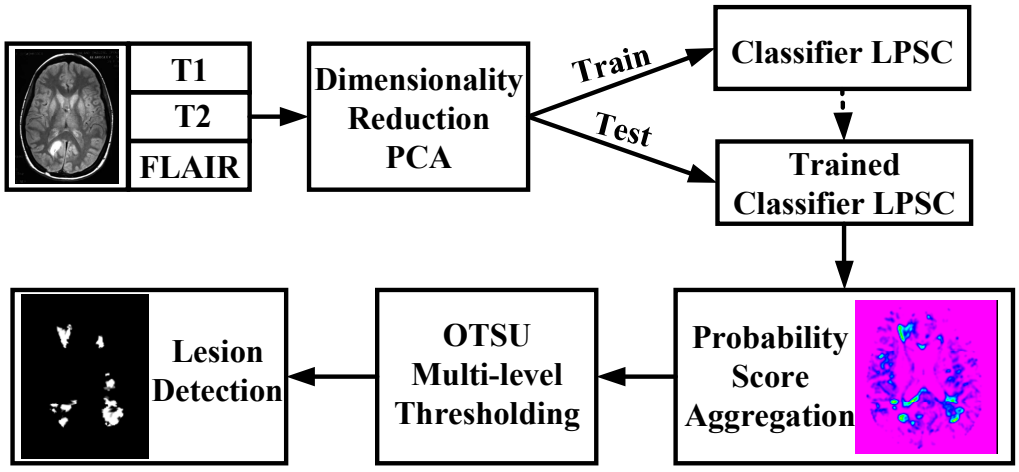


Figure 7.1: Workflow for the proposed MS Lesions detection.

7.3.2 Probabilistic Classification

We consider training data formed at each voxel by stacking local intensities from multiple modalities inside a surrounding patch of M voxels. To reduce the dimensionality of each vector, PCA is performed on these vectors, leading to a training set $D = \left\{(\mathbf{x}_i, y_i)|\mathbf{x}_i \in \mathcal{R}^d, y_i \in \mathcal{Y}\right\}_{i=1}^{N}$ where $y_i \in \mathcal{Y}$ is the corresponding class label and $\mathcal{Y} = (-1, +1)$ is the set of classes. The objective of probabilistic classification is to learn the class-posterior probability $p(y|\mathbf{x})$ of the training samples D. Based on the class-posterior probability, classification of a new sample $\mathbf{x}$ can be carried out

$$\widehat{y} := \underset{y \in \{-1,+1\}}{\arg\max}\, p(y|\mathbf{x}) \tag{7.10}$$

with confidence $p(\widehat{y}|\mathbf{x})$. One approach to tackle this problem is Least Squares Probabilistic Classification (LSPC) [Sugiyama 2010]. It employs the linear combination of kernel functions and its parameters are learned by regularized least-squares fitting of the true class-posterior probability.

It is possible to construct functions $q(y = i|\mathbf{x}, \theta_i)$ to estimate $p(y = i|\mathbf{x})$ for each $i \in \mathcal{Y}$ using the approximation form

$$q(y = i|\mathbf{x}, \theta_i) = \theta_i^T \phi(\mathbf{x}) \tag{7.11}$$

where $\theta_i = (\theta_{i,1}, \ldots, \theta_{i,B})^T \in \mathcal{R}^B$ for some number of B parameters, and $\phi(\mathbf{x}) = (K(\mathbf{x}, \mathbf{x_1}), \ldots, K(\mathbf{x}, \mathbf{x_B})^T \in \mathcal{R}^B)$ is a vector of kernel basis functions. Here, the Gaussian kernel $K(\mathbf{x}, \mathbf{x}') = exp(-\frac{1}{2\rho^2} \left\|\mathbf{x} - \mathbf{x}'\right\|^2)$, where ρ is the width of the kernel. By setting $B = N$, a kernel basis function at every training data sample is computed and for $B < N$, some random subset of the training data samples is used to obtain the vector of kernel basis function.

The analytical solution for θ_i is given by

$$\widehat{\theta_i} = (\Phi^T \Phi + \rho I_B)^{-1} \Phi \pi_i \tag{7.12}$$

where $\Phi = (\phi(\mathbf{x_1}), \ldots, \phi(\mathbf{x_N}))^T$ is the design matrix, I_B denotes B dimensional identity matrix, π_i is a column vector indicating membership of class i such that the j th element is one if $y_j = i$ and zero otherwise. In Equation (7.12), the effect of increasing the size of parameter ρ is both to regularize and to decrease the sensitivity to outliers. Finally, a posterior is obtained by normalizing over all classes as follows For $y \in \{-1, +1\}$, $p(y|\mathbf{x})$ can be estimated as follows

$$\widehat{p}(y = i|\mathbf{x}) = \frac{q(y = i|\mathbf{x}, \widehat{\theta_i})}{\sum_{j \in \{-1,+1\}} q(y = j|\mathbf{x}, \widehat{\theta_j})} \tag{7.13}$$

LSPC is a consistent estimator and is very fast to compute in practice, finding a global optimum for θ in a single step.

7.4 Lesion Detection Model

The framework in Section 7.3 was extended to one class learning in [Quinn 2014]. The basic premise holds an assumption that MS Lesions occupy low-density regions of the data space and that a kernel model can be used to characterize the high-density regions of NABT. We consider the case where MS Lesions are not present in the training data but present in the test data. Let $y = \{-1, +1\}$ be the NABT and MS Lesions class respectively. The task of MS Lesions detection is to assign a value to the estimate $\widehat{p}(y = +1|\mathbf{x})$ for test data $\mathbf{x}$ given

training data. The conditional probability of MS Lesions $p(y = +1|\mathbf{x}, \theta)$ with $q(y = +1|\mathbf{x}, \theta_{+1}) = 1 - \theta_{-1}^{T}\phi(x)$ can be expressed as

$$q(y = +1|\mathbf{x}, \theta_{+1}) = 1 - q(y = -1|\mathbf{x}, \theta_{-1}) \tag{7.14}$$

This can be estimated as discussed in Section 7.3. The parameters learned to model the NABT can therefore be used for MS Lesions detection.

7.4.1 Aggregate Probability Score

Since the probability map generated by the classification output scores of LSPC is noisy, we adopted the technique from [Reddy 2012] to rectify the probability score obtained from the classifier by smoothing it using a 3D Gaussian. For each voxel, its probability score is propagated to the neighborhood by using an isotropic Gaussian kernel. It produces a weighted average of each voxel's neighborhood, with the average weighted more towards the value of the central voxels. The Gaussian kernel for a voxel has a zero mean and standard deviation defined by the probability score of that particular voxel. The advantage of this method is that it establishes the spatial connectivity among voxels.

7.4.2 Thresholding Guided Detection

Automatic thresholding is used in final lesion detection. The basic idea is to automatically select an optimal probability threshold value for separating MS Lesions from the NABT based on their probability distribution. A common thresholding technique, the Otsu method [Otsu 1979], provides satisfactory results for thresholding an image with a histogram of bimodal distribution. This method however fails if the histogram is unimodal or close to unimodal. We used the revised Otsu method [Ng 2004] for selecting optimal threshold values for both unimodal and bimodal distributions. Figure 7.2 shows the pictorial representation of the proposed framework.

7.5 Experiments

7.5.1 Selection of NABT Patches

Over the years, researchers have proposed numerous approaches to represent visual information in form of feature representation spanning a wide spectrum, from local to global level and from low-level (bottom-up) to semantics (top-down). As far as spatial resolution is concerned, at one end, a voxel can be considered for representing the visual information. But there is generally not

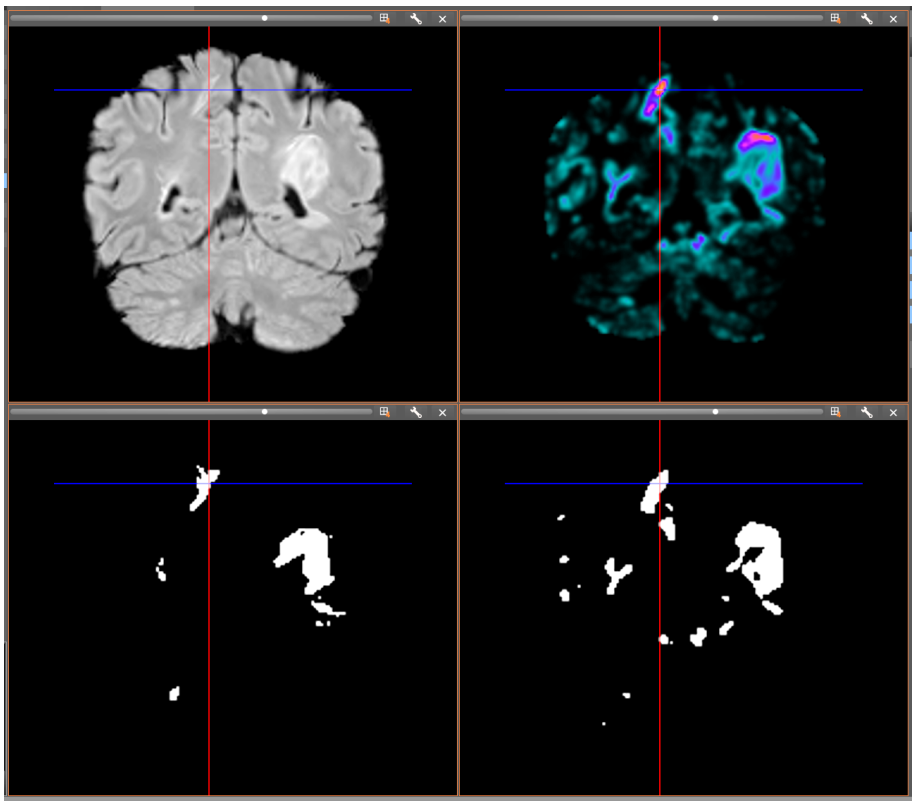


Figure 7.2: From top to bottom and left to right: Slice of FLAIR, aggregated probability map, ground truth and obtained lesion mask after Otsu thresholding.

enough information at the voxel level to make a useful feature for further learning frameworks. At other end, features computed on a whole image can be incorporated for training procedure, as for instance, Gabor and Wavelets based features. However, such a framework requires extraordinarily large amounts of training data, since one needs to represent all possible spatial configurations (scales and rotations) of images. Consequently, most of the studies use feature at intermediate scale i.e. features based on image patches. We use a representation based on the appearance of the patches. These patches can correspond to parts, objects called as visual words or visual phrases. They possess both representative and discriminative property such that they can be detected in a large number of images with high recall and precision.

The selection of patches plays a crucial role in training the classifier. Random patch selection may lead to over-fitting or under-fitting the classifier. Patches should cover all the regions of the brain anatomy. We select the NABT patches from the brain using the super-voxel technique. We briefly discuss the super-voxel technique. Simple Linear Iterative Clustering (SLIC) [Achanta 2012] is a simple and efficient method to decompose an image in visually homogeneous regions which yield super-voxels. It is based on a spatially localized version of k-means clustering.

The first phase is the initialization phase. The creation of cluster centers begins with an initialization of k clusters equally spaced throughout the volume, where k represents the desired approximate number of equally sized super-voxels. Taking I_{ni} to be the MRI signal intensity for the cluster center i in modality n and x, y, z being the spatial coordinates, the cluster centers $C_i = [I_{1i} I_{2i} ... I_{ni} x_i y_i z_i]^T$ are sampled on a lattice, spaced S voxels apart, where $S = \sqrt[3]{\frac{N}{k}}$, and N is the total number of voxels in the volume. Centers are moved to the lowest gradient position within their local 3^3 neighborhood to avoid centering a super-voxel on an edge and to reduce the chance that the seed voxel is a noisy voxel.

The next phase is the assignment phase, where each voxel i is associated with the nearest cluster center whose search region overlaps voxel i. The distance measure D determines the nearest cluster center for each voxel. Given that the expected spatial extent of a super-voxel is a volume of approximately S^3, the search for similar voxels is performed in a region of around $2S \times 2S \times 2S$ the super-voxel center. SLIC combines the intensity distance d_i and spatial distance d_s into a single metric, D. In order to combine the two distances into this single metric it is necessary to normalize both the spatial and intensity distances by their respective maximum distances within a cluster N_s and N_i. Therefore the normalized metric $D^{'}$ is formed by combining the spatial and

intensity distances:

$$d_i = \sqrt{(I_{1i} - I_{1j})^2 + (I_{2i} - I_{2j})^2 + ... + (I_{ni} - I_{nj})^2} \tag{7.15}$$

$$d_s = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \tag{7.16}$$

$$D^{'} = \sqrt{\left(\frac{d_i}{N_i}\right)^2 + \left(\frac{d_s}{N_s}\right)^2} \tag{7.17}$$

The maximum spatial distance expected within a given super-voxel corresponds to the sampling interval

$$S = \sqrt[3]{\frac{N_s}{k}} \tag{7.18}$$

Determining the maximum intensity distance is not as straightforward however, as intensity distances vary from super-voxel to super-voxel. This problem is solved by setting the normalized intensity distance to a constant m so that the distance equation becomes:

$$D = \sqrt{d_i^2 + \frac{d_s^2}{D^{'}} m^2} \tag{7.19}$$

Creating the distance measure in this way allows for m to be used to weigh the relative importance of the intensity similarity and the spatial proximity. When m is large, spatial proximity is weighed more heavily and this produces compact super-voxels. Conversely, when m is small the super-voxels created follow image boundaries closely and have more irregular sizes and shapes. S and m can be thought of as the average expected spatial and intensity distances inside a super-voxel, respectively. Compact super-voxels can be desirable because they more often correspond to a lattice structure in the volume, and their boundaries are simpler which leads to more regular neighborhood relationships. The compactness parameter m can also be described in relation to the intensity range in the volume $m = range_{\%} \cdot (max(I) - min(I))$. This definition yields a more intuitive procedure for the user as now the user needs to input a target average intensity range for the super-voxels, and this is scaled to each particular volume. The assignment and update phases are repeated iteratively until the L^2 norm error between the new cluster center locations and the old locations converge. Alternatively, a fixed number of iterations can be used.

The final phase is a post-processing phase to enforce connectivity by reassigning any disjoint voxels to nearby super-voxels, and ensuring that all voxels within a label are connected components within 26-neighbor connectivity. In this process, a minimum super-voxel size can also be enforced.

The patches from boundary of super-voxels and its centroid are chosen as shown in the Figure 7.3. We used the python-implementation of SLIC from sci-kit image [van der Walt 2014]. The number of super-voxels k in the segmented output image is chosen as 5000. The compactness factor m is kept as 30. The contours of the segmented output image is obtained by connected component labeling. The patches which overlap by more than 50% are removed from the training.

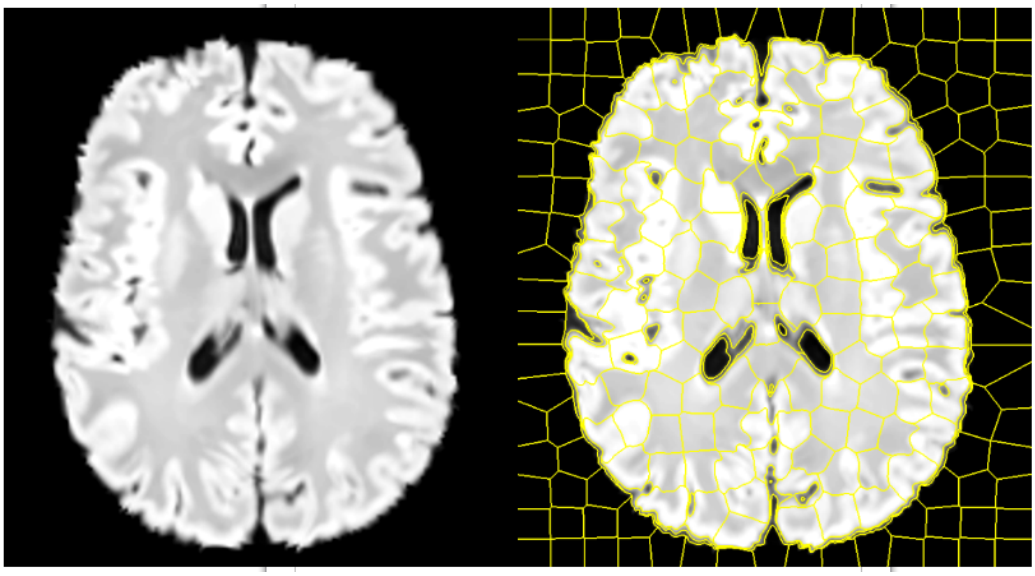


Figure 7.3: Super-voxels using SLIC on control subject on FLAIR sequence.

7.5.2 Data

Two datasets were chosen to evaluate intensity normalization. The first dataset consists of 18 MS patients while the second dataset consists of 40 patients. Whole-brain MR images were acquired on MS patients in both datasets. T1-w MPRAGE, T2-w and FLAIR modalities were chosen for the experiment. Expert annotations of lesions were carried out by an expert radiologist on all MS patients. The volume size for T1-w MPRAGE and FLAIR is 256×256×160 and voxel size is 1×1×1 mm^3. For T2-w, the volume size is 256×256×44 and voxel size is 1×1×3 mm^3. All imaging experiments for dataset 1 were performed on a 3T Siemens Verio (VB17) scanner with a 32-channel head coil. Dataset 2 consists of multi site data in which imaging experiments were performed on two 3T Siemens Verio (VB17) scanners and a 3T Philips Achieva scanner, both of them having 32-channel head coil. MR images from each patient are de-noised [Coupe 2008], bias field corrected [Tustison 2010] and registered with respect to T1-MPRAGE volume. All images are processed to

extract the intra-cranial region. Intensity normalization with respect to an unbiased atlas is performed as described in chapter 5.

7.5.3 Experimental Setup

NABT patches were collected from 20 healthy volunteers. The trained classifiers were tested on MS subjects. The features are intensity values extracted from 3^3 patches. The concatenated feature vector was formed using voxel values by extracting patches from FLAIR, T2-w and T1-w MPRAGE. The dimensionality reduction of feature vector was done using PCA. The number of components was decided by keeping 90% of the total variance.

In this experiment, in context of Equation (7.12), we searched the regularization parameter ρ and Gaussian kernel width σ over a range of $10^{[-3:1:2]}$ and $\{m/10, m/5, m/2, m, 3m/2\}$ respectively, where m is the median of kernel matrix $\phi(x)$ as described in Section 7.2.2. We chose the best ρ according to the validation set. This set consists of two MS patients which are not included in the test dataset.

We compared results with other benchmark methods: One Class SVM (OSVM) [Section 7.2.3] and the Minimum Covariance Determinant estimator (MCD) [Section 7.2.4]. For both OSVM and MCD, we chose the best parameters ν for OSVM and concentration of contamination according to the validation set. The validation set is same as used in proposed method. For OSVM, the Gaussian width is set to the median distance between samples, which has been shown to be a useful heuristic [Schölkopf 2001]. We report the results for $\nu = 0.2$ for OSVM. For MCD, the amount of contamination of the data set is assumed to be 0.30.

7.6 Results and Discussion

Table 7.1: Performance analysis for lesion detection on dataset 1 with intensity normalization.

	$\varphi = 0.1$		$\varphi = 0.2$		$\varphi = 0.4$	
Method	Precision	Recall	Precision	Recall	Precision	Recall
OCSVM	0.32±0.04	0.52±0.08	0.63±0.01	0.65±0.02	0.56±0.03	0.60±0.03
Proposed	**0.35±0.08**	**0.55±0.04**	**0.79±0.02**	**0.94±0.03**	**0.7±0.02**	**0.83±0.01**
MCD	0.28±0.04	0.39±0.05	0.55±0.02	0.65±0.03	0.40±0.03	0.52±0.05

In order to investigate the effect of intensity normalization on classifier to distinguish between lesion and NABT, we conducted lesion detection experiments with and without intensity normalization. We report the quantitative

Table 7.2: Performance analysis for lesion detection on dataset 1 without intensity normalization.

	$\varphi = 0.1$		$\varphi = 0.2$		$\varphi = 0.4$	
Method	Precision	Recall	Precision	Recall	Precision	Recall
OCSVM	0.27±0.02	0.54±0.09	0.58±0.02	0.64±0.05	0.53±0.04	0.58±0.02
Proposed	**0.30±0.09**	**0.53±0.04**	**0.75±0.04**	**0.89±0.02**	**0.72±0.04**	**0.79±0.02**
MCD	0.26±0.04	0.40±0.05	0.51±0.02	0.61±0.04	0.35±0.05	0.45±0.05

Table 7.3: Performance analysis for lesion detection on dataset 2 with intensity normalization.

	$\varphi = 0.1$		$\varphi = 0.2$		$\varphi = 0.4$	
Method	Precision	Recall	Precision	Recall	Precision	Recall
OCSVM	0.38±0.08	0.55±0.06	0.66±0.04	0.70±0.07	0.54±0.08	0.63±0.04
Proposed	**0.42±0.06**	**0.52±0.06**	**0.76±0.04**	**0.96±0.04**	**0.70±0.05**	**0.82±0.04**
MCD	0.26±0.03	0.46±0.06	0.56±0.02	0.65±0.06	0.48±0.02	0.49±0.04

Table 7.4: Performance analysis for lesion detection on dataset 2 without intensity normalization.

	$\varphi = 0.1$		$\varphi = 0.2$		$\varphi = 0.4$	
Method	Precision	Recall	Precision	Recall	Precision	Recall
OCSVM	0.29±0.04	0.40±0.05	0.62±0.04	0.64±0.04	0.51±0.05	0.58±0.03
Proposed	**0.55±0.04**	**0.48±0.05**	**0.72±0.05**	**0.82±0.03**	**0.64±0.06**	**0.76±0.04**
MCD	0.25±0.02	0.44±0.04	0.53±0.01	0.62±0.04	0.44±0.04	0.45±0.07

improvement for identification of lesions for two datasets. For dataset 1, we report in Tables 7.1 and 7.2 the precision and recall score average across 16 patients with and without intensity normalization respectively. For dataset 2, we report in Tables 7.3 and 7.4 the precision and recall score average across 40 patients with and without intensity normalization respectively. The lesion is said to be detected if $\frac{R_c \cap R_{GT}}{R_{GT}} \geq \varphi$ where R_c, R_{GT} and φ are respectively the candidate region in the image, the ground truth and a threshold. As from the figures, our approach outperforms other methods. For intensity-normalized dataset 1 and dataset 2, we have a very high recall of 0.94 and 0.96 at $\varphi = 0.2$ and even 0.83 and 0.82 at $\varphi = 0.4$. On other hand, For un-normalized dataset 1 and dataset 2, we have a very high recall of 0.89 and 0.82 at $\varphi = 0.2$ and even 0.79 and 0.76 at $\varphi = 0.4$. From all the tables, it is clear that intensity normalization plays crucial role in our framework. For evaluating the quality of MS lesion detection proposals, we follow the overlap criterion as described above. We compare the different benchmark methods using Receiver Operating Characteristic (ROC) for Precision Vs 1−Recall curves. We report Area Under Curve (AUC) as described for various overlap thresholds (φ) ranging from 0.1 to 0.8. Figures 7.5 and 7.4 demonstrate the ROC-AUC for proposed framework of 0.71 and 0.65 for dataset 1 with and without intensity normalization. Figures 7.7 and 7.6 demonstrate the ROC-AUC for proposed framework of 0.76 and 0.56 for dataset 2 with and without intensity normalization.

Both benchmark methods have a lower performance. A potential reason for OSVM to perform slightly worse is because it is not easily scalable. It becomes very slow when there are more than 10 or 20 thousand training points. Besides, the NABT class has a large intra class variance. MCD also results in lower performance as it may fail to capture the multi-modal distribution of data. Our framework alleviates these problems by including a large training dataset estimating correct data distribution.

Figures 7.8 and 7.9 show the lesion detection results for FLAIR sequence . The detections are shown in three different orientations. In each of the image, the top row represents the image, image embedded with ground truth and lesion detections obtained by OCSVM. Bottom row shows detections obtained by MCD and proposed framework respectively. These figure demonstrates visually the ability of our approach to detect lesions. As seen from the last column of bottom row, there is considerable improvement of lesion detection, thanks to the proposed framework.

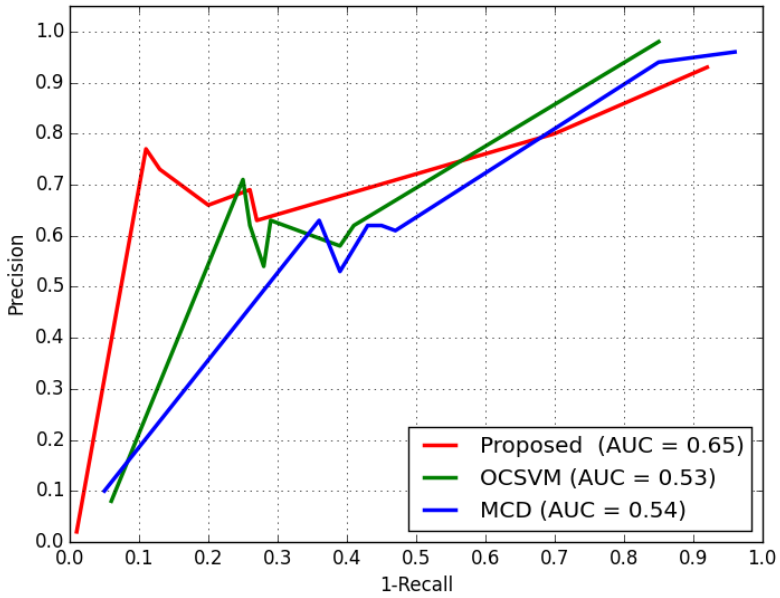


Figure 7.4: Precision vs 1-Recall curve for dataset 1 without intensity normalization.

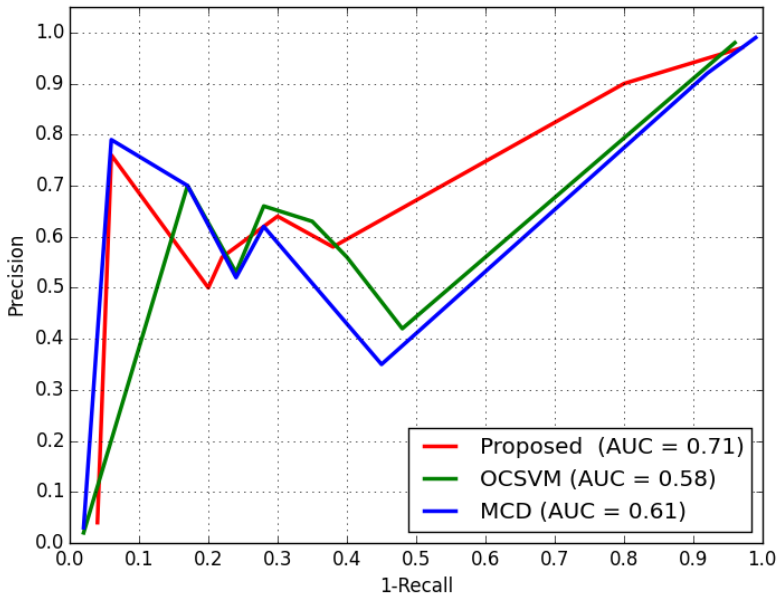


Figure 7.5: Precision vs 1-Recall curve for dataset 1 with intensity normalization.

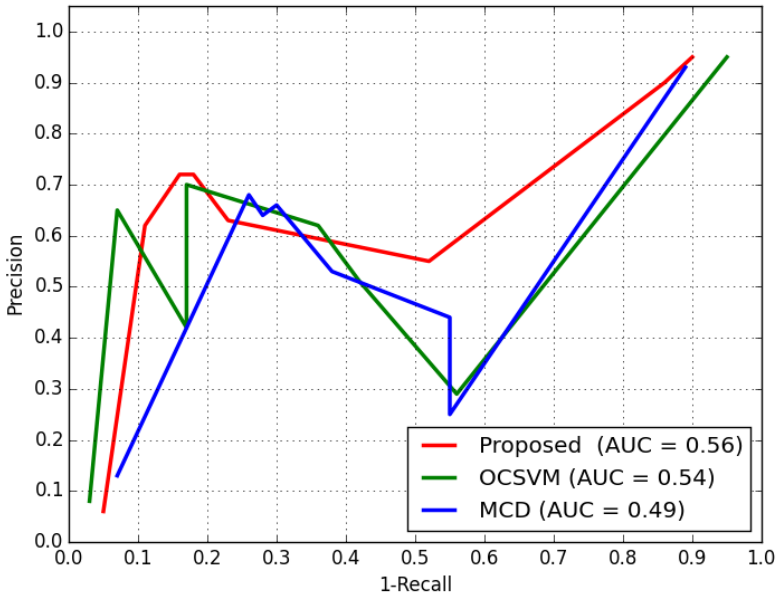


Figure 7.6: Precision vs 1-Recall curve for dataset 2 without intensity normalization.

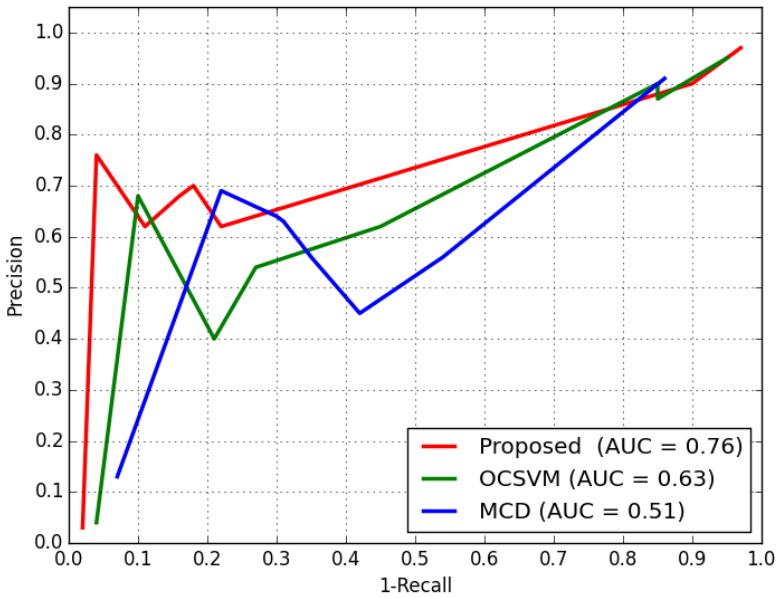
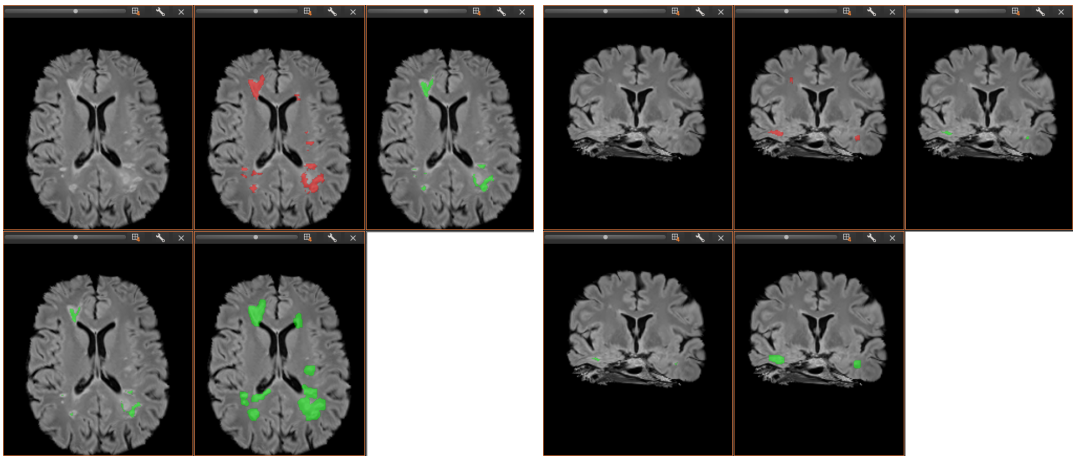
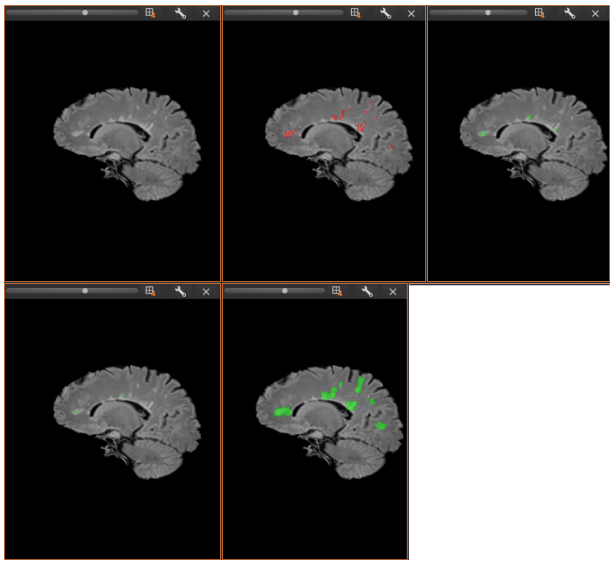


Figure 7.7: Precision vs 1-Recall curve for dataset 2 with intensity normalization.

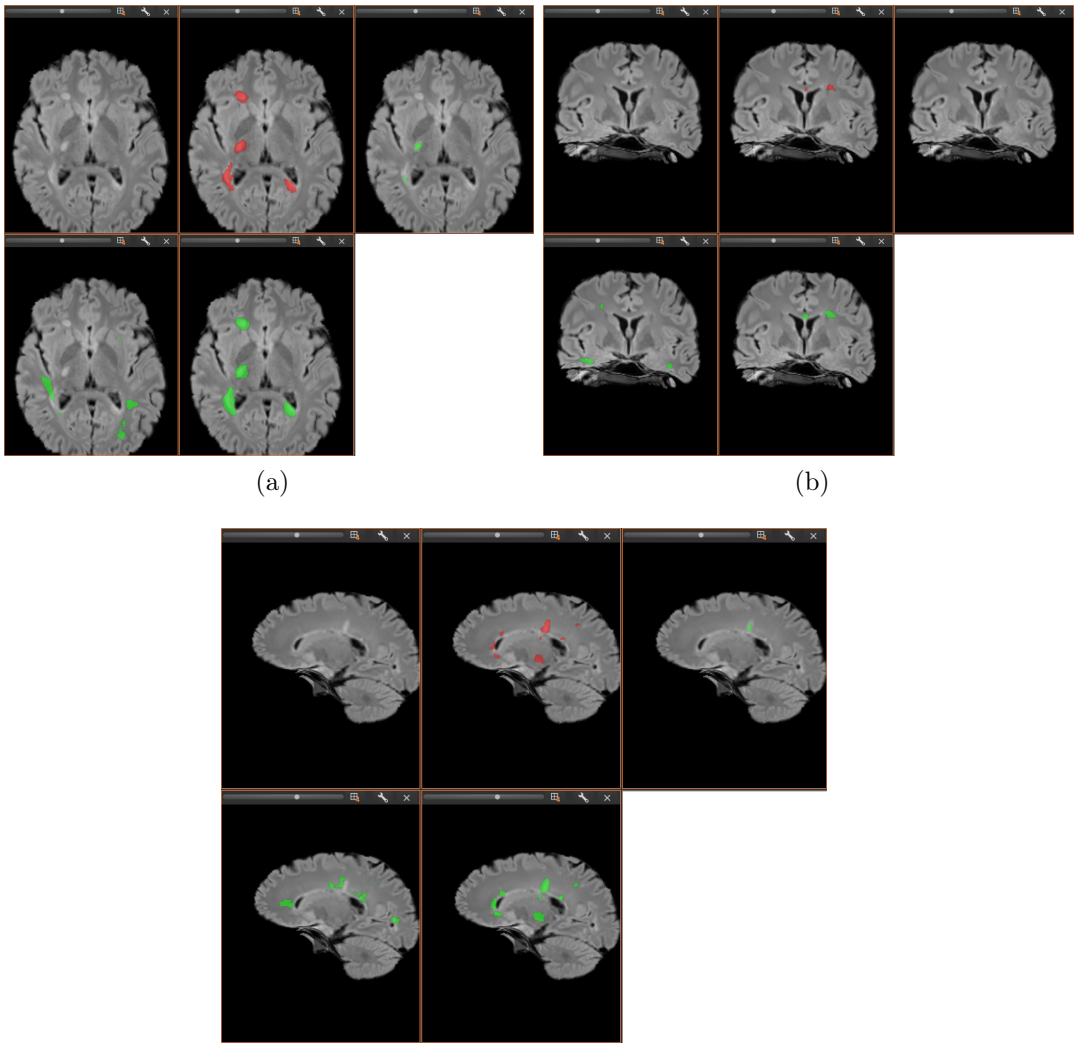


(a) (b)



(c)

Figure 7.8: From dataset 1, a slice of FLAIR with lesion detection in three different orientations : (a) axial; (b) coronal ; (c) sagittal. In each image: (1) top row: a slice of FLAIR, same slice overlayed with ground truth (red), detections (green) obtained by OCSVM; (2) Bottom row: from left to right, detection obtained by MCD and proposed method.



(a) (b)

(c)

Figure 7.9: From dataset 2, a slice of FLAIR with lesion detection in three different orientations : (a) axial; (b) coronal ; (c) sagittal. In each image: (1) top row: a slice of FLAIR, same slice overlayed with ground truth (red), detection (green) obtained by OCSVM; (2) Bottom row: from left to right, detection obtained by MCD and proposed method.

7.7 Conclusion

We have introduced a novel method for MS Lesions detection based on Least Squares Probabilistic Classifier. The efficacy of our method was evaluated though rigorous evaluation on multi-site multi-scanner data from multi-sequence MRI volumes of patients with variable MS lesion load. We have demonstrated that our method achieves better performance compared to benchmark methods: OSVM and MCD. This observation is consistent in both type of data e.g. intensity-normalized and un-normalized. Our results demonstrate that intensity normalization has a instrumental role to play in both enabling efficient NABT learning and classification ability. Our methodology is more suitable for MS lesion analysis because of its ability to capture NABT distribution correctly. This performance suggests that it can provide valuable assistance in detecting the MS lesions in clinical routine with high reliability.

The method has wide applicability as was demonstrated with experiments on two datasets. The proposed framework is generic in nature and can be extended beyond MS Lesions detection e.g. ischemic strokes, tumor. The framework described here allows for exploration of additional MR sequences with or without contrast agents.

Chapter 8

Summary and Perspective

In this thesis, we have presented methods for longitudinal intensity normalization, change detection of lesion activity and MS lesion detection. Image representation and learning is a critical component of any pathology recognition task and has been an active research problem in the medical vision community. The publications which were stemmed out of this thesis so far are as follows:

- Christian Barillot, Yogesh Karpate, Alessandro Crimi and Olivier Commowick. *Analyse d'images spatio-temporelles dans la Sclérose en Plaques*. In Reconnaissance de Formes et Intelligence Artificielle (RFIA) 2014, Rouen, France, June 2014

- Yogesh Karpate, Olivier Commowick, Christian Barillot and Gilles Edan. *Longitudinal Intensity Normalization in Multiple Sclerosis Patients*. In Clinical Image-Based Procedures. Translational Research in Medical Imaging held in conjunction with MICCAI 2014, volume 8680 of *Lecture Notes in Computer Science*, pages 118–125. Springer International Publishing, 2014

- Yogesh Karpate, Olivier Commowick and Christian Barillot. *Robust detection of multiple sclerosis lesions from intensity-normalized multi-channel MRI*. In Proc. SPIE, volume 9413, pages 941314–941314–7, 2015

- Yogesh Karpate, Olivier Commowick and Christian Barillot. *Probabilistic One Class Learning for Automatic Detection of Multiple Sclerosis Lesions*. In IEEE International Symposium on Biomedical Imaging (ISBI), pages 486–489, Brooklyn, United States, April 2015

We now summarize the contributions made in this thesis and then present some directions of possible future work.

8.1 Longitudinal Intensity Normalization

In Chapter 5, we proposed a novel framework for longitudinal intensity normalization to derive a more discriminative image representation for tracking

disease evolution. The proposed method addresses two limitations of the standard techniques i.e. (i) robust yet sensitive estimation of tissue parameters with γ divergence, (ii) disease evolution tracking based on subtraction and thresholding.

We showed that the method performs better than current state of the art algorithms. The method has general applicability to other tasks where the tracking of lesions is important e.g. we demonstrated the method on active lesions with T1-Gd as well.

Future work: We considered a single value of γ-parameter as we expected the scans to adapt to the content and hence adapt automatically to the modes in the distribution of tissue parameters, even if multiple tuning parameters provide different information. However, in more complicated scenarios, we can consider different values of tuning parameters for the same image. The best one or a weighted combination of all of them may then be considered for the final decision.

Another path could be a simple extension to validate the framework on another contrast agent ultra small super-paramagnetic iron oxide (USPIO) which generally exhibits hypo intense lesions on T1 image. This is particularly challenging because CSF and lesions enhanced by USPIO share a similar intensity space. Currently, the proposed intensity normalization method considers only a single sequence at a time. A possible direction could be to extend this representation for multi sequence approach by considering their information simultaneously. This may lead to better intensity normalization.

8.2 Robust Detection of Multiple Sclerosis Lesions from Intensity-Normalized Multi-Channel MRI

It aims at studying the benefits of using multi-channel MRI to detect statistically significant differences between each individual MS patient and a database of control subjects. This framework consists in two components. First, intensity standardization is conducted to minimize the inter-subject intensity difference arising from variability of the acquisition process and different scanners. Based on 5, the intensity normalization method maps parameters obtained using a robust Gaussian Mixture Model (GMM) estimation not affected by the presence of MS lesions. The second part studies the comparison of multi-channel MRI of MS patients with respect to an atlas built from the control subjects, thereby allowing us to look for differences in normal appearing white matter, in and around the lesions of each patient. Experimental results demonstrate that

our technique accurately detects significant differences in lesions consequently improving the results of MS lesion detection.

Future Work: The work presented demonstrated that statistical comparisons based on voxel wise comparison achieve very high performance, however, the overall global structure is still missing. One can consider to infuse statistical comparison of patches using kernel methods, i.e. two sample test based on kernel methods.

8.3 Probabilistic One Class Learner for MS Lesion Detection

In Chapter 7, we proposed a method to learn a bag of words representation from multi-sequence MRI for a given classification task. We demonstrated that, only by learning on data of healthy volunteers by one class learner, we can detect the MS lesions from a patient. This is particularly important because standard binary classifiers suffer from class imbalance problem generating bias towards healthy class under-fitting the minority class i.e MS lesions. We proposed to learn such discriminative distribution for MS lesion detection. The proposed method addresses an important issue of adaptation of the representation pathological tissues based on the specific distribution of discriminative information in healthy tissues. We showed that the method achieves better or comparable results with respect to benchmark methods on our dataset.

Future work: The method uses T1-w-MPRAGE, T2-w and FLAIR-w sequences. One can include more sequences like DTI and MTR. The probability maps obtained from classifier are smoothed using simple Gaussian kernel to establish spatial connectivity. This can be replaced with a more sophisticated approach like Conditional Random Fields. This is an interesting direction to follow. Besides, the evaluation of performance metrics for different patch sizes will yield an insight into classifier behavior.

With existing framework, only raw intensity patches are taken into account. Another possible direction is to extend this framework on data based on features. For example, one can think to include scale and rotational invariant features of images for better detection of MS lesions.

8.4 Conclusion

In the present thesis we have mainly addressed the problem of image representation, with MS lesion detection. Many longitudinal MS investigations are coming up. MR images play crucial role in diagnosis and better therapeutic

management of the disease. In the near future, it will be critical to have good representations and detection frameworks in order to be able to accurately analyze and understand MS lesions using automatic statistical and machine learning based techniques. Analyzing longitudinal MR scans will also become an important technology owing to its numerous important applications e.g. tracking disease evolution. Towards these goals, we have proposed image representations and learning methods which extend the current state-of-the-art either by better exploiting the spatial distribution of discriminative information or by providing a statistical test of highly local pixel neighborhoods. We have demonstrated, with experiments on challenging datasets, that our proposed methods perform better or comparable to the state-of-the-art methods.

We have proposed three different MS lesions detection approaches which can be utilized in different scenarios. First method is useful to track the lesion evolution in longitudinal studies. Though this method detects appearing or vanishing lesions quite effectively, there is no provision to detect static lesions. The second method is based on patient to group comparison methodology, which can detect lesions on voxel level. This method works very well in a controlled environment. The method is highly dependent on data pre-processing pipeline. Another factor which may affect this method is atlas construction. The last one is based on probabilistic one class learner. The prime advantage of this approach is that there is no need to include MS lesion data in training. The approach is flexible to accommodate the different nature of dataset. The parameter tuning of this approach need to be handled carefully. The choice of methodology for MS lesion detection between last two approaches depends on the different use cases. They may be used as complementary to each other. For ex. detections obtained by patient to group comparison method can be used as a prior in probabilistic one class learning framework. Consequently, this will reduce the search space to look for probable lesion locations and will accelerate the method. Patient to group comparison method can be extended to another application. It may be used to detect other abnormal changes in the brain e.g. atrophy.

An efficient MS lesion segmentation can be obtained by detection. The core idea is to generate around few hundred candidate MS lesion segments per image that with high probability coarsely capture most of the lesions in the image. MS lesion detection search must be performed over position, scale and aspect ratio. The resulting four dimensional search space is large and difficult to search over exhaustively. For MSL detection, the most of the state-of-the-art is based on exhaustive search. In order to enable the use of the more expensive features and classifiers, selective search sampling can be used instead of sliding window technique. The candidate segments themselves may be noisy and overlapping and in general need not capture the objects perfectly.

Then such detection can be improved to yield MS lesions segmentation masks. The proposed framework can also be used for localization of MS lesions where detected lesions can be represented by the bounding boxes. The appearance and contextual knowledge of anatomy can be effectively incorporated to obtain a better MS lesion segmentation.

Bibliography

[Abdullah 2011] B Abdullah, Younis PPA and Saraf Lavi E. *Textural based SVM for MS lesion segmentation in FLAIR MRIs.* Open Journal of Medical Imaging, vol. 1, no. 1, pages 26–42, 2011. (Cited on page 29.)

[Achanta 2012] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua and Sabine. *SLIC Superpixels Compared to State-of-the-Art Superpixel Methods.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 11, pages 2274–2282, 2012. (Cited on page 86.)

[AïtAli 2005] L.S. AïtAli, S. Prima, P. Hellier, B. Carsin, G. Edan and C. Barillot. *STREM: A Robust Multidimensional Parametric Method to Segment MS Lesions in MRI.* In JamesS. Duncan and Guido Gerig, editeurs, Medical Image Computing and Computer-Assisted Intervention - MICCAI 2005, volume 3749 of *Lecture Notes in Computer Science*, pages 409–416. Springer Berlin Heidelberg, 2005. (Cited on page 28.)

[Akaike 1974] H. Akaike. *A new look at the statistical model identification.* Automatic Control, IEEE Transactions on, vol. 19, no. 6, pages 716–723, Dec 1974. (Cited on page 47.)

[Akselrod-Ballin 2006] Ayelet Akselrod-Ballin, Meirav Galun, Ronen Basri, Achi Brandt, Moshe John Gomori, Massimo Filippi and Paula Valsasina. *An Integrated Segmentation and Classification Approach Applied to Multiple Sclerosis Analysis.* In CVPR (1), pages 1122–1129. IEEE Computer Society, 2006. (Cited on page 30.)

[Aymerich 2010a] F Aymerich, E Montseny, P Sobrevilla and A Rovira. *A fuzzy localbased approach for detecting cerebrospinal fluid regions in presence of MS lesions.* Proceedings of the 13th International Conference, Part II, IPMU: 2010; Dortmund, 2010. (Cited on page 29.)

[Aymerich 2010b] F Aymerich, P Sobrevilla, E Montseny and A Rovira. *Filtering false detections of small multiple sclerosis lesions using fuzzy regional analysis.* 2010 IEEE International Conference on Fuzzy Systems (FUZZ); July 18-23 2010; Barcelona, pages pp. 1–8, 2010. (Cited on page 29.)

[Bakshi 2008] Rohit Bakshi, Alan J Thompson, Maria A Rocca, Daniel Pelletier, Vincent Dousset, Frederik Barkhof, Matilde Inglese, Charles RG Guttmann, Mark A Horsfield and Massimo Filippi. *MRI in multiple*

sclerosis: current status and future prospects. The Lancet Neurology, vol. 7, no. 7, pages 615–625, jul 2008. (Cited on page 23.)

[Barillot 2014] Christian Barillot, Yogesh Karpate, Alessandro Crimi and Olivier Commowick. *Analyse d'images spatio-temporelles dans la Sclérose en Plaques*. In Reconnaissance de Formes et Intelligence Artificielle (RFIA) 2014, Rouen, France, June 2014. (Not cited.)

[Benjamini 1995] Yoav Benjamini and Yosef Hochberg. *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing*. Journal of the Royal Statistical Society Series B (Methodological), vol. 57, no. 1, pages 289–300, 1995. (Cited on page 65.)

[Biediger 2014] Daniel Biediger, Christophe Collet and Jean-Paul Armspach. *Multiple sclerosis lesion detection with local multimodal Markovian analysis and cellular automata 'GrowCut'*. Journal of Computational Surgery, vol. 1, no. 1, page 3, 2014. (Cited on page 26.)

[Bishop 2006] Christopher M. Bishop. Pattern recognition and machine learning (information science and statistics). Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006. (Cited on page 61.)

[Bosc 2003] Marcel Bosc, Fabrice Heitz, Jean-Paul Armspach, Izzie Namer, Daniel Gounot and Lucien Rumbach. *Automatic change detection in multimodal serial MRI: application to multiple sclerosis lesion evolution*. NeuroImage, vol. 20, no. 2, pages 643 – 656, 2003. (Cited on page 50.)

[Boykov 2006] Yuri Boykov and Gareth Funka-Lea. *Graph Cuts and Efficient N-D Image Segmentation*. Int. J. Comput. Vision, vol. 70, no. 2, pages 109–131, November 2006. (Cited on page 26.)

[Brück 2005] Wolfgang Brück. *The pathology of multiple sclerosis is the result of focal inflammatory demyelination with axonal damage*. Journal of Neurology, vol. 252, no. 5, pages v3–v9, 2005. (Cited on page 14.)

[Cárdenes 2003] Rubén Cárdenes, Simon K. Warfield, Elsa M. Macías, Jose Aurelio Santana and Juan Ruiz-Alzola. *An Efficient Algorithm for Multiple Sclerosis Segmentation from Brain MRI*. In Computer Aided Systems Theory - EUROCAST 2003, 9th International Workshop on Computer Aided Systems Theory, Las Palmas de Gran Canaria, Spain, February 24-28, 2003, Revised Selected Papers, pages 542–551, 2003. (Cited on page 25.)

[Chawla 2005] NiteshV. Chawla. *Data Mining for Imbalanced Datasets: An Overview*. In Oded Maimon and Lior Rokach, editeurs, Data Mining and Knowledge Discovery Handbook, pages 853–867. Springer US, 2005. (Cited on pages 2 and 36.)

[Commowick 2008] Olivier Commowick, Pierre Fillard, Olivier Clatz and Simon K. Warfield. *Detection of DTI White Matter Abnormalities in Multiple Sclerosis Patients*. In Proceedings of the 11th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'08), Part I, volume 5241 of *LNCS*, pages 975–982, September 2008. (Cited on page 64.)

[Commowick 2012a] O. Commowick, N. Wiest-Daessle and S. Prima. *Block-matching strategies for rigid registration of multimodal medical images*. In Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on, pages 700–703, May 2012. (Cited on pages 53 and 66.)

[Commowick 2012b] Olivier Commowick, Nicolas Wiest-DaesslÃľ and Sylvain Prima. *Automated diffeomorphic registration of anatomical structures with rigid parts: Application to dynamic cervical MRI*. In Medical Image Computing and Computer-Assisted Intervention (MICCAI'12), October 2012. (Cited on page 66.)

[Commowick 2015] Olivier Commowick, Adil Maarouf, Jean-Christophe Ferré, Jean-Philippe Ranjeva, Gilles Edan and Christian Barillot. *Diffusion MRI abnormalities detection with orientation distribution functions: A multiple sclerosis longitudinal study*. Medical Image Analysis, vol. 22, no. 1, pages 114–123, May 2015. (Cited on page 64.)

[Compston 2008] Alastair Compston and Alasdair Coles. *Multiple sclerosis*. The Lancet, vol. 372, no. 9648, pages 1502 – 1517, 2008. (Cited on page 15.)

[Cotton 2003] F. Cotton, H. L. Weiner, F. A. Jolesz and C. R. Guttmann. *MRI contrast uptake in new lesions in relapsing-remitting MS followed at weekly intervals*. Neurology, vol. 60, no. 4, pages 640–646, Feb 2003. (Cited on page 21.)

[Coupe 2008] P. Coupe, P. Yger, S. Prima, P. Hellier, C. Kervrann and C. Barillot. *An Optimized Blockwise Nonlocal Means Denoising Filter for 3-D Magnetic Resonance Images*. Medical Imaging, IEEE Transactions on, vol. 27, no. 4, pages 425–441, April 2008. (Cited on pages 53, 66 and 88.)

[Cox 1995] I.J. Cox, S. Roy and S.L. Hingorani. *Dynamic histogram warping of image pairs for constant image brightness.* In Image Processing, 1995. Proceedings., International Conference on, volume 2, pages 366–369 vol.2, Oct 1995. (Cited on page 41.)

[Curati 1996] W. L. Curati, E. J. Williams, A. Oatridge, J. V. Hajnal, N. Saeed and G. M. Bydder. *Use of subvoxel registration and subtraction to improve demonstration of contrast enhancement in MRI of the brain.* Neuroradiology, vol. 38, no. 8, pages 717–723, 1996. (Cited on page 49.)

[Dalal 2005] Navneet Dalal and Bill Triggs. *Histograms of Oriented Gradients for Human Detection.* In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01, CVPR '05, pages 886–893, Washington, DC, USA, 2005. IEEE Computer Society. (Cited on pages 3 and 37.)

[Dempster 1977] A. P. Dempster, N. M. Laird and D. B. Rubin. *Maximum likelihood from incomplete data via the EM algorithm.* JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B, vol. 39, no. 1, pages 1–38, 1977. (Cited on page 46.)

[Derraz 2010] Foued Derraz, Laurent Peyrodie, Antonio Pinti, Azzeddine Chikh and Patrick Hautecoeur. *Semi-automatic Segmentation of Multiple Sclerosis Lesion Based Active Contours Model and Variational Dirichlet Process.* CMES:Computer Modeling in Engineering and Sciences, vol. 67, no. 2, 2010. (Cited on page 25.)

[Deshpande 2014] Hrishikesh Deshpande, Pierre Maurel and Christian Barillot. *Detection of Multiple Sclerosis Lesions using Sparse Representations and Dictionary Learning.* In 2nd International Workshop on Sparsity Techniques in Medical Imaging (STMI), MICCAI 2014, numéro 71-79, Boston, MA, United States, September 2014. (Cited on page 31.)

[Deshpande 2015] Hrishikesh Deshpande, Pierre Maurel and Christian Barillot. *Adaptive Dictionary Learning for Competitive CLassification of Multiple Sclerosis Lesions.* In International Symposium on Biomedical Imaging, ISBI 2015, numéro x-x, Newyork, United States, 2015. (Cited on page 31.)

[DH 2005] Gilden DH. *Infectious causes of multiple sclerosis.* Lancet Neurology, vol. 4, no. 3, pages 195–205, 2005. (Cited on page 14.)

[Dugas-Phocion 2004] Guillaume Dugas-Phocion, Miguel Ángel González Ballester, Grégoire Malandain, Christine Lebrun and Nicholas Ayache. *Improved EM-Based Tissue Segmentation and Partial Volume Effect Quantification in Multi-sequence Brain MRI.* In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2004, 7th International Conference Saint-Malo, France, September 26-29, 2004, Proceedings, Part I, pages 26–33, 2004. (Cited on pages 2, 28 and 36.)

[Dyment 2004] David A Dyment, George C Ebers and A Dessa Sadovnick. *Genetics of multiple sclerosis.* The Lancet Neurology, vol. 3, no. 2, pages 104 – 110, 2004. (Cited on page 15.)

[Ebers 2008] George Ebers. *Environmental factors and multiple sclerosis.* The Lancet Neurology, vol. 7, no. 3, pages 268 – 277, 2008. (Cited on page 14.)

[Eguchi 2010] Shinto Eguchi and Shogo Kato. *Entropy and Divergence Associated with Power Function and the Statistical Application.* Entropy, vol. 12, no. 2, pages 262–274, 2010. (Cited on page 44.)

[Felzenszwalb 2010] Pedro F. Felzenszwalb, Ross B. Girshick, David A. McAllester and Deva Ramanan. *Object Detection with Discriminatively Trained Part-Based Models.* IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 9, pages 1627–1645, 2010. (Cited on pages 3 and 37.)

[Filippi 2001] M Filippi. *Diffusion Tensor Magnetic Resonance Imaging in Multiple Sclerosis.* Neurology, vol. 56, no. 3, pages 304–3011, 2001. (Cited on pages 19 and 64.)

[Filippi 2007] Massimo Filippi and Federica Agosta. *Magnetization Transfer MRI in Multiple Sclerosis.* Journal of Neuroimaging, vol. 17, pages 22S–26S, 2007. (Cited on page 19.)

[Fiot 2008] JB Fiot, L Cohen, Raniga P and Fripp. *Efficient lesion segmentation using support vector machines.* volume 1, pages 26–42, 2008. (Cited on page 29.)

[Fox 2011] Robert J. Fox, Erik Beall, Pallab Bhattacharyya, Jacqueline T. Chen and Ken Sakaie. *Advanced {MRI} in Multiple Sclerosis: Current Status and Future Challenges.* Neurologic Clinics, vol. 29, no. 2, pages 357 – 380, 2011. Multiple Sclerosis. (Cited on page 19.)

[Freifeld 2009] Oren Freifeld, Hayit Greenspan and Jacob Goldberger. *Multiple Sclerosis Lesion Detection Using Constrained GMM and Curve*

Evolution. Journal of Biomedical Imaging, vol. 2009, pages 14:1–14:13, January 2009. (Cited on page 28.)

[Fromont 2010] Agnes Fromont, Christine Binquet, Erik A. Sauleau, Isabelle Fournel, Audrey Bellisario, Johan Adnet, Alain Weill, Sandra Vukusic, Christian Confavreux, Marc Debouverie, Laurence Clerc, Claire Bonithon-Kopp and Thibault Moreau. *Geographic variations of multiple sclerosis in France.* Brain, vol. 133, no. 7, pages 1889–1899, 2010. (Cited on page 15.)

[Garcia-Lorenzo 2011] Daniel Garcia-Lorenzo, Sylvain Prima, Douglas L. Arnold, D. Louis Collins and Christian Barillot. *Trimmed-Likelihood Estimation for Focal Lesions and Tissue Segmentation in Multisequence MRI for Multiple Sclerosis.* IEEE Trans. Med. Imaging, vol. 30, no. 8, pages 1455–1467, 2011. (Cited on pages 27 and 28.)

[Garcia-Lorenzo 2013] Daniel Garcia-Lorenzo, Simon Francis, Sridar Narayanan, Douglas L. Arnold and D. Louis Collins. *Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging.* Medical Image Analysis, vol. 17, no. 1, pages 1 – 18, 2013. (Cited on pages 22 and 26.)

[Ge 2006] Y. Ge. *Multiple Sclerosis: The Role of MR Imaging.* American Journal of Neuroradiology, vol. 27, no. 6, pages 1165–1176, 2006. (Cited on page 18.)

[Geremia 2010] Ezequiel Geremia, Bjoern H. Menze, Olivier Clatz, Ender Konukoglu, Antonio Criminisi and Nicholas Ayache. *Spatial Decision Forests for MS Lesion Segmentation in Multi-channel MR Images.* In Proceedings of the 13th International Conference on Medical Image Computing and Computer-assisted Intervention: Part I, MICCAI'10, pages 111–118, Berlin, Heidelberg, 2010. Springer-Verlag. (Cited on page 30.)

[Guimond 2000] A. Guimond, J. Meunier and J.-P. Thirion. *Average Brain Models: A Convergence Study.* Computer Vision and Image Understanding, vol. 77, no. 2, pages 192–210, 2000. (Cited on page 66.)

[Harmouche 2006] Rola Harmouche, D. Louis Collins, Douglas L. Arnold, Simon J. Francis and Tal Arbel. *Bayesian MS Lesion Classification Modeling Regional and Local Spatial Information.* In ICPR (3), pages 984–987. IEEE Computer Society, 2006. (Cited on page 28.)

[He 2002] Renjie He and Ponnada A. Narayana. *Automatic delineation of Gd enhancements on magnetic resonance images in multiple sclerosis.* Medical Physics, vol. 29, no. 7, pages 1536–1546, 2002. (Cited on page 31.)

[Hellier 2003] P. Hellier. *Consistent intensity correction of MR images.* In Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on, volume 1, pages I–1109–12 vol.1, Sept 2003. (Cited on pages 6, 42 and 43.)

[Horsfield 2007] Mark A. Horsfield, Rohit Bakshi, Marco Rovaris, Mara A. Rocca, Venkata S. R. Dandamudi, Paola Valsasina, Elda Judica, Fulvio Lucchini, Charles R. G. Guttmann, Maria P. Sormani and Massimo Filippi. *Incorporating Domain Knowledge Into the Fuzzy Connectedness Framework: Application to Brain Lesion Volume Estimation in Multiple Sclerosis.* IEEE Transactions on Medical Imaging, vol. 26, no. 12, pages 1670–1680, December 2007. (Cited on page 25.)

[Jäger 2006] Florian Jäger, Yu Deuerling-Zheng, Bernd Frericks, Frank Wacker and Joachim Hornegger. *A new Method for MRI Intensity Standardization with Application to Lesion Detection in the Brain.* In L. Kobbelt, T. Kuhlen, T. Aach and R. Westermann, editeurs, Vision Modeling and Visualization 2006, pages 296–276, Berlin, 2006. (Cited on page 41.)

[Jäger 2007] Florian Jäger, Laszlo Nyul, Bernd Frericks, Frank Wacker and Joachim Hornegger. *Whole Body MRI Intensity Standardization.* In Alexander Horsch, ThomasM. Deserno, Heinz Handels, Hans-Peter Meinzer and Thomas Tolxdorff, editeurs, Bildverarbeitung för die Medizin 2007, Informatik aktuell, pages 459–463. Springer Berlin Heidelberg, 2007. (Cited on page 41.)

[Kadoury 2012] Kadoury. *Manifold-constrained embeddings for the detection of white matter lesions in brain MRI.* In 9th IEEE,ISBI, pages 562–565, 2012. (Cited on page 32.)

[Karimaghaloo 2012] Zahra Karimaghaloo, Douglas L. Arnold, D. Louis Collins and Tal Arbel. *Hierarchical Conditional Random Fields for Detection of Gad-enhancing Lesions in Multiple Sclerosis.* In Proceedings of the 15th International Conference on Medical Image Computing and Computer-Assisted Intervention - Volume Part II, MICCAI'12, pages 379–386, Berlin, Heidelberg, 2012. Springer-Verlag. (Cited on page 31.)

[Karimaghaloo 2013] Zahra Karimaghaloo, Hassan Rivaz, Douglas L. Arnold, D. Louis Collins and Tal Arbel. *Adaptive Voxel, Texture and Temporal*

Conditional Random Fields for Detection of Gad-Enhancing Multiple Sclerosis Lesions in Brain MRI. In Medical Image Computing and Computer-Assisted Intervention - MICCAI 2013 - 16th International Conference, Nagoya, Japan, September 22-26, 2013, Proceedings, Part III, pages 543–550, 2013. (Cited on page 31.)

[Karpate 2014] Yogesh Karpate, Olivier Commowick, Christian Barillot and Gilles Edan. *Longitudinal Intensity Normalization in Multiple Sclerosis Patients.* In Clinical Image-Based Procedures. Translational Research in Medical Imaging held in conjunction with MICCAI 2014, volume 8680 of *Lecture Notes in Computer Science*, pages 118–125. Springer International Publishing, 2014. (Not cited.)

[Karpate 2015a] Yogesh Karpate, Olivier Commowick and Christian Barillot. *Probabilistic One Class Learning for Automatic Detection of Multiple Sclerosis Lesions.* In IEEE International Symposium on Biomedical Imaging (ISBI), pages 486–489, Brooklyn, United States, April 2015. (Not cited.)

[Karpate 2015b] Yogesh Karpate, Olivier Commowick and Christian Barillot. *Robust detection of multiple sclerosis lesions from intensity-normalized multi-channel MRI.* In Proc. SPIE, volume 9413, pages 941314–941314–7, 2015. (Not cited.)

[Kayhan 2008] Kayhan. *Multiparametric tissue abnormality characterization using manifold regularization.* In SPIE Medical Imaging,, 2008. (Cited on page 32.)

[Khayati 2008] R. Khayati, M. Vafadust, F. Towhidkhah and S. Nabavi. *A novel method for automatic determination of different stages of multiple sclerosis lesions in brain MR FLAIR images.* Computerized Medical Imaging and Graphics, vol. 32, no. 2, pages 124–133, March 2008. (Cited on pages 27 and 29.)

[Lassmann 2008] H. Lassmann. *Models of multiple sclerosis: new insights into pathophysiology and repair.* Curr. Opin. Neurol., vol. 21, no. 3, pages 242–247, Jun 2008. (Cited on page 21.)

[Lecoeur 2009] J. Lecoeur, J.-C. Ferre, D. L. Collins, S. Morrissey and C. Barillot. *Multi Channel MRI Segmentation with Graph Cuts using Spectral Gradient and Multidimensional Mixture Mode.* In Proceedings of SPIE Medical Imaging 2009, volume 7259, page 72593X, February 2009. (Cited on page 26.)

[Leemput 2001] Koen Van Leemput, Frederik Maes, Dirk Vandermeulen, Alan C. F. Colchester and Paul Suetens. *Automated Segmentation of Multiple Sclerosis Lesions by Model Outlier Detection.* IEEE Trans. Med. Imaging, vol. 20, no. 8, pages 677–688, 2001. (Cited on page 28.)

[Levin 2005] Lynn Levin, Kassandra Munger, Mark Rubertone, Charles Peck, Evelyne Lennette and Dona Spiegleman. *Temporal relationship between elevation of Epstein-Barr virus antibody titers and initial onset of neurological symptoms in multiple sclerosis.* Journal of the American Medical Association, vol. 293, pages 2496–2500, 2005. (Cited on page 15.)

[Lincoln 2005] Mathew Lincoln, Alexandre Montpetit, M Zameel Cader, Janna Saarela, David A Dyment, Milvi Tiisla, Vincent Ferretti, Pentti J Tienari, A Dessa Sadovnick, Leena Peltonen, George C Ebers and Thomas J Hudson. *A predominant role for the HLA class II region in the association of the MHC region with multiple sclerosis.* Nature Genetics, vol. 37, no. 2, pages 1108 – 1112, 2005. (Cited on page 15.)

[Lladó 2012a] Xavier Lladó, Onur Ganiler, Arnau Oliver, Robert MartÃŋ, Jordi Freixenet, Laia Valls, JoanC. Vilanova, Lluís Ramió-Torrentí and ílex Rovira. *Automated detection of multiple sclerosis lesions in serial brain MRI.* Neuroradiology, vol. 54, no. 8, pages 787–807, 2012. (Cited on pages 48 and 49.)

[Lladó 2012b] Xavier Lladó, Arnau Oliver, Mariano Cabezas, Jordi Freixenet, Joan C. Vilanova, Ana Quiles, Laia Valls, Lluís Ramió-Torrentí and ílex Rovira. *Segmentation of Multiple Sclerosis Lesions in Brain MRI: A Review of Automated Approaches.* Inf. Sci., vol. 186, no. 1, pages 164–185, March 2012. (Cited on page 26.)

[Lublin 1996] Lublin and SC Reingold. *Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis.* Neurology, vol. 46, no. 4, pages 907–11, 1996. (Cited on page 16.)

[MacQueen 1967] J. B. MacQueen. *Some Methods for Classification and Analysis of MultiVariate Observations.* In L. M. Le Cam and J. Neyman, editeurs, Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability, volume 1, pages 281–297. University of California Press, 1967. (Cited on page 54.)

[McDonald 2001] W. Ian McDonald, Alistair Compston, Gilles Edan, Donald Goodkin, Hans-Peter Hartung, Fred D. Lublin, Henry F. McFarland,

Donald W. Paty, Chris H. Polman, Stephen C. Reingold, Magnhild Sandberg-Wollheim, William Sibley, Alan Thompson, Stanley Van Den Noort, Brian Y. Weinshenker and Jerry S. Wolinsky. *Recommended diagnostic criteria for multiple sclerosis: Guidelines from the international panel on the diagnosis of multiple sclerosis.* Annals of Neurology, vol. 50, no. 1, pages 121–127, 2001. (Cited on page 19.)

[McLachlan 2008] Geoffrey J. McLachlan and Thriyambakam Krishnan. The em algorithm and extensions. Wiley series in probability and statistics. Wiley, Hoboken, NJ, 2. ed édition, 2008. (Cited on page 28.)

[Meier 2003] Dominik S. Meier and Charles R.G. Guttmann. *Time-series analysis of {MRI} intensity patterns in multiple sclerosis.* NeuroImage, vol. 20, no. 2, pages 1193 – 1209, 2003. (Cited on page 51.)

[Meier 2007a] DominikS. Meier, HowardL. Weiner and CharlesR.G. Guttmann. *Time-series modeling of multiple sclerosis disease activity: A promising window on disease progression and repair potential?* Neurotherapeutics, vol. 4, no. 3, pages 485–498, 2007. (Cited on page 21.)

[Meier 2007b] D.S. Meier, H.L. Weiner and C.R.G. Guttmann. *MR Imaging Intensity Modeling of Damage and Repair In Multiple Sclerosis: Relationship of Short-Term Lesion Recovery to Progression and Disability.* American Journal of Neuroradiology, vol. 28, no. 10, pages 1956–1963, 2007. (Cited on page 21.)

[Moraal 2009] Bastiaan Moraal, Dominik S. Meier, Peter A. Poppe, Jeroen J. G. Geurts, Hugo Vrenken, William M. A. Jonker, Dirk L. Knol, Ronald A. van Schijndel, Petra J. W. Pouwels, Christoph Pohl, Lars Bauer, Rupert Sandbrink, Charles R. G. Guttmann and Frederik Barkhof. *Subtraction MR Images in a Multiple Sclerosis Multicenter Clinical Trial Setting.* Radiology, vol. 250, no. 2, pages 506–514, 2009. PMID: 19037018. (Cited on page 50.)

[Moraal 2010] Bastiaan Moraal, Mike P. Wattjes, Jeroen J. G. Geurts, Dirk L. Knol, Ronald A. van Schijndel, Petra J. W. Pouwels, Hugo Vrenken and Frederik Barkhof. *Improved Detection of Active Multiple Sclerosis Lesions: 3D Subtraction Imaging.* Radiology, vol. 255, no. 1, pages 154–163, 2010. PMID: 20308453. (Cited on page 50.)

[Mumford 1992] C.J. Mumford, M.B. Fraser, N.W. Wood and D.A. Compston. *Multiple sclerosis in the Cambridge health district of East Anglia.* Journal of Neurology Neurosurgery and Psychiatry, vol. 55, pages 877 – 882, 1992. (Cited on page 15.)

[Neykov 2007] N. M. Neykov, Peter Filzmoser, R. Dimova and P. N. Neytchev. *Robust fitting of mixtures using the trimmed likelihood estimator.* Computational Statistics and Data Analysis, vol. 52, no. 1, pages 299–308, 2007. (Cited on pages 28 and 61.)

[Ng 2004] Hui-Fuang Ng. *Automatic thresholding for defect detection.* In Third International Conference on Image and Graphics, ICIG 2004, Hong Kong, China, December 18-20, 2004, pages 532–535. IEEE Computer Society, 2004. (Cited on page 84.)

[Notsu 2014] Akifumi Notsu, Osamu Komori and Shinto Eguchi. *Spontaneous Clustering via Minimum Gamma-divergence.* Neural Comput., vol. 26, no. 2, pages 421–448, February 2014. (Cited on pages 46 and 47.)

[Nyul 2000] L.G. Nyul, J.K. Udupa and Xuan Zhang. *New variants of a method of MRI scale standardization.* Medical Imaging, IEEE Transactions on, vol. 19, no. 2, pages 143–150, Feb 2000. (Cited on pages 6 and 40.)

[Otsu 1979] N. Otsu. *A Threshold Selection Method from Gray-Level Histograms.* IEEE Trans. on Systems, Man and Cybernetics, vol. SMC-9, no. 1, pages 62–66, 1979. (Cited on pages 53 and 84.)

[Ourselin 2000] S. Ourselin, A. Roche, S. Prima and N. Ayache. *Block Matching: A General Framework to Improve Robustness of Rigid Registration of Medical Images.* In ScottL. Delp, AnthonyM. DiGoia and Branislav Jaramaz, editeurs, Medical Image Computing and Computer-Assisted Intervention âĂŞ MICCAI 2000, volume 1935 of *Lecture Notes in Computer Science*, pages 557–566. Springer Berlin Heidelberg, 2000. (Cited on pages 53 and 66.)

[Pekmezovic 2006] T Pekmezovic and M Milenkovic M. *Lifestyle Factors and Multiple Sclerosis: A Case-Control Study in Belgrade.* Neuroepidemiology, vol. 27, pages 212 – 216, 2006. (Cited on page 15.)

[Pietrangelo 2015] Ann Pietrangelo and Valencia Higuera. *Multiple Sclerosis Global Prevalence*, 2015. (Cited on page 16.)

[Polman 2005] Chris H. Polman, Stephen C. Reingold, Gilles Edan, Massimo Filippi, Hans-Peter Hartung, Ludwig Kappos, Fred D. Lublin, Luanne M. Metz, Henry F. McFarland, Paul W. O'Connor, Magnhild Sandberg-Wollheim, Alan J. Thompson, Brian G. Weinshenker and Jerry S. Wolinsky. *Diagnostic criteria for multiple sclerosis: 2005 revisions to the âĂIJMcDonald CriteriaâĂİ.* Annals of Neurology, vol. 58, no. 6, pages 840–846, 2005. (Cited on page 19.)

[Polman 2011] Chris H. Polman, Stephen C. Reingold, Brenda Banwell, Michel Clanet, Jeffrey A. Cohen, Massimo Filippi, Kazuo Fujihara, Eva Havrdova, Michael Hutchinson, Ludwig Kappos, Fred D. Lublin, Xavier Montalban, Paul O'Connor, Magnhild Sandberg-Wollheim, Alan J. Thompson, Emmanuelle Waubant, Brian Weinshenker and Jerry S. Wolinsky. *Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria.* Annals of Neurology, vol. 69, no. 2, pages 292–302, 2011. (Cited on page 19.)

[Quinn 2014] John A. Quinn and Masashi Sugiyama. *A Least-squares Approach to Anomaly Detection in Static and Sequential Data.* Pattern Recogn. Lett., vol. 40, pages 36–40, April 2014. (Cited on page 83.)

[Reddy 2012] K.K. Reddy, Berkan Solmaz, Pingkun Yan, N.G. Avgeropoulos, D.J. Rippe and M. Shah. *Confidence guided enhancing brain tumor segmentation in multi-parametric MRI.* In Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on, pages 366–369, May 2012. (Cited on page 84.)

[Robitaille 2012] Nicolas Robitaille, Abderazzak Mouiha, Burt CrÃĽpeault, Fernando Valdivia and Simon Duchesne. *Tissue-Based MRI Intensity Standardization: Application to Multicentric Datasets.* Int. J. Biomedical Imaging, vol. 2012, 2012. (Cited on page 42.)

[Rousseau 2013] François Rousseau, Estanislao Oubel, Julien Pontabry, Marc Schweitzer, Colin Studholme, Mériam Koob and Jean-Louis Dietemann. *BTK: An open-source toolkit for fetal brain MR image processing.* Computer Methods and Programs in Biomedicine, vol. 109, no. 1, pages 65–73, jan 2013. (Cited on page 23.)

[Rousseeuw 1999] Peter J. Rousseeuw and Katrien Van Driessen. *A Fast Algorithm for the Minimum Covariance Determinant Estimator.* Technometrics, vol. 41, no. 3, pages 212–223, August 1999. (Cited on page 81.)

[Rovira 2013] Alex Rovira, Cristina Auger and Juli Alonso. *Magnetic Resonance Monitoring of Lesion Evolution in Multiple Sclerosis.* Ther Adv Neurol Disorders, vol. 6, no. 5, pages 298 – 310, 2013. (Cited on page 22.)

[Roy 2013a] Snehashis Roy, Aaron Carass and Jerry L. Prince. *Patch based intensity normalization of brain MR images.* In ISBI, pages 342–345, 2013. (Cited on page 43.)

[Roy 2013b] Snehashis Roy, Aaron Carass, Navid Shiee, Dzung L. Pham, Peter A. Calabresi, Daniel S. Reich and Jerry L. Prince. *Longitudinal intensity normalization in the presence of multiple sclerosis lesions*. In ISBI, pages 1384–1387, 2013. (Cited on page 43.)

[Roy 2013c] Snehashis Roy, Amod Jog, Aaron Carass and Jerry L. Prince. *Atlas Based Intensity Transformation of Brain MR Images*. In Li Shen, Tianming Liu, Pew-Thian Yap, Heng Huang, Dinggang Shen and Carl-Fredrik Westin, editeurs, MBIA, volume 8159 of *Lecture Notes in Computer Science*, pages 51–62. Springer, 2013. (Cited on page 43.)

[Sadovnick 1993] Sadovnick and Ebers GC. *Epidemiology of multiple sclerosis: a critical overview*. Canadian Journal of Neuroscience, vol. 20, pages 17–29, 1993. (Cited on page 13.)

[Schmidt 2012] Paul Schmidt, Christian Gaser, Milan Arsic, Dorothea Buck, Annette FÃűrschler, Achim Berthele, Muna Hoshi, RÃijdiger Ilg, Volker J. Schmid, Claus Zimmer, Bernhard Hemmer and Mark MÃijhlau. *An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis*. NeuroImage, vol. 59, no. 4, pages 3774 – 3783, 2012. (Cited on page 29.)

[Schölkopf 2001] Bernhard Schölkopf, John C. Platt, John Shawe-Taylor, Alex J. Smola and Robert C. Williamson. *Estimating the Support of a High-Dimensional Distribution*. Neural Computation, vol. 13, no. 7, pages 1443–1471, 2001. (Cited on pages 81 and 89.)

[Shawe-Taylor 2004] John Shawe-Taylor and Nello Cristianini. Kernel methods for pattern analysis. Cambridge University Press, New York, NY, USA, 2004. (Cited on pages 30 and 80.)

[Shiee 2010] Navid Shiee, Pierre-Louis Bazin, Arzu Ozturk, Daniel S. Reich, Peter A. Calabresi and Dzung L. Pham. *A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions*. NeuroImage, vol. 49, no. 2, pages 1524–1535, 2010. (Cited on page 43.)

[Smith 2002] S.M. Smith. *Fast robust automated brain extraction*. Human Brain Mapping, vol. 17, no. 3, pages 143–155, 2002. (Cited on pages 53 and 66.)

[Souplet 2008] J. Souplet, C. Lebrun, N. Ayache and G. Malandain. *An Automatic Segmentation of T2-FLAIR Multiple Sclerosis Lesions*. 07 2008. (Cited on page 28.)

[Styner 2008] M. Styner, J. Lee, B. Chin, M.S. Chin, O. Commowick, H. Tran, S. Markovic-Plese, V. Jewells and S.K. Warfield. *3D Segmentation in the Clinic: A Grand Challenge II: MS Lesion Segmentation*. In MIDAS Journal, Special Issue on 2008 MICCAI Workshop - MS Lesion Segmentation, pages 1–5, September 2008. (Cited on page 35.)

[Sugiyama 2010] Masashi Sugiyama. *Superfast-Trainable Multi-Class Probabilistic Classifier by Least-Squares Posterior Fitting*. IEICE Transactions, vol. 93-D, no. 10, pages 2690–2701, 2010. (Cited on page 83.)

[Tan 2002] I. Leng Tan, Ronald A. van Schijndel, Franz Fazekas, Massimo Filippi, Peter Freitag, David H. Miller, Tarek A. Yousry, Petra J. W. Pouwels, Herman J. AdÃĺr and Frederik Barkhof. *Image registration and subtraction to detect active T2 lesions in MS: an interobserver study*. Journal of Neurology, vol. 249, no. 6, pages 767–773, 2002. (Cited on page 50.)

[Tax 2004] DavidM.J. Tax and RobertP.W. Duin. *Support Vector Data Description*. Machine Learning, vol. 54, no. 1, pages 45–66, 2004. (Cited on pages 3 and 36.)

[Thirion 1999] J.-P. Thirion and G. Calmon. *Deformation analysis to detect and quantify active lesions in three-dimensional medical image sequences*. Medical Imaging, IEEE Transactions on, vol. 18, no. 5, pages 429–441, May 1999. (Cited on page 50.)

[Tustison 2010] N.J. Tustison, B.B. Avants, P.A. Cook, Yuanjie Zheng, A. Egan, P.A. Yushkevich and J.C. Gee. *N4ITK: Improved N3 Bias Correction*. Medical Imaging, IEEE Transactions on, vol. 29, no. 6, pages 1310–1320, June 2010. (Cited on pages 53, 66 and 88.)

[Uijlings 2013] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers and A. W. M. Smeulders. *Selective Search for Object Recognition*. International Journal of Computer Vision, vol. 104, no. 2, pages 154–171, 2013. (Cited on pages 3 and 37.)

[van der Walt 2014] Stéfan van der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, Tony Yu and the scikit-image contributors. *scikit-image: image processing in Python*. PeerJ, vol. 2, page e453, 6 2014. (Cited on page 88.)

[Vapnik 1995] Vladimir N. Vapnik. The nature of statistical learning theory. Springer-Verlag New York, Inc., New York, NY, USA, 1995. (Cited on pages 29 and 79.)

[Vukusic 2003] Sandra Vukusic and Christian Confavreux. *Prognostic factors for progression of disability in the secondary progressive phase of multiple sclerosis.* Journal of the Neurological Sciences, vol. 206, no. 2, pages 135 – 137, 2003. The Progressive Phase of Multiple Sclerosis: Pathology and Treatment. 7th European Charcot foundation Symposium. (Cited on page 16.)

[Wang 1998] Liqun Wang, H-Ming Lai, Gareth J. Barker, David H. Miller and Paul S. Tofts. *Correction for variations in MRI scanner sensitivity in brain studies with histogram matching.* Magnetic Resonance in Medicine, vol. 39, no. 2, pages 322–327, 1998. (Cited on page 42.)

[Weiner 2004] Howard Weiner. *Multiple Sclerosis Is an Inflammatory T-cell Mediated Autoimmune Disease.* Archives of Neurology, vol. 61, pages 1613–1615, 2004. (Cited on page 14.)

[Weisenfeld 2004] N.I. Weisenfeld and S.K. Warfteld. *Normalization of joint image-intensity statistics in MRI using the Kullback-Leibler divergence.* In Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on, pages 101–104 Vol. 1, April 2004. (Cited on page 42.)

[Weiss 2013] Nick Weiss, Daniel Rueckert and Anil Rao. *Multiple Sclerosis Lesion Segmentation Using Dictionary Learning and Sparse Coding.* In Kensaku Mori, Ichiro Sakuma, Yoshinobu Sato, Christian Barillot and Nassir Navab, editeurs, Medical Image Computing and Computer-Assisted Intervention âĂŞ MICCAI 2013, volume 8149 of *Lecture Notes in Computer Science*, pages 735–742. Springer Berlin Heidelberg, 2013. (Cited on page 31.)

[Wels 2008] M. Wels, M. Huber and J. Hornegger. *Fully automated segmentation of multiple sclerosis lesions in multispectral MRI.* Pattern Recognition and Image Analysis, vol. 18, no. 2, pages 347–350, June 2008. (Cited on page 30.)

Résumé: La sclérose en plaques (SEP) est une maladie auto-immune inflammatoire du jeune adulte causant des handicaps variables et progressifs irréversibles. Cette maladie est présente de manière prépondérante dans l'hémisphère nord. Cette thèse s'attache à la caractérisation et à la modélisation de signatures IRM multimodales des lésions de sclérose en plaques. L'objectif est d'améliorer les modèles de représentation de l'image et d'adapter les méthodes d'apprentissage pour la reconnaissance visuelle, dans le cas où des informations de haut niveau telles que les lésions SEP incluses dans l'IRM sont extraites.

Nous proposons dans cette thèse un nouvel algorithme de normalisation d'intensité en IRM, particulièrement centré sur la normalisation d'images longitudinales multimodales, afin de produire des détections d'évolution de lésion robustes. Cette normalisation est centrée sur la modélisation de l'histogramme de l'image par un modèle de mixture de Gaussiennes robuste à la présence de lésions. Faisant suite à cet algorithme, nous proposons également deux nouvelles méthodes de détection de lésions SEP basées sur (1) une comparaison statistique du patient vis à vis d'une population de sujets contrôle et (2) un cadre probabiliste de détection basé sur un apprentissage d'une classe (tissus sains). Nous avons évalué les algorithmes proposés sur plusieurs jeux de données multi-centriques et vérifié leur efficacité dans la détection de lésions.
Mot clés: IRM, sclérose en plaques, statistiques, apprentissage.

Abstract: Multiple Sclerosis (MS) is an acquired inflammatory disease, which causes disabilities in young adults and it is common in northern hemisphere. This PhD work focuses on characterization and modeling of multidimensional MRI signatures in MS Lesions (MSL). The objective is to improve image representation and learning for visual recognition, where high level information such as MSL contained in MRI are automatically extracted.

We propose a new longitudinal intensity normalization algorithm for multichannel MRI in the presence of MSL, which provides consistent and reliable longitudinal detections. This is based on learning the tissue intensities from multichannel MRI using robust Gaussian Mixture Modeling. Further, we proposed two MSL detection methods based on a statistical patient to population comparison framework and probabilistic one class learning. We evaluated our proposed algorithms on multi-center databases to verify its efficacy.
Key Words: MRI, Multiple Sclerosis, Statistics, Machine Learning.