Universidade de São Paulo
FFCLRP - Departamento de Física
Pós-graduação em Física aplicada à Medicina e Biologia

MEHRAN AZIMBAGIRAD

Segmentação por entropia de Tsallis através de MRF para o Parcelamento de Ressonância Magnética Cerebral

Tese apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo, como parte das exigências para a obtenção do título de Doutor em Ciências, Área: Física aplicada à Medicina e Biologia.

Ribeirão Preto - SP
2019

MEHRAN AZIMBAGIRAD

Tsallis-Entropy Segmentation through MRF for Brain Magnetic Resonance Parcellation

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Faculty of Philosophy, Sciences and literature of Ribeirão Preto, University of São Paulo.

Area:

Applied Physics in Medicine and Biology.

Supervisor:

Luiz Otavio Murta Junior.

Ribeirão Preto - SP
2019

FICHA CATALOGRÁFICA

Azimbagirad, Mehran

Segmentação por entropia de Tsallis através de MRF para o Parcelamento de Ressonância Magnética Cerebral / Mehran Azimbagirad; orientador Luiz Otavio Murta Junior. Ribeirão Preto - SP, 2019.

49 f.:il.

Tese (Doutorado - Programa de Pós-Graduação em Física Aplicada à Medicina e Biologia) - Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo, 2019.

1.Segmentação de Imagem 2.Expectativa-Maximização 3.Algoritmo K-mean 4.q-entropia 5.Markov Random Field 6.Ressonância magnética

AZIMBAGIRAD, M. **Tsallis-Entropy Segmentation through MRF for Brain Magnetic Resonance Parcellation**. 2019. 49 f. Thesis (Ph.D. - Postgraduate Program in Physics Applied to Medicin and Biology) - Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto - SP, 2019.

ERRATUM

Page	Line	Change	To

Nome: Azimbagirad, Mehran

Título: Segmentação por entropia de Tsallis através de MRF para o Parcelamento de Ressonância Magnética Cerebral

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Faculty of Philosophy, Sciences and literature of Ribeirão Preto, University of São Paulo.

Approved in:____/____/____.

Examination Board

Prof. Dr. : ______________________ Institution:____________

Judgement:______________________ Signature: ____________

Prof. Dr. : ______________________ Institution:____________

Judgement:______________________ Signature: ____________

Prof. Dr. : ______________________ Institution:____________

Judgement:______________________ Signature: ____________

Prof. Dr. : ______________________ Institution:____________

Judgement:______________________ Signature: ____________

Prof. Dr. : ______________________ Institution:____________

Judgement:______________________ Signature: ____________

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor Prof. Luiz Otavio Murta Jr for the continuous support of my Ph.D study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study (muitíssimo obrigado).

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Carlos Ernesto Garrido Salmon, Prof. Letícia Rittner, Prof. Renata Ferranti Leoni, and Prof. Dr. Antonio Carlos dos Santos, for their insightful comments and encouragement, but also for the hard questions which incented me to widen my research from various perspectives.

My sincere thanks also goes to Prof. Théo Zeferino Pavan , Prof. Antonio Adilton Oliveira Carneiro, and Prof. Antonio Carlos dos Santos, who provided me an opportunity to join their team, and who gave access to the laboratory and research facilities. Without they precious support it would not be possible to conduct this research.

I thank my fellow labmates in for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last four years. Also I thank my friends in CSIM lab. In particular, I am grateful to Dr. Antonio Carlos da Silva Senra Filho, Dr. Fabrício Henrique Simozo and Dr. Gustavo Canavaci Barizon for enlightening me the first glance of research.

Last but not the least, I would like to thank my family: my parents and to my brothers and sister for supporting me spiritually throughout writing this thesis and all of my friends, specially, Larissa, Leonardo, Yasr, Jessica, Juliano, Felipe, Guilherme, Hadi, Morteza, Kourosh, Ali, Yousof, Vivian, Nilza.

“ ‘Tentar’ não existe.

Faça.

Ou não faça.”

Mestre Yoda

Resumo

AZIMBAGIRAD, M. **Segmentação por entropia de Tsallis através de MRF para o Parcelamento de Ressonância Magnética Cerebral**. 2019. 49 f. Tese (Doutorado - Programa de Pós-Graduação em Física Aplicada à Medicina e Biologia) - Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto - SP, 2019.

A quantificação das alterações do volume do tecido intracraniano na ressonância magnética (RM) auxilia os especialistas a analisar os efeitos das alterações naturais ou patológicas. Como essas alterações podem ser sutis, a precisão do método de compartimentação influencia os estudos para analisar e quantificar os tecidos cerebrais. Nesta tese, revisamos os métodos recentes de segmentação do cérebro usados em ferramentas de imagens médicas. Em seguida, investigando a origem dos erros que podem ocorrer nos algoritmos de segmentação revisados, um pipeline híbrido é proposto para mitigar a influência desses erros. No primeiro capítulo, alguns pré-requisitos sobre estatística e modelos estatísticos e, em seguida, dois estimadores mais utilizados para os parâmetros do modelo são ilustrados. O segundo capítulo explica o uso de um modelo estatístico para segmentar imagens cerebrais. Além disso, as desvantagens desses métodos são discutidas. No terceiro capítulo, propomos um método de segmentação baseado na q-entropia modificada através de um campo aleatório modificado de Markov (Mqe-MMRF) para melhorar a precisão da parcela dos tecidos cerebrais. No último capítulo, os métodos propostos foram submetidos a duas estratégias para avaliar Mqe-MMRF, ou seja, uma simulação de diferentes níveis de ruído em dados de ressonância magnética e um conjunto de vinte dados de ressonância magnética disponíveis a partir de MRBrainS13 como desafio de segmentação de tecido cerebral. Nós acessamos nove métricas de qualidade

de segmentação em comparação com delineamentos de tecidos de referência para avaliar o Mqe-MMRF. As simulações de ruído de ressonância magnética mostraram apenas 4,8 % de decréscimo nas métricas de pontuação de segmentação após a adição de artefatos de ruído de 40 % e 9 % de não uniformidade e de ruído Gaussiano, respectivamente. Para cinco sujeitos de treinamento, encontramos melhoras significantes médias nas métricas de similaridade, para cérebro inteiro 0,78, Matéria Branca 2,91, Matéria Cinzenta 3,85 e Líquido Cefalorraquidiano 3,83 % (p-valores <0,02) nas métricas quando o Mqe-MMRF é comparado a métodos estado da arte. O Mqe-MMRF foi realizado em 15 outros sujeitos reais no desafio on-line MRBrainS13, e os resultados mantiveram uma classificação mais alta do que as ferramentas de referência, ou seja, FreeSurfer, SPM e FSL. Como o método proposto melhorou a precisão da segmentação do cérebro e classificou o melhor desempenho para GM, ele pode ser usado em estudos morfológicos quantitativos do cérebro.

Palavras-chave: 1.Segmentação de Imagem 2.Expectativa-Maximização 3.Algoritmo K-mean 4.q-entropia 5.Markov Random Field 6.Ressonância magnética

Abstract

AZIMBAGIRAD, M. **Tsallis-Entropy Segmentation through MRF for Brain Magnetic Resonance Parcellation**. 2019. 49 f. Thesis (Ph.D. - Postgraduate Program in Physics Applied to Medicin and Biology) - Faculty of Philosophy, Sciences and Literature, University of São Paulo, Ribeirão Preto - SP, 2019.

Quantifying the intracranial tissue volume changes in magnetic resonance imaging (MRI) assists specialists to analyze the effects of natural or pathological changes. Since these changes can be subtle, the accuracy of the compartmentalization method influences studies to analyze and quantify brain tissues. In this thesis, we review the recent brain segmentation methods used in medical imaging tools. Then by investigating the source of mistakes which may happen in the reviewed segmentation algorithms, a hybrid pipeline is proposed in order to mitigate the influence of such mistakes. In the first chapter, some prerequisites about statistics and statistical models and then two most used estimators for the model parameters are illustrated. Second chapter, explained using statistical model to segment brain images. In addition, drawbacks of these methods are discussed. In the third chapter, we propose a segmentation method based on modified q-entropy through a modified Markov random field (Mqe-MMRF) to improve the accuracy of brain tissues parcellation. In the last chapter, we underwent two strategies to evaluate Mqe-MMRF, i.e., a simulation of different levels of noise on MRI data and a set of twenty MRI data available from MRBrainS13 as brain tissue segmentation challenge. We accessed nine segmentation quality metrics compared to reference tissues delineations to evaluate Mqe-MMRF. MRI noise simulations showed only 4.8 % decreasing for segmentation scores metrics after adding 40 % and 9 % non-uniformity and Gaussian noise artifacts, respectively. For five training subjects,

we, found significant mean improvements in the similarity metrics, for whole brain 0.78, White Matter 2.91, Gray Matter 3.85 and Cerebrospinal Fluid 3.83 % (p-values < 0.02) in the metrics when Mqe-MMRF is compared to the other state of the art methods. The Mqe-MMRF was performed on 15 other real subjects in MRBrainS13 online challenge, and the results held a higher rank than the reference tools, i.e., FreeSurfer, SPM, and FSL. Since the proposed method improved the precision of brain segmentation and ranked the best performance for GM, it can be used in morphological brain quantitative studies.

Keywords: 1.Image Segmentation 2.Expectation-Maximization 3.K-mean Algorithm 4.q-entropy 5.Markov Random Field 6.MRI

List of Figures

List of Tables

List of Abbreviations

EM	Expectation-Maximization
GMM	Gaussian Mixture Model
PDF	Probability Distribution Function
MLE	Maximum Likelihood Estimation
MME	Method of Moments Estimation
BGS	Boltzmann-Gibbs-Shannon
MRF	Markov Random Field
ICM	Iterated Conditional Modes
WM	White Matter
GM	Gray Matter
CSF	Cerebrospinal Fluid
MRI	Magnetic Resonance Image
MAP	Maximum A Posteriori
Mqe	Modified q-entropy
MMRF	Modified Markov Random Field
DICE, DC	Dice Coefficient
JACCARD	Jaccard Coefficient
AUC	Area under ROC Curve

KAPPA	Cohen Kappa
RNDIND	Rand Index
ADJRIND	Adjusted Rand Index
ICCORR	Interclass Correlation
VOLSMTY	Volumetric Similarity Coefficient
MUTINF	Mutual Information
CSO	Contour Segmentation Objects
AVD	Absolute Volume Difference
HD	Hausdorff Distance

Contents

Chapter 1

Statistical Prerequisites

In this chapter we review some prerequisites about statistics and statistical models, specifically, Gaussian models. Then two most used estimators for the parameters of the exemplified models is presented. We analyze the performance of estimators on the models and compare them. In the end, three mechanical statistics models are reviewed which used to describe segmentation systems in next chapters.

1.1 Introduction

In almost all science, researchers try to reveal undiscovered relations or at least conditional relations[1] that govern the behavior of the object under investigation. Since these relations are not directly accessible, researchers try to formulate them in terms of hypotheses or models. In this way, mathematical modeling helps us to formulate interested phenomena.

A mathematical model is usually defined as a combination of relations and variables. The relations can be described by operators, such as algebraic operators, functions, differential operators and so on. The variables are abstractions of the components of a system, that can be quantified. The way that the variables and the relations are combined is according to their structure like; Linear, Nonlinear, Static, Dynamic, Statistic (Stochastic), etc. Essence and symptoms of a system lead us to use one of these models or a combination of them in order to predict the behavior of the system.

[1] Mostly finding a global rule works for all conditions may be hard or even impossible

Among these classes, statistical models cover a wide variety of those phenomena which have an inherently stochastic nature. For instance in physics, biology, chemistry, neurology and even social science, several phenomena can be modeled in this class. Although we are not restricted to use one class of the models in this thesis, the statistical model are explained as follows.

A statistical model is a class of mathematical model, which depicts a set of variables and relations concerning the generation of some sample data, and similar data from a larger population. There are three purposes for a statistical model, according to Konishi and Kitagawa [1]:

- Predictions
- Extraction of information
- Description of stochastic structures

Generally, these aims are interested in all studies that need to be modeled.

We leave the elementary concepts and definitions about statistics and probability provided in almost all fundamental textbooks, e.g., [2–4] and focus on statistical models.

1.2 Statistical Model

In mathematical terms, a statistical model is usually thought of as a pair (S, P), where S is the set of possible observations, i.e., the sample space, and P is a set of probability distributions (relations) on S [5]. Probability distributions function (PDF or shortly f) as a part of the statistical model includes a function and one (or more) parameter(s) known by θ. Usually pair (f, θ) are used as PDF of a S or population to describe the relations between its members. Parameters are descriptive measures of an entire population. However, their values are usually unknown because it is infeasible to measure an entire population. Therefore, due to describe a model, usually, a sample is acquired from the population. Then a model is guessed for the sample. In order to estimate the parameters of the model, usually, two methods are used:

- Maximum Likelihood Estimation (MLE)
- Method of Moments Estimation (MME)

MME method is discussed in [6] which is beyond of this study. MLE will shortly be explained in next section along with two examples describing the implementation of MLE.

1.2.1 Parameter Estimation by MLE

MLE as an estimator method for the parameters of a statistical model has widely been used. For an instance (in Wikipedia), one may be interested to know the mean heights of adult penguins in a region. But this is not possible to measure the height of each penguin in a population due to cost, time and other limitations. Assume that the heights has an PDF, e.g., Gaussian (Normal) model ($\mathcal{N}(\mu, \sigma)$) with unknown parameters mean (μ) and variance (σ). The μ and σ can be estimated with MLE while only knowing the heights of some sample of the overall population. MLE would accomplish this by taking the mean and variance as parameters and finding particular parametric values that make the observed results the most probable given the model.

Suppose we have a random variables $X_1, X_2, ..., X_n$ by a known (or guessed) PDF (f) depends on some unknown parameter θ, written $f(X = X_i; \theta)$. Primary goal here is the estimation of the θ, such that $f(X = x_i; \theta)$ is a *good* estimate of the model. This goodness can be defined as a specific error calculated by any method for the guessed PDF with estimated parameters. If the random variables are assumed to be normally distributed with mean μ and variance σ^2 and $x_1, x_2, ..., x_n$ are the observed values of the random sample, then our goal is to find a good estimate of μ and σ^2. It seems reasonable that a good estimate of $\theta = (\mu, \sigma)$ would be the value such that maximize the *joint density function*[2] for all observations

$$p(X_1 = x_1, X_2 = x_2, ..., X_n = x_n; \theta) = f(X = x_1, x_2, ..., x_n; \theta). \tag{1.1}$$

Now we look at this function from a different perspective by considering the observed values $x_1, x_2, ..., x_n$ to be fixed parameters, whereas θ be the function's variable and

[2]is a PDF which gives the probability that each of X_i falls in any particular range or discrete set of values specified for that variable

allowed to vary freely

$$L(\theta) = f(X = x_1, x_2, ..., x_n; \theta) = f(x_1; \theta) \times f(x_2; \theta)...f(x_n; \theta) \tag{1.2}$$

$$\Rightarrow L(\theta) = \prod_{i=1}^{n} f(x_i; \theta) \tag{1.3}$$

this function is called the *likelihood*. Now, in light of the basic idea of MLE, one reasonable way to proceed is to treat the likelihood function $L(\theta)$ as a function of θ, and find the value of $\tilde{\theta}$ which maximizes it. In practice, it is often more convenient when working with the natural logarithm of the likelihood function, called the *log-likelihood* defined as follow:

$$\ln L(\theta) = \ln \prod_{i=1}^{n} f(x_i; \theta) = \sum_{i=1}^{n} \ln f(x_i; \theta) \tag{1.4}$$

Several properties of this function illustrated in [7]. There are also studies have been working on different types of MLE and its application (refer to the last few [8–13]). In the following, we provide two examples to explain how MLE works and may not be applicable easily for parameter estimation of all models.

1.2.2 MLE for Gaussian distribution

Assume that the sample $x_1, x_2, ..., x_n$ are chosen from a population with Gaussian distribution and unknown parameters mean μ and variance σ^2, i.e., $\mathcal{N}(x_i; \mu, \sigma)$. Let $\mathcal{N}(x_i; \mu, \sigma)$ rewrite as a bivariate function of $\theta_1 = \mu$ and $\theta_2 = \sigma^2$, then

$$f(x_i; \theta_1, \theta_2) = \frac{1}{\sqrt{\theta_2}\sqrt{2\pi}} e^{[\frac{-(x_i-\theta_1)^2}{2\theta_2}]} \tag{1.5}$$

for $-\infty < \theta_1 < \infty$ and $0 < \theta_2 < \infty$. Therefore, likelihood function will be:

$$L(\theta) = L(\theta_1, \theta_2) = \prod_{i=1}^{n} f(x_i; \theta_1, \theta_2) = \theta_2^{\frac{-n}{2}} (2\pi)^{\frac{-n}{2}} e^{[\frac{-1}{2\theta_2}\Sigma_{i=1}^{n}(x_i-\theta_1)^2]} \tag{1.6}$$

As a bivariate function in calculus, to find the maximizer point of $L(\theta_1, \theta_2)$, we need to take partial derivative and setting it to zero to find $(\tilde{\theta_1}, \tilde{\theta_2})$. Since this function has nonlinear form and a lots of terms, logarithm form of the function will be much easier to take partial derivative. Therefore, the log-likelihood function is:

$$\ln L(\theta_1, \theta_2) = -\frac{n}{2} \ln \theta_2 - \frac{n}{2} \ln(2\pi) - \frac{\sum_{i=1}^{n}(x_i - \theta_1)^2}{2\theta_2} \tag{1.7}$$

Now, upon taking the partial derivative of the log-likelihood with respect to θ_1, and setting to 0, we have:

$$\frac{d \ln L(\theta_1, \theta_2))}{d\theta_1} = -\frac{-2 \times \sum_{i=0}^{n}(x_i - \theta_1)}{2\theta_2} = 0 \tag{1.8}$$

Now, multiplying through by θ_2, and distributing the summation, we get:

$$(\sum_{i=1}^{n} x_i) - n\theta_1 = 0 \tag{1.9}$$

by solving for θ_1 and putting on its tilde, we have shown that the maximum likelihood estimate of θ_1 is:

$$\tilde{\theta_1} = \tilde{\mu} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} \tag{1.10}$$

Now partial derivitive respect to θ_2 and set to 0 is:

$$\frac{d \ln L(\theta_1, \theta_2))}{d\theta_2} = -\frac{n}{2\theta_2} + \frac{2 \times \sum_{i=1}^{n}(x_i - \theta_1)^2}{4\theta_2^2} = 0 \tag{1.11}$$

multiplying through by $2\theta_2^2$ we have:

$$-n\theta_2 + \sum_{i=1}^{n}(x_i - \theta_1)^2 = 0 \tag{1.12}$$

therefore, the maximum likelihood estimate of θ_2 is:

$$\tilde{\theta_2} = \tilde{\sigma}^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} \tag{1.13}$$

Clearly the second partial derivative of the log-likelihood is negative. Therefore, $(\tilde{\theta_1}, \tilde{\theta_2})$ maximizes $L(\theta_1, \theta_2)$. In summary, we showed that MLE estimated the two parameters of Gaussian model as follow:

$$\tilde{\mu} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} \ , \ \tilde{\sigma}^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}. \tag{1.14}$$

what we expect as an inductive reasoning[3] for the population.

[3]Inductive reasoning is a method of reasoning in which the premises are viewed as supplying some evidence for the truth of the conclusion

1.2.3 MLE for Gaussian Mixture Model

Let we have a random variables $X_1, X_2, ..., X_n$ with a PDFdepends on some unknown parameter θ, $f(X = X_i; \theta)$, i.e., a Gaussian Mixture Model (GMM) with three components[4] $C = \{C_1, C_2, C_3\}$. Therefore, each component has own normal distribution $\mathcal{N}(x_i; \mu_k, \sigma_k^2)$ for $k = 1, 2, 3$. The mixture component weights (proportions) are defined as:

$$Pro_k = p(X \in C_k), s.t \sum_{k=1}^{3} Pro_k = 1 \tag{1.15}$$

then

$$p(X \in C) = \prod_{i=1}^{3} Pro_k^{d_k} \tag{1.16}$$

which $d_j = 1$ if $C = C_j$ and for $k \neq j$, $d_k = 0$. Then the PDF of GMM by three components is:

$$f(X = x_i; \theta = \{\theta_1, \theta_2\}) = \sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_i; \mu_k, \sigma_k) \tag{1.17}$$

where $\theta_1 = \{\mu_1, \mu_2, \mu_3\}$ and $\theta_2 = \{\sigma_1, \sigma_2, \sigma_3\}$. We can define a conditional probability

$$p(X = x_i \mid x_i \in C_k) = \mathcal{N}(x_i; \mu_k, \sigma_k) \tag{1.18}$$

then,

$$p(X = x_i \mid x_i \in C) = \prod_{k=1}^{3} \mathcal{N}(x_i; \mu_k, \sigma_k)^{d_k}. \tag{1.19}$$

Now the likelihood for the GMM is

$$L(\theta) = P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n; \theta) = \prod_{i=1}^{n} \sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_i; \mu_k, \sigma_k) \tag{1.20}$$

and log-likelihood for GMM is

$$\ln L(\theta) = \ln \prod_{i=1}^{n} \sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_i; \mu_k, \sigma_k) = \sum_{i=1}^{n} \ln[\sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_i; \mu_k, \sigma_k)]. \tag{1.21}$$

Now taking the partial derivative of the log-likelihood with respect to μ_1

$$\frac{d \ln L(\theta)}{d\mu_1} = \sum_{i=1}^{n} \frac{Pro_1 \times \frac{x_i - \mu_1}{\sigma_1^2} \times \mathcal{N}(x_i; \mu_1, \sigma_1)}{\sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_i; \mu_k, \sigma_k)} \tag{1.22}$$

[4]It can be any number, but we only use three component in this thesis

now we need to set Eq.(1.22) to zero to find $\tilde{\mu}_1$ maximizing $L(\theta)$. It will not be easy to solve recent equation. Beside numerical methods, Expectation-Maximization (EM) algorithm is an iterative method to find the maximizer of likelihood, i.e., the parameters of GMM. Using EM for finding the parameters of GMM are illustrated as follows.

1.3 EM Algorithm

The EM algorithm is a very general iterative algorithm for parameter estimation by maximum likelihood when some of the random variables involved are not observed or calculating $\frac{d \ln L(\theta)}{d\theta} = 0$ has not analytical answer. This idea has been used for many years before [14] in a problem in missing information principle provided the theoretical foundation of the underlying idea. Later, the term EM was introduced in [15] where proof of general results about the behavior of the algorithm was first given as well as a large number of applications. EM algorithm, uses an initial guess, i.e., $\theta^{(0)}$ for θ, then on the first iteration compute

$$Q(\theta; \theta^{(0)}) = E_{\theta^{(0)}}[\ln L(\theta)] \tag{1.23}$$

where $Q(\theta; \theta^{(0)})$ is now maximized with respect to θ, that is, $\theta^{(1)}$ is found such that

$$Q(\theta^{(1)}; \theta^{(0)}) \geq Q(\theta; \theta^{(0)}) \tag{1.24}$$

for all possible θ. Thus the EM algorithm consists of an E-step (Expectation step) followed by an M-step (Maximization step) defined as follows:

I. **E-Step**: Compute $Q(\theta; \theta^{(t)})$ where

$$Q(\theta; \theta^{(t)}) = E_{\theta^{(t)}}[\ln L(\theta)] \tag{1.25}$$

II. **M-Step**: Find $\theta^{(t+1)}$ such that

$$Q(\theta^{(t+1)}; \theta^{(t)}) \geq Q(\theta; \theta^{(t)}) \tag{1.26}$$

for all possible θ. The E-step and the M-step are repeated alternately until the difference $L(\theta^{(t+1)}) - L(\theta^{(t)})$ is less than *epsilon*, where *epsilon* is a prescribed

small quantity. The computation of these two steps simplify a great deal with finding the maximizer of the log-likelihood in analytical way for θ. There are several examples discussed to illustrate these steps in different cases [16]. As a general algorithm available for complex maximum likelihood computations, the EM algorithm has several appealing properties relative to other iterative algorithms such as Newton-Raphson. First, it is typically easily implemented because the E-step of each iteration only involves taking expectations over complete-data conditional distributions and the M-step only requires complete-data maximum likelihood estimation. Secondly, it is numerically stable and converges, to a local maximum or saddle point of MLE . A disadvantage of EM is the rate of convergence which can be extremely slow if many data are missing. Dempster, Laird, and Rubin [15] showed that convergence is linear with the rate proportional to the fraction of information. In the next section, EM algorithm is used to estimate the parameters for GMM which was not easy to estimate by MLE.

1.3.1 EM algorithm for GMM

Let $X_1, X_2, ..., X_n$ come from three sets $C = \{C_1, C_2, C_3\}$ and each set has own normal distribution $\mathcal{N}(x_i; \mu_k, \sigma_k)$ consequently we have a GMM distribution as defined in Section 1.2.3. Consider the following conditional probability[5] and using Eqs. (1.15) and (1.17) and Bayes rule [6] we have

$$p(x_i \in C_j \mid X = x_i) = \frac{p(x_i \in C_j) \times p(X = x_i \mid x_i \in C_j)}{p(X = x_i)} \tag{1.27}$$

$$= \frac{Pro_1 \times \mathcal{N}(x_i; \mu_1, \sigma_1)}{\sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_i; \mu_k, \sigma_k)} = w_{ji}, \ for\ j = 1, 2, 3\ and\ i = 1, 2, ..., n \tag{1.28}$$

Assume that the w_{ji} are given which are a part of $\frac{d \ln L(\theta)}{d\mu_j} = 0$ (in Eq. (1.22)). By substitution of these w_{ji} in Eq. (1.22) we have

$$\frac{d \ln L(\theta)}{d\mu_j} = \sum_{i=1}^{n} w_{ji} \times \frac{x_i - \mu_j}{\sigma_j^2} = 0 \tag{1.29}$$

[5] or posterior probability which is the probability of the parameters θ given the observed data X; $p(\theta \mid X)$

[6] $P(B \mid A) = \frac{P(A \mid B) \times P(B)}{P(A)}$

$$\Rightarrow \frac{1}{\sigma_j^2}(\sum_{i=0}^{n} w_{1i}x_i - \sum_{i=0}^{n} w_{ji}\mu_j) = 0 \Rightarrow \mu_j = \frac{\sum_{i=0}^{n} w_{ji}x_i}{\sum_{i=0}^{n} w_{ji}} \tag{1.30}$$

then we will have

$$\tilde{\mu}_j = \frac{\sum_{i=0}^{n} w_{ji}x_i}{\sum_{i=0}^{n} w_{ji}}, \quad j = 1, 2, 3 \tag{1.31}$$

By taking derivative on another parameter, i.e., σ we have

$$\tilde{\sigma}_j = \frac{\sum_{i=0}^{n} w_{ji}(x_i - \mu_j)^2}{\sum_{i=0}^{n} w_{ji} - \frac{\sum_{i=0}^{n} w_{ji}^2}{\sum_{i=0}^{n} w_{ji}}}, \quad j = 1, 2, 3 \tag{1.32}$$

and

$$\tilde{Pro_j} = \frac{\sum_{i=0}^{n} w_{ji}}{n}, \quad j = 1, 2, 3 \tag{1.33}$$

Therefore, we found $\tilde{\theta} = (\tilde{\theta}_1, \tilde{\theta}_2) = (\tilde{\mu}_j, \tilde{\sigma}_j)$ to maximize $L(\theta)$. This estimate can be performed if we have that posterior probability introduced in Eq. (1.27). Indeed, we do not have that weights, but by giving them we can estimate the parameters and by having parameters we can calculate the weights. EM algorithm helps to solve this problem by these two steps:

I. **E-Step**: Calculate weights (by having initial guess for the parameters $\tilde{\theta}$)

II. **M-Step**: Calculate parameters which maximize $L(\theta)$.

We can perform and repeat these two steps until the estimations for parameters do not significantly change in respect to previous performance. There are two questions about implementation of EM algorithm to estimate parameters of GMM. First, how expectation step lead us to calculate the weights? To answer this question, assume that $c = c_1, c_2, ..., c_n$ are hidden data which show the classification of $X = x_1, x_2, ..., x_n$ belong to one of $\{C_1, C_2, C_3\}$. We said that (in Eq. (1.25)) E-step is calculating $E_{c|X,\theta}[log(L(\theta; X, C))]$. Then by using Eqs. (1.16) and (1.19)

$$E_{c|X,\theta}[log(L(\theta; X, C))] = E_{c|X,\theta}[log(p(X, C))] = E_{c|X,\theta}[log(p(X \mid C) \times p(X \in C))]$$

$$= E_{c|X,\theta}[log(\prod_{i=1}^{n}\prod_{k=1}^{3} \mathcal{N}(x_i; \mu_k, \sigma_k)^{d_k} \times Pro_k^{d_k})]$$

$$= E_{c|X,\theta}[\sum_{i=1}^{n}\sum_{k=1}^{3} log(\mathcal{N}(x_i; \mu_k, \sigma_k) \times Pro_k)^{d_k}]$$

$$= E_{c|X,\theta}[\sum_{i=1}^{n}\sum_{k=1}^{3} d_k log(\mathcal{N}(x_i;\mu_k,\sigma_k) \times Pro_k)]$$

$$= E_{c|X,\theta}[\sum_{i=1}^{n}\sum_{k=1}^{3} d_k (log\mathcal{N}(x_i;\mu_k,\sigma_k) + logPro_k)] \tag{1.34}$$

now Expectation operator moves[7] inside the summations and Eq. (1.3.1) is equal to

$$= \sum_{i=1}^{n}\sum_{k=1}^{3} E_{c_i|X,\theta}[d_k](log\mathcal{N}(x_i;\mu_k,\sigma_k) + logPro_k) \tag{1.35}$$

which $E_{c|X,\theta}$ is the probability of $x_i \in C_k$ if $X = x_i$. It means

$$E_{c_i|X,\theta}[d_k] = p(x_i \in C_1 \mid X = x_i) = \frac{Pro_1 \times \mathcal{N}(x_i;\mu_1,\sigma_1)}{\sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_i;\mu_k,\sigma_k)} = w_{1i}. \tag{1.36}$$

Therefore, E-step leads to Eq. (1.36) which is the weights. Second step in EM algorithm was finding the maximization point. Then, let

$$\frac{dE_{c|X,\theta}[log(L(\theta;X,C))]}{d\mu_1} = 0 \tag{1.37}$$

equally, by taking the partial derivative of Eq. (1.35) and set it to 0 we have

$$\sum_{i=1}^{n} w_{1i}(\frac{x_i - \mu_1}{\sigma_1^2}) = 0 \tag{1.38}$$

which is exactly Eq. (1.29). Second question is why finding the maximum of $E_{c|X,\theta}[\log(L(\theta;X,C))]$ is equal to find the maximum of $\log(L(\theta))$. Shortly $E_{c|X,\theta}[\log(L(\theta;X,C))]$ is a lower bound of $\log(L(\theta))$. Therefore, if we increase $E_{c|X,\theta}$ means we increase $\log(L(\theta))$. In Jensen inequality it has been proved

$$E[\log(X)] \leq \log(E[X]). \tag{1.39}$$

but we have

$$L(\theta) = p(X,C) = \sum_{k=1}^{3} p(X,C_i). \tag{1.40}$$

By taking logarithm of Eq. (1.40) and multiply $\frac{p(X \in C_i)}{p(X \in C_i)}$ to each terms

$$\log L(\theta) = \log \sum_{k=1}^{3} p(X,C_i) = \log(\sum_{k=1}^{3} p(X,C_i) \times \frac{p(X \in C_i)}{p(X \in C_i)})$$

[7] Linear property of Expectation operator

$$= \log(E_C[\frac{p(X,C)}{p(X \in C)}]) \geq E_C[\log(\frac{p(X,C)}{p(X \in C)})], \quad (using\ (1.39)) \tag{1.41}$$

which is equal to

$$= E_C[\log p(X,C)] - E_C[\log p(X \in C)] \tag{1.42}$$

The second term above does not depend on θ. Therefore, we only need to maximize the first term on θ and done. Then M-Step is

$$\tilde{\theta} \leftarrow \underset{\theta}{\operatorname{argmax}}\ E_C[\log p(X,C)]$$

To find $\tilde{\theta}$ which maximize (1.41)

$$E_C[\log(\frac{p(X,C)}{p(X \in C)})] = E_C[\log(\frac{p(X)p(C \mid X)}{p(X \in C)})] = \log p(X) + E_C[\log \frac{p(X \mid C)}{p(X \in C)}]$$

$$= \log p(X) - E_C[\log \frac{p(X \in C)}{p(X \mid C)}] \tag{1.43}$$

Eq. (1.43) will be maximized if

$$p(X \in C) \leftarrow p(C \mid X)$$

which is E-step.

1.3.2 EM algorithm for GMM by q-log

In the MLE method for parameters estimation of a model, the likelihood function is employed to be maximized. Maximizing the likelihood gives the optimum parameters, but the maximum likelihood is not always easy obtained. Hence, logarithm of likelihood can be used instead. Though we showed even log-likelihood may not help here we use q-log-likelihood to see a possible difference between these two function. Before using q-logarithm inside EM algorithm, let to review some definitions about q-logarithm. The q-exponential is a deformation of the exponential function using the real parameter q

$$e_q(x) = [1 + (1-q)x]^{1/(1-q)} \quad if\ q \neq 1\ and\ 1 + (1-q)x > 0, \tag{1.44}$$

and the q-logarithm is the inverse of q-exponential and a deformation of the logarithm using the real parameter q

$$\ln_q(x) = \frac{x^{1-q} - 1}{1-q} \quad if\ q \neq 1\ and\ x \geq 0, \tag{1.45}$$

Now instead of log-likelihood, we calculate q-log of likelihood for GMM. We know likelihood for GMM is

$$L(\theta) = P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n; \theta) = \prod_{i=1}^{n} \sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_i; \mu_k, \sigma_k) \quad (1.46)$$

then

$$\ln_q L(\theta) = \frac{[\prod_{i=1}^{n} \sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_i; \mu_k, \sigma_k)]^{1-q} - 1}{1-q}$$

$$= \frac{\prod_{i=1}^{n} [\sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_i; \mu_k, \sigma_k)]^{1-q} - 1}{1-q}$$

$$= \frac{\prod_{i=1}^{n} [Pro_1 \times \mathcal{N}(x_i; \mu_1, \sigma_1) + Pro_2 \times \mathcal{N}(x_i; \mu_2, \sigma_2) + Pro_3 \times \mathcal{N}(x_i; \mu_3, \sigma_3)]^{1-q} - 1}{1-q} \quad (1.47)$$

for taking derivative of multiplication of functions we have

$$u_1^n \times u_2^n \times ... \times u_m^n \overset{\frac{d}{dx}}{\Longrightarrow} [n \times \frac{d}{dx} u_1 \times u_1^{n-1}] \times u_2^n \times ... \times u_m^n$$

$$+ u_1^n \times [n \times \frac{d}{dx} u_2 \times u_2^{n-1}] \times u_3^n \times ... \times u_m^n + ... \quad (1.48)$$

therefore, taking derivative of $\ln_q L(\theta)$ over one of parameters for example, μ_1 is

$$\frac{d \ln_q L(\theta)}{d\mu_1} = \frac{(1-q)[Pro_1 . \frac{x_1 - \mu_1}{\sigma_1^2} . \mathcal{N}(x_1; \mu_1, \sigma_1) \times (\sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_1; \mu_k, \sigma_k))^{-q}] \times}{1-q}$$

$$\underline{\prod_{i=1, i\neq 1}^{n} [\sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_i; \mu_k, \sigma_k)]^{1-q} + (1-q)[Pro_1 . \frac{x_2 - \mu_1}{\sigma_1^2} . \mathcal{N}(x_2; \mu_1, \sigma_1) \times}$$

$$\underline{(\sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_2; \mu_k, \sigma_k))^{-q}] \times \prod_{i=1, i\neq 2}^{n} [\sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_i; \mu_k, \sigma_k)]^{1-q} + ...} \quad (1.49)$$

by simplify and separate terms we have

$$\frac{d \ln_q L(\theta)}{d\mu_1} =$$

$$[Pro_1 . \frac{x_1 - \mu_1}{\sigma_1^2} . \mathcal{N}(x_1; \mu_1, \sigma_1) \times (\sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_1; \mu_k, \sigma_k))^{-q}] \times$$

$$\prod_{i=1, i\neq 1}^{n} [\sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_i; \mu_k, \sigma_k)]^{1-q} + [Pro_1 . \frac{x_2 - \mu_1}{\sigma_1^2} . \mathcal{N}(x_2; \mu_1, \sigma_1) \times$$

$$(\sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_2;\mu_k,\sigma_k))^{-q}] \times \prod_{i=1,i\neq 2}^{n} [\sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_i;\mu_k,\sigma_k)]^{1-q} + ... \quad (1.50)$$

now multiplying each term by

$$\frac{[\sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_i;\mu_k,\sigma_k)]^{1-q}}{[\sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_i;\mu_k,\sigma_k)]^{1-q}} \quad (1.51)$$

which for the first term $i = 1$ for the second one $i = 2$ and so on. Then we have

$$\frac{d \ln_q L(\theta)}{d\mu_1} =$$

$$[\frac{Pro_1.\frac{x_1-\mu_1}{\sigma_1^2}.\mathcal{N}(x_1;\mu_1,\sigma_1)}{\sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_1;\mu_k,\sigma_k)}] \times \prod_{i=1}^{n} [\sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_i;\mu_k,\sigma_k)]^{1-q}$$

$$+[\frac{Pro_1.\frac{x_2-\mu_1}{\sigma_1^2}.\mathcal{N}(x_2;\mu_1,\sigma_1)}{\sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_2;\mu_k,\sigma_k)}] \times \prod_{i=1}^{n} [\sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_i;\mu_k,\sigma_k)]^{1-q}$$

$$+ ... \quad (1.52)$$

but we know

$$\frac{Pro_1 \times \mathcal{N}(x_i;\mu_1,\sigma_1)}{\sum_{k=1}^{3} Pro_k \times \mathcal{N}(x_i;\mu_k,\sigma_k)} = w_{1i} \quad (1.53)$$

then by substitution and remove the same factor in all terms

$$\frac{d \ln_q L(\theta)}{d\mu_1} = 0$$

$$\frac{d \ln_q L(\theta)}{d\mu_1} = [\frac{x_1-\mu_1}{\sigma_1^2}.w_{11}] + [\frac{x_2-\mu_1}{\sigma_1^2}.w_{12}] + ... + [\frac{x_n-\mu_1}{\sigma_1^2}.w_{1n}] = 0$$

$$= \sum_{i=1}^{n} w_{1i} \times \frac{x_i-\mu_1}{\sigma_1^2} = 0 \quad (1.54)$$

but recent equation is exactly equal to Eq. (1.29). We proved the substitution of $\ln_q$ by ln, in MLE for GMM parameter estimation for any q result in the same answer. It is an example to show that using q-log may not change the results. Also, it can be proved without details that

$$\frac{d \ln_q L(\theta)}{d\mu_1} = \frac{(1-q)\acute{L}(\theta)L(\theta)^{-q}}{(1-q)} = \frac{\acute{L}(\theta)}{L(\theta)^q} = 0 \quad \Rightarrow \quad \acute{L}(\theta) = 0 \quad (1.55)$$

indeed, q is not important to find the maximization of $\ln_q L(\theta)$.

1.4 K-mean Algorithm

The K-means algorithm is another iterative method to simply estimate the parameters, specifically, the means of a population. This algorithm has been discovered by several researchers across different disciplines, introduced by Lloyd [17], and then Forgey [18] and others [19, 20]. A detailed history of K-means along with descriptions of several variations are given in [20], but here we review this algorithm very shortly. Let we have a population including three components as the same example used in MLE estimator for GMM. We use K-means algorithm to estimate the means of the components. The algorithm is initialized by three (K) giving points as the initial means representatives or *centroids* for the components. Then the algorithm iterates between two steps till converges to the best estimations:

I. **Data Assignment**: By using the given centroids, each member of the population is assigned to its closest centroid. Then this step clusters the population to three groups.

II. **Relocation of means**: For each cluster (group), its mean is recalculated.

The algorithm converges when the assignments (and hence the centroids) no longer change. The number of iterations required for convergence varies and may depend on population, but as a first cut, this algorithm can be considered linear in the dataset size. In respect to the other estimators like MLE, K-mean is faster. Nevertheless, the precision of the estimated parameters may not be acceptable in those studies looking for a high precision approximation. This issue is explained in the next chapter in a specific usage.

1.5 Additive and Nonadditive entropy

A statistical model, estimating how much information is required, on average, to identify random samples from a distribution is defined by Shannon entropy. Equally, the Shannon entropy equation provides a way to estimate the average minimum number of bits needed to encode a string of symbols, based on the frequency of the symbols. Since we desire to model such systems (any system where

has a PDF with GMM to estimate the probability of its states), it could be defined as a statistical model. We review the Shannon entropy and its generalized version aiming to estimate the parameters of a GMM as follows.

1.5.1 Additive entropy

The entropy of a system is often obtained from a probability distribution, where $p = \{p_i\}$ is the probability of finding the system in each possible state i. Therefore, $0 \leq p_i \leq 1$ and $\sum_{i=1}^{n} p_i = 1$ where n is the total number of states. For example, Shannon entropy described as

$$S = -\sum_{i=1}^{n} p_i \ln(p_i) \tag{1.56}$$

This formalism is restricted to the domain of validity of the Boltzmann-Gibbs-Shannon (BGS) statistics [21]. Assume that a BGS system can be decomposed into two independent statistical subsystems A and B. Then, the probability of the composite system is $P^{A+B} = P^A.P^B$. It has been verified that the Shannon entropy has the additive property (or extensivity in some sense)

$$S(A+B) = S(A) + S(B) \tag{1.57}$$

hence for three subsystems $C_1,\ C_2$ and C_3 it is defined

$$S(C_1 + C_2 + C_3) = S(C_1) + S(C_2) + S(C_3) \tag{1.58}$$

In order to calculate the Shannon entropy of each subsystem, it is needed to divide those probabilities to three parts. For example, $C_1 = \{p_1, p_2, ...p_t\}$, $C_2 = \{p_{t+1}, p_{t+2}, ...p_k\}$ and $C_3 = \{p_{k+1}, p_{k+2}, ...p_n\}$ then

$$\begin{aligned} S(C_1) &= -\sum_{i=1}^{t} \frac{p_i}{p_{C_1}} \ln(\frac{p_i}{p_{C_1}}) \\ S(C_2) &= -\sum_{i=t+1}^{k} \frac{p_i}{p_{C_2}} \ln(\frac{p_i}{p_{C_2}}) \\ S(C_3) &= -\sum_{i=k+1}^{n} \frac{p_i}{p_{C_3}} \ln(\frac{p_i}{p_{C_3}}) \end{aligned} \tag{1.59}$$

where p_{C_i} is sum of all probabilities in C_i. Finding the optimum t and k which maximize the Shannon entropy is desired to estimate the best statistical model describing the system. This entropy concept is used to describe an image segmentation process [21] in section (2.2.3).

1.5.2 Nonadditive entropy

For a certain class of physical systems, which entail long-range interactions, long time memory and fractal-type structures, some kind of extension appears to become necessary [22]. Inspired by multifractals concepts, Tsallis has proposed a generalization of the BGS statistics [23]. Tsallis statistics is currently considered useful in describing the thermostatical properties of nonadditive systems, and it is based on a generalized entropic form,

$$S_q = \frac{1 - \sum_{i=1}^{n} p_i^q}{q - 1} \tag{1.60}$$

where n is the total number of possibilities of the system and the real number q is an entropic index that characterizes the degree of nonextensivity. This expression meets the BGS entropy in the limit $q \to 1$ [23]. Tsallis entropy is nonextensive in such a way that for a statistical independent system (including two subsystems A and B), the entropy of the system is defined by the following pseudo additivity entropic rule

$$S_q(A + B) = S_q(A) + S_q(B) + (1 - q) \times S_q(A).S_q(B) \tag{1.61}$$

then

$$S_q(A + B) = \frac{1 - \sum_{i=1}^{t} p_i^q}{q - 1} + \frac{1 - \sum_{i=t+1}^{n} p_i^q}{q - 1} + (1 - q) \times \frac{1 - \sum_{i=1}^{t} p_i^q}{q - 1} \times \frac{1 - \sum_{i=t+1}^{n} p_i^q}{q - 1} \tag{1.62}$$

Equally, Tsallis entropy can be defined for a system composed of three subsystem using Eq. (1.61). Here also, finding the optimum t and k which maximize the Tsallis entropy is desired which can be extended to image processing areas, specifically for the image segmentation. This implementation is illustrated in section (2.2.4).

1.6 Markov Random Field

The last statistical model described in this thesis is Markov Random Field (MRF). This model is a stochastic process that specifies the local characteristics of a system [24]. A random field can be considered as a MRF if its probability distribution at any site depends only upon its neighborhood [25]. The Cliff-Hammersley theorem [26], says any MRF can be described by a probability distribution of the Gibbs form

$$P(f) = \frac{1}{Z} e^{-U(f)/T} \tag{1.63}$$

where Z is a normalization constant, T is a quantity analogous to temperature in statistical mechanics and $U(f) = \sum_i U_i(f)$ is an energy function that can be computed as the sum of local contribution of each neighborhood i. In all statistical approaches, when we look for a PDF to model the data, Maximum A Posteriori (MAP) is desired. Therefore, finding the best PDF will maximize $U(f)$ as well. Hence $U(f)$ can be rewritten as summation of two separated functions

$$U(f) = U_t(f) + \alpha . U_l(f) \tag{1.64}$$

where $U_t(f)$ is the total and $U_l(f)$ is local energy of a system. Parameter α is used here to control the effect of local characteristics. There are many different ways to define total energy function $U_t(f)$ and local energy function $U_l(f)$. The total energy function can be defined by Gaussian, GMM, entropy, q-entropy or any other models. Local energy can be defined based on the neighborhood specifications depends on the system. Using an MRF to describe an image segmentation method is illustrated in section (2.3).

1.7 Conclusion

In the first chapter, we described two estimators, i.e., EM and K-mean for the parameters of a model. besides, three statistical models, i.e., entropy, q-entropy, and MRF were explained for a system. In the next chapter, these estimators and models will be used to segment a brain image. Cons and Pros of these model and estimators will be discussed as well.

Chapter 2

Brain Segmentation Methods

In this chapter, a pipeline for brain segmentation is illustrated. Several preprocessing, labeling and post processing are explained using the models introduced in the first chapter. By a deep analysis of these techniques, their advantages and drawbacks are discussed.

2.1 Introduction

2.1.1 Image Segmentation

Image segmentation is the process of partitioning an image into multiple regions such that each region includes a set of pixels (in 3D image namely voxels). The goal of segmentation is to simplify and/or change the representation of an image into something more meaningful and easier to analyze [27]. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images [28]. More precisely, image segmentation is the process of assigning label to pixels such that the pixels with the same label represent certain characteristics such as color, intensity, or texture. When a segmentation applied to a stack of images, typical in medical imaging, the resulting contours can be used to create 3D reconstructions with the help of interpolation algorithms [28].

Several general-purpose algorithms and techniques have been developed for image segmentation [29] like:

- Thresholding
- Clustering methods
- Compression-based methods

- Histogram-based methods
- Edge detection
- Region-growing methods
- Partial differential equation-based methods
- Parametric methods
- Level set methods
- Fast marching methods
- Graph partitioning methods
- Watershed transform
- Model-based segmentation
- Multi-scale segmentation
- One-dimensional hierarchical signal segmentation

and many new methods which are monthly even daily have been provided by researchers in a specific purpose. In this thesis, we focused on medical image segmentation and the more specific area, i.e., Magnetic Resonance Image (MRI) brain segmentation. Therefore, an illustration of all segmentation methods, which may only be for a defined purpose, is beyond of this scope.

2.1.2 Brain Image Segmentation

MRI occupies a prominent role in medical sciences due to the capabilities of obtained images in diagnostic and treatment follow-up procedures. Measurement of brain tissue volumes, like White Matter (WM), Gray Matter (GM), Cerebrospinal Fluid (CSF) and lesions can help physicians to analyze the progress of diseases and treatments. For instance, brain (cerebral) atrophy is a common feature of many of the diseases that affect the brain [30]. In order to measure these volumes, a segmentation process is needed. Segmentation methods for brain are divided into three main groups; manual, semi-automatic and automatic. Although manual segmentation is the most trusted one, it is hugely time-consuming, especially, if segmentation of the whole brain volume is needed. Furthermore, inter- and intra-operator variability negatively impacts the reliability of such studies [31]. Therefore, automatic and semi-automatic methods are important and capable of assisting many clinical procedures involving image analysis. Since around 1989 [32](see also Figure 2.1), many automatic and semi-automatic brain segmentation methods have been provided by researchers. Simultaneously with the increasing number of such methods, precision, computation time or both have also been

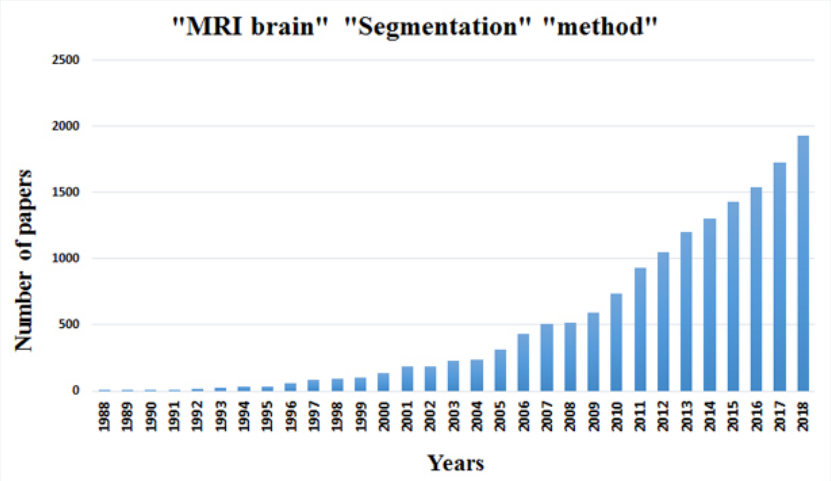


Figure 2.1: *By searching keywords "MRI brain" "Segmentation" "method" in google scholar year by year, it can be seen that the number of papers significantly increased during the years, similar to many subjects in researches. With the increase in numbers of studies, it is also expected the increase in quality and precision of developed methods.*

improved. Many of these algorithms are implemented inside packages, software and applications in order to be used in clinical institutions and research centers [33–37].

Commonly, used automatic segmenters for brain MR images identify GM, WM, and CSF with acceptable performance [38]. Nevertheless, neuroscientists are in great need of reliable segmentations of substructures of major tissue compartments for analyzing group differences in anatomy. Unfortunately, identifying these structures poses a serious challenge as their boundaries do not always coincide with tissue boundaries and may, therefore, not be visible in MR images. Automatic segmenters typically rely on prior information in order to delineate these invisible boundaries [39]. However, the use of prior information is not without risk as it can bias the automatic segmentation results [40]. Some successful attempts have been made to solve this problem, although current techniques are tailored and tuned to specific segmentation problems [41]. Moreover, the tuning results in assumptions making the implementations often challenging to adjust to other segmentation problems that deviate from the specific scenario. For example, the segmentation of compartments with large spatial variations within a population, such as lesions

in elderly patients, is difficult to segment with methods based on label propagation via image registration of a template [42]. Furthermore, many of these methods are restricted to MR images from specific acquisition protocols.

In order to decrease the effects of the issues above, the techniques sequence or pipeline for an acceptable MRI brain segmentation often includes three steps, i.e., preprocessing, labeling and post processing. The preprocessing prepares MRI image for labeling by decreasing the artifacts and selecting the region of interest. Then, a labeling technique is applied to label the tissues voxels in the brain image. Finally, the label map is adjusted by the post processing technique.

Three preprocessing techniques are applied for most brain MRI segmentation pipelines, i.e., brain extraction [43], noise reduction [44] and inhomogeneous field correction [45]. More details about preprocessing techniques can be found in medical image processing books, but for the labeling and the post processing sufficient details are provided in the next sections.

2.2 Image Labeling

WM, GM and CSF could be almost entirely distinguished by a different signal intensity range, i.e., grey level of pixels in MRI submodalities, namely T1 and T2 weighted images. We assume that all voxels in a 3D MRI (e.g., T1) are the members of a population. Then each voxel has a specific level of intensity in a defined range (e.g., 0 to 255). We can use a GMM to model the range intensity and the frequency of each level (figure 2.2, on the right the green curve is a GMM) of an image. Therefore, the labeling can be used to separate the tissues and creates a label map such that each tissue has a different label. Labeling for such image uses two thresholds (t_1,t_2) to separate the three tissues. WM, GM, and CSF intensity range from brightest to darkest grey levels respectively can be seen in Fig. 2.2, on the left. When labeling a T1 image, the voxels having intensity values smaller than t_1 are labeled as CSF, between t_1 and t_2 labeled as GM and greater than t_2 labeled as WM as seen in Fig. 2.2, on the right.

The parameters of the GMM must be estimated good enough to model the intensity-frequency graph (histogram) of an image. In the next section, EM

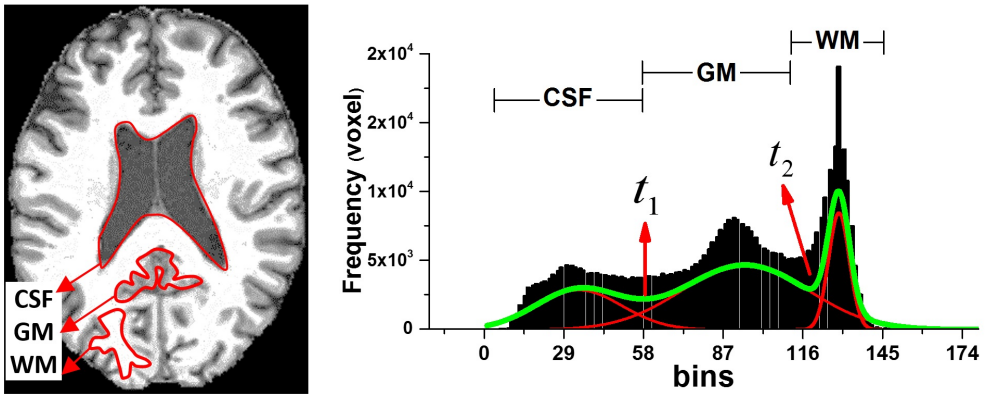


Figure 2.2: *Left: Red regions show a brain slice of a 3D MRI, three samples of WM, GM and CSF. Right: Gray level intensity histogram of the image, 0 is the darkest and 255 is the brightest level. The green curve is a GMM fitted to the histogram.* t_1 *and* t_2 *are two thresholds.*

estimator is used to estimate the parameters and then label the image.

2.2.1 EM Brain Segmentation

EM segmentation algorithms have become a popular tool to perform MR image segmentation [46, 47]. They not only segment the image into a label map, which is the unknown data in the EM formulation, but also incorporate estimation of parameters related to either the image acquisition (e.g., intensity inhomogeneity [47]), or the segmentation process (e.g., atlas registration parameters [48]). To achieve this, the EM algorithm alternates between computing the unknown data (its expectation) given the parameters and observed data and refining the estimation of the parameters based on the observed data (see section 1.3). For instance, let a given brain image includes three parts (components) WM, GM, and CSF. If we use a Gaussian mixture model to fit a cure to histogram of this image, nine parameters include three means (μ_1, μ_2, μ_3), three standard deviation $(\sigma_1^2, \sigma_2^2, \sigma_3^3)$ and three proportions (π_1, π_2, π_3) are needed to be estimated by EM algorithm. Here, μ_1 is the mean of those intensity voxels which belongs to CSF, σ_1^2 is the standard deviation of them and π_1 is the proportion of CSF in the histogram and so on. When EM algorithm is converged, for each voxel, three weights are calculated (Eq. (1.28)). The decision for labeling the voxels is based on the maximum weight. The pseudo code for EM algorithm for brain segmentation can be seen in Appendix (A.1).

In EM algorithm, it is assumed the input image includes only three main tissues (brain) and background intensity zero produced by a brain extraction. Any brain extraction method can do this task [43, 49, 50]. In addition, the conditions to stop EM algorithm, i.e., a *epsilon* for two successive iterations and/or a limitation for a number of iterations, can be chosen by users.

2.2.1.1 Pros and Cons

Although the convergence of EM algorithm was proven, it may take a lot of iteration to achieve this when choosing not precise enough initial parameters. In addition, parameters estimation is computationally expensive to calculate. Furthermore, this method needs nine inputs which are strictly depend how to choose initial parameters. It may converge to local optimum instead of global [15]. Nevertheless, EM approach still is one of the most trustable parameter estimators.

2.2.2 K-mean brain Segmentation

The K-mean clustering algorithm is described in detail by Hartigan [51]. This algorithm aims to divide n points to k cluster so that the distance between each point and the center of its cluster is minimized. The center of each cluster is defined as the mean of that cluster what we explained in Section 1.4. Yet, assume an image with three clusters as WM, GM, and CSF. First, the mean intensity of each cluster are given. Then K-mean algorithm calculates the distance between each intensity voxel and the means. The minimum distance shows each voxel belongs to which cluster. The pseudocode of this algorithm is presented in Appendix (A.2).

In the K-mean algorithm, Euclidean distance, here, assumed the absolute difference value of means and the intensity voxels. *error* and *epsilon* can be chosen the same as described in EM algorithm.

2.2.2.1 Pros and Cons

K-mean algorithm needs less initial inputs than EM, only three values. In addition, the expense of calculations is less than EM, because in EM algorithm we need to calculate standard deviation and proportion as well. However, it should not be neglected a considerable mistake in K-mean may happen in some situations. In

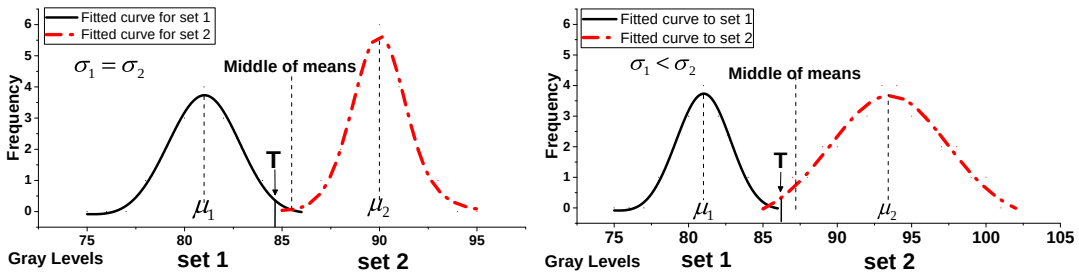


Figure 2.3: *Representation of two situations that may occur when analyzing the voxel values of two different tissues.*

figure 2.3 (a) and (b), two Gaussian curves (black and red dash) were fitted to set 1 and set 2 (each set contains the gray level intensities of each tissue). In (a), knowing that $\sigma_1 = \sigma_2$, a decision based on K-mean or MAP will be the same for level T, i.e., both methods label T as set 1. But for different standard deviations, i.e., (b), K-mean labels T as belonging to set 1, while MAP labels T as belonging to set 2. This difference shows that, in some cases, K-Mean might not be enough precise to label a voxel. Therefore, in such situations, K-mean make mistake in labeling a voxel.

2.2.3 Entropy brain Segmentation

Using entropy to label an image was introduced by Kapur et.al [21] in 1985. Many years later, entropy was used to segment a brain image into three main tissues [52]. Assume that the brain image is a system and each level of its intensity as a state has a value (probability) defined by

$$p_i = \frac{f_i}{N} \tag{2.1}$$

where i is an intensity level, f_i is its associated frequency counted in the image and N is the total number of voxels. In an MRI by 8 bits levels, i can be from $1, ..., n = 2^8 = 256$ or $i \in [0, 255]$. Again let this image includes three subsystems (tissues). Then, Shannon entropy (see section 1.5.1) of this image using the probability definition Eq. (2.1) can be defined

$$S(CSF + GM + WM) = S(CSF) + S(GM) + S(WM) \tag{2.2}$$

or equivalently

$$S_{CSF+GM+WM} = -\sum_{i=1}^{t} \frac{p_i}{p_{CSF}} \ln(\frac{p_i}{p_{CSF}}) - \sum_{i=t+1}^{k} \frac{p_i}{p_{GM}} \ln(\frac{p_i}{p_{GM}}) - \sum_{i=k+1}^{n} \frac{p_i}{p_{WM}} \ln(\frac{p_i}{p_{WM}}) \tag{2.3}$$

where p_{CSF} is sum of all probabilities of levels in CSF and so on. In order to calculate Eq. (2.3), it is needed to determine t and k, where $1 < t < k < n$. Finding the optimum t and k which maximize the Shannon entropy of the image is a subproblem. These t and k indeed are two thresholds to divide levels to three parts. That is all levels less than t are belong to CSF, the levels between t and k are belong to GM and the rest (more than k) are belong to WM. Therefore, solving the subproblem

$$\arg\max_{t,k} S(CSF + GM + WM) \tag{2.4}$$

gives us $\tilde{t}$ and $\tilde{k}$ in order to segment the brain image into CSF, GM and WM. Subproblem (2.4) can be solved by any method. The first way to find $\tilde{t}$ and $\tilde{k}$ is to check all levels one by one. Another way is assuming the image has two tissue. Then we need just one threshold to segment the image. Next, we segment the segmented part which includes two parts by another entropy brain segmenter. In addition to these methods, the bisection search method can be used to find the optimum thresholds. The pseudo code of entropy labeling method can be seen in Appendix (A.3).

2.2.3.1 Pros and Cons

Although the first approach in order to solve Eq. (2.4) is computationally expensive, convergence is absolutely accessible (the number of all combinations of t and k is finite). In addition, by using recent CPUs with high speed of computing, this method will not be very slow in the low range of levels. In respect to K-mean and EM, entropy does not need any initial parameter.

2.2.4 q-entropy brain Segmentation

Albuquerque et al. [22], introduced entropy-based thresholding using Tsallis entropy (also known as q-entropy). Later, Deniz et al. [52], proposed the q-entropy

maximization as a criterion to segment the brain tissues in a patient with Multiple Sclerosis (MS) disease. Since then, q-entropy was reportedly used for segmenting lesion [17] and tumor [18] as well as the main brain tissues [16].

In section 1.5.2, it was explained how to calculate q-entropy of a system including two subsystems. Using the same definitions in section 2.2.3, q-entropy for the brain image [22] is

$$\begin{aligned} S_q(CSF+GM+WM) = S_q(CSF) + S_q(GM) + (1-q) \times [S_q(CSF).S_q(GM) + \\ S_q(CSF).S_q(WM) + S_q(GM).S_q(WM)] \\ + (1-q)^2 \times S_q(CSF).S_q(GM).S_q(WM) \end{aligned} \tag{2.5}$$

where $S_q(C)$ is defined by Eq. (1.60). In addition to t and k and p_is required in entropy brain segmentation, a q entropic index is needed to calculate Eq. (2.5). Then subproblem in q-entropy brain segmentation is

$$\arg\max_{q,t,k} S_q(CSF+GM+WM) \tag{2.6}$$

The pseudo code of q-entropy brain segmentation is presented in Appendix A.4.

2.2.4.1 Pros and Cons

In q-entropy labeling (like entropy algorithm), the optimum $\tilde{t}$ and $\tilde{k}$ can be chosen by the ways mentioned earlier. Choosing the optimum entropic index q and solving the maximization step may be time-consuming. Nevertheless, dependence on only one parameter q is an important advantage of q-entropy labeling. Furthermore, nonadditive statistics has proved to be appropriate to multifractal geometry condition which is a property of brain image structures.

2.3 Post Processing

MRF segmentation [53] is capable of classifying an image on its own, but is often used for modifying an initial segmentation as well [33, 34, 54–57]. In section 1.6, it was presented a MRF can be described by a probability distribution of the Gibbs form

$$P(f) = \frac{1}{Z} e^{-U(f)/T} \tag{2.7}$$

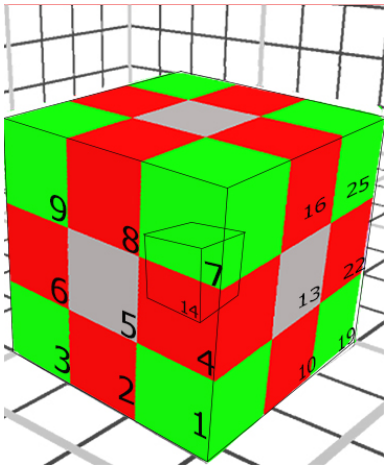
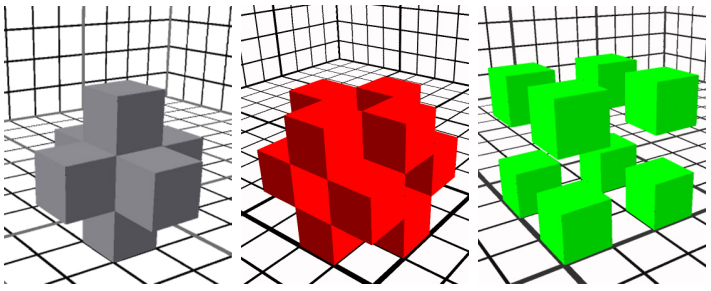


Figure 2.4: *A common site for the voxel in 3D images. If we enumerate the voxels from 1 to 25 in a 3x3x3 cubic, the number of center voxel will be 14. Three different colors are chosen for three kinds of voxels based on the distance from the center.*

where

$$U(f) = U_t(f) + \alpha.U_l(f). \tag{2.8}$$

In 3D MRI the neighborhood of a voxel is defined by a cubic surrounded the voxel. This cubic can be defined by different size or distance. For example, a common neighborhood for the voxels is all voxels around the center by Euclidean distance less than or equal $\sqrt{3}$ unit of voxel size. This neighborhood is presented in figure 2.4. Assume that a brain image is already labeled by any method. Therefore, for



(a) $d = 1$ (b) $d = \sqrt{2}$ (c) $d = \sqrt{3}$

Figure 2.5: *Three different sets of voxels in neighbors of a voxel. Distance is calculated by using Euclidean distance between centers of voxels.*

each labeled tissue, the parameters of a Gaussian PDF can be calculated. Let this Gaussian PDF as $U_t(f)$ and K-mean algorithm as $U_l(f)$ which are two common

selections for Eq. (2.7) [33, 34, 37, 56]). Since the best $U_t(f)$ is desired, Iterated Conditional Modes (ICM) is an optimization method [58] to minimize $U(f)$, because this maximize $P(f)$. For the voxel v_i in a brain image, ICM calculates three weights (using Gaussian PDFs) and three distances (using K-mean on only neighbors of v_i) corresponding to the possible cases in which v_i belongs to each tissue (CSF, GM or WM). Then by using Eq. (2.8) and finding the minimum value, v_i is relabeled. It should be mentioned when using MRF that K-mean mistakes (explanted in section 2.2.2.1) may propagate in MRF. In addition, we assume that in a site of a voxel, the labels (or at least majority of them) are labeled correctly else this wrong labeling may relabel a true label. The algorithm of ICM is presented in Appendix (A.5).

2.4 Conclusion

In this chapter four labeling and one refining labeling were introduced. Also, their advantages and disadvantages were discussed. The analysis of the methods leads us to find the source of failures and to improve their performance. We provide two modifications for q-entropy and MRF to decrease such that drawbacks and combine them as a new brain image segmentation method in the next chapter.

Chapter 3

Improving Segmentation Methods

In chapter 2. four well-accepted labeling methods were introduced. In this chapter, we propose and then evaluate a brain segmentation method based on modified q-entropy (Mqe) through a modified Markov random field (MMRF) to improve the accuracy of brain tissues parcellation (Mqe-MMRF). This method has been submitted to MRBrains13 [59] global challenge and the results held higher rank than the reference methods, specially, in Gray matter.

3.1 Introduction

Quantifying the intracranial tissue volume changes in magnetic resonance imaging (MRI) assists specialists to analyze the effects of natural or pathological changes. Since these changes can be subtle, the accuracy of the compartmentalization method influences studies to analyze and quantify brain tissues. Although there are already several methods to segment brain structures, these methods still present critical errors when dealing with sensitive applications. Such methods must be as precise as possible in segmenting and quantifying brain tissues to detect diseases and evaluate treatments accurately.

There have been several attempts to improve the performance of labeling and MRF by enhancing their accuracy and speed for specific applications [51, 56, 60–64]. By considering the drawbacks of such methods we also provide two modifications of q-entropy, and MRF for brain image segmentation as follows.

3.2 Modified q-entropy Segmentation

The q-entropy segmentation algorithm (see section 2.2.4) uses the normalized histogram (Eq.(2.1)) as the PDF which is highly sensitive to noise and inhomogeneity because these factors skew the distributions towards one gray level value. To ensure a good estimation of two thresholds, we used a GMM to fit to the histogram and then use it as the PDF to calculate q-entropy. By using this GMM, it is possible to decrease both inhomogeneity and noise effects simultaneously. This modification of q-entropy segmentation is presented in Appendix A.6 as a new algorithm.

Estimation of the GMM parameters by EM is a time-consuming process. Therefore, we provided a fast solution instead. For a chosen pair of thresholds (t_1, t_2), we classify the intensities belonging to each of the three tissues (explained in section 2.2). By using all levels in each class, the mean (μ_i) and STD (σ_i) values can be calculated as well as the proportion of each tissue (π_i) which is the sum of the frequencies in each class divided by total frequency of the intensities. As in algorithm (A.4), no parameter other than q is required here, but the calculation time will be increased though it is negligible.

3.3 Modified MRF segmentation

To reduce the influence of wrong decisions that may happen in ICM (section 2.3) using K-mean, we fitted a local Gaussian PDF to the histogram intensity of each tissue in the neighborhood of each voxel v_i. Then, instead of calculating the distances between v_i and the means for each tissue (explained in K-mean algorithm section 2.2.2), the MAP of v_i, using the new local PDFs, as local weights are calculated. Because ICM tries to minimize the energy function, but the maximum weight is desired, we inverted the local weight values in order to be considered as the minimum. Since this approach uses both mean and standard deviation of each tissue distribution, it yields a more robust result than K-mean which uses only the mean values. This modification is implemented in the algorithm (A.7) in rhe Appendix.

In practice, when calculating the local weights, the STD for one or more local PDFs might be very small or equal to zero. This situation happens when the number of the voxels belonging to that tissue in the neighborhood (S_i) is small

(in the borders), or the intensities of the voxels are very similar. In these cases, a default value of 1 for STD helps to overcome this mathematical problem. Although calculation will be twofold for each neighborhood, it is not considered yet.

3.4 Modified q-entropy through Modified MRF

Investigating the source of mistakes which may happen in well-accepted segmentation algorithms, a hybrid pipeline with modified q-entropy and modified MRF (Mqe-MMRF) was proposed in order to mitigate the influence of such mistakes. Now a methodology is needed to evaluate Mqe-MMRF to present its robustness or even its weakness.

3.5 Methodology

We evaluate Mqe-MMRF in two stage; first by artifact challenge and the second in real MR brain image. To see the effects of noise and bias field on Mqe-MMRF performance, we used a simulated MR brain T1-weighted image (dimension=$181 \times 217 \times 181$ voxel, Voxel size=$1 \times 1 \times 1$ mm, TE=10 s, TR=18 s, FA=30°) from BrainWeb dataset [65] with seven percentages of artifacts (noise from 0 to 9 percent and bias effect from 0 to 40 percent). By performing Mqe-MMRF on the simulations, the produced label maps were evaluated by nine similarity metrics, i.e., Dice Coefficient (DICE), Jaccard Coefficient (JACCARD), Area under ROC Curve (AUC), Cohen Kappa (KAPPA), Rand Index (RNDIND), Adjusted Rand Index (ADJRIND), Interclass Correlation (ICCORR), Volumetric Similarity Coefficient (VOLSMTY) and Mutual Information (MUTINF) [66].

In order to compare all considered methods in this study, i.e., EM, K-means, q-entropy (qe) and modified q-entropy algorithms were considered as four labeling techniques (P_1 to P_4) performed on five training images from MRBrainS13 [59]. Then, we used MRF to refine the borders of the label map produced by P_1 to P_4, providing 4 new hybrid pipelines namely EM-MRF, K-MRF, qe-MRF, and Mqe-MRF respectively P_5 to P_8. Moreover, MMRF was also applied on P_1 to P_4 to provide other four pipelines namely EM-MMRF, K-MMRF, qe-MMRF and Mqe-MMRF (P_9 to P_{12}). We defined Improvement, which is the subtraction of

the similarity metric values for Mqe-MMRF from the mean of others to compare the pipelines. For example, for WM we calculated DICE value for 11 label maps produced by P_1 to P_{11}, and then used their mean value to be subtracted from the DICE value calculated for Mqe-MMRF, i.e.,

$$Improvment_{WM} = DICE^{WM}_{Mq-MMRF} - DICE^{WM}_{meanofP1-P11} \tag{3.1}$$

In the same way, Eq. (3.1) is applied to other subjects and tissues. By this definition for $Improvment$, a positive value shows a higher accuracy for Mqe-MMRF than the average of other eleven pipelines and a negative value *vice versa*. In order to show this $Improvement$ is significant, the paired sample t-test was performed (p-value 0.02) for each pipeline (P_i), individually, against Mqe-MMRF (P_{12}) by Null hypothesis

$$H_0 = meanP_i - meanP_{12} \geq 0 \tag{3.2}$$

against alternate hypothesis

$$H_1 = meanP_i - meanP_{12} < 0, for\ i = 1, 2, ..., 11 \tag{3.3}$$

In the classical and modified MRF, a weighted neighborhood (defined by a $3 \times 3 \times 3$ kernel or all voxels with $d \leq 2$), was used. For those pipelines where qe or Mqe were used in, we adopted a value of $q = -3.4$. This value was chosen empirically by comparing the label map of a segmented image using q-entropy with manual segmentation (see also [66]). The second evaluation step using the five training datasets [59] are provided with manual segmentation to use as training data to tune the segmentation algorithm. Manual segmentations used as ground truth were drawn on the thick-slice scans, i.e., 3 mm slice thickness, using a developed manual segmentation tool based on the contour segmentation objects (CSO) tool. A freehand spline drawing technique was used to segment all structures in the brain. The outline of each structure was delineated, starting with the deep structures. By iteratively subtracting delineations to create holes, binary images were created for each structure. Segmentations were performed in a darkened room with optimal viewing conditions. All segmentations were inspected for correctness by an expert not involved in the segmentation procedure, and corrections were made if needed. A third expert-approved all final segmentation. Evaluation using the remaining

fifteen scans are provided as test data. Only the scans are provided, not the manual segmentations. Authors can submit the segmentation results of their algorithms, after which the evaluation results will be sent to them by e-mail. The results are published on the website. One must submit the labels volumes following a standard format. Submitted segmentation results, i.e., label set, compared to the manually obtained reference standard. For each tissue label sets, i.e., GM, WM, and CSF, an automatic tool calculates the Dice coefficient (DC), the 95th-percentile of the Hausdorff distance (HD-95) and the absolute volume difference (AVD). The final ranking provided by MRBrainS13 challenge considers evaluation results of all 15 test datasets and is determined as follows: For each evaluation measure, i.e., DC, HD-95, and AVD, the mean value over all fifteen datasets is determined for WM, GM, and CSF. Each evaluated method receives a rank ranging up to the unit for each tissue type and each evaluation measure based on the mean value of the evaluation measures over the fifteen datasets. MRBrainS13 challenge determined the final score by adding the ranks of all tissue types and evaluation measures for each method.

3.6 Partial Volume Effect

In neurodegenerative clinical studies, patients are usually followed up for several years frequently examined in different MRI scanners and base magnetic fields. Technical specification differences in MRI scanners, e.g., acquisition protocols and spatial resolution, are two crucial limitations for longitudinal long-term cerebral atrophy investigations. Although widely known that MRI base magnetic field (B0) boldly affects brain tissue volume measurements, no systematic study has been proposed to maintain brain volume consistency through longitudinal exams. As the last part of the segmentation pipeline, we proposed a method to convert the measured volumes produced by brain segmentation methods with lower B0 and resolutions to volumes consistent with a higher resolution if the dataset includes different scanners.

The proposed partial volume transfer (PVT) method is consistent with partial volume effect to correct the volume measures. The PVT demands at least one pair of simultaneous acquisitions on both MRI scanners to estimate three unknown

coefficients completing the PVT system. We selected ten healthy subjects from a standard dataset on the web, ten other healthy subjects, and ten patients with multiple sclerosis diseases scanned by two different scanners, i.e., 1.5 T and 3.0 T from local hospital dataset, to evaluate the PVT. We obtained the mean relative errors ranging from 0.28 to 1.14 % of total brain volume using the PVT conversion for brain compartments, significantly less than interpolation methods which were more than 6.54%. More details about this PVT can be found in [67].

3.7 Conclusion

We investigated the source of mistakes where may occur in q-entropy and MRF. Then two modifications in order to decrease the effects of the mistakes were proposed. After, a methodology to evaluate the modifications was provided including noise effects challenge and comparison with well-accepted method as well as the methods before modifying. Finally, we suggest a conversion for those longitudinal studies in which the dataset includes two different sources of image acquisitions. This conversion can be used for any other segmentation methods. In the final chapter, our suggestion for brain image segmentation is evaluated and results are discussed.

Chapter 4

Results and Discussion

In this chapter, the proposed method, i.e., Mqe-MMRF in the third chapter is evaluated. The results of the evaluation are presented here. Also, the results of performing Mqe-MMRF can be seen online in MRBrains13 [59] global challenge website in the results page.

4.1 Introduction

A hybrid approach, i.e., Mqe-MMRF, to parcellate brain compartment in Chapter 3. was introduced. We underwent two strategies to evaluate Mqe-MMRF, i.e., a simulation of different levels of noise on MRI data, and a set of twenty MRI data available from MRBrainS13 [59] as brain tissue segmentation challenge. We accessed different segmentation quality metrics compared to reference tissues delineations to evaluate Mqe-MMRF. Two evaluation steps, assessed the proposed method, i.e., artifact effect and real subjects. In this chapter, the results after performing Mqe-MMRF in two steps are discussed in detail.

4.2 Artifact effect evaluation

The simulation images (S_N^{RF}), from the online dataset [65], by different percentages of noise (N) and bias field (RF) with their histogram based on the frequency of intensity levels (bins), are shown in Fig. 4.1.

As it can be seen the histogram of S_0^0 has lower noise than the other histograms. Fig. 4.2 shows the similarity metrics calculated for Mqe performed

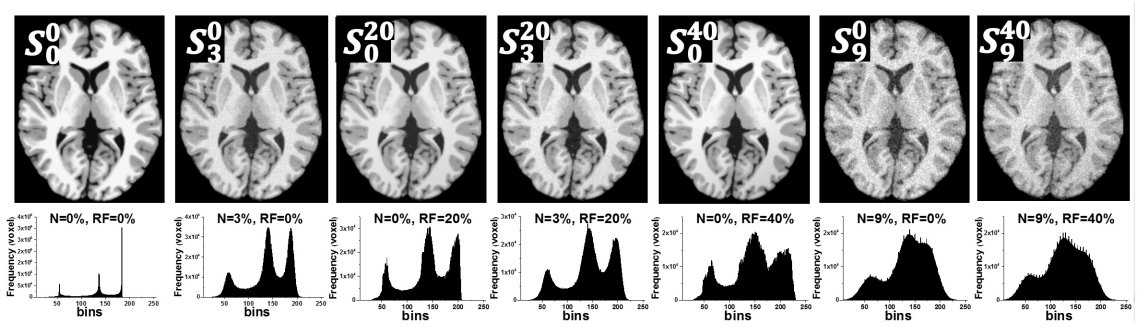


Figure 4.1: *One slice of each simulation, with its image histogram, indicated below. N=noise, RF= intensity non-uniformity.* S_0^0 *represents the simulation image with 0% noise and 0% intensity non-uniformity, and* S_3^0 *by 3% noise and 0%RF and so on.*

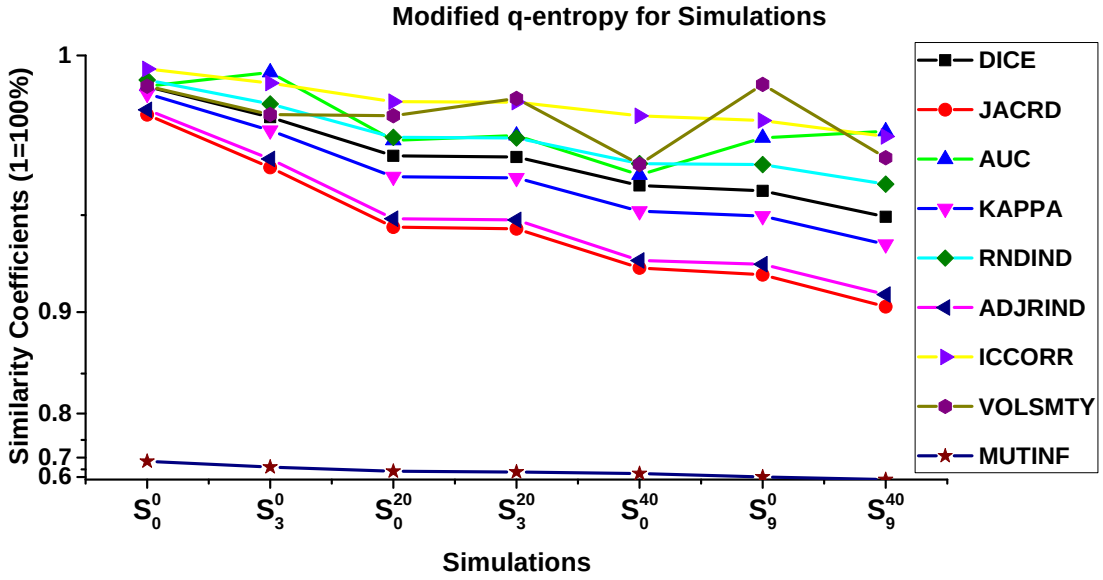


Figure 4.2: *Changes in similarity metrics when using Mqe by adding Noise and RF in simulations. Trends of similarity metrics show a very slight decrease (here y-axis scaled exponentially to show the difference between metrics, else all were impact together).*

on the simulations.

Reasonably, by adding artifacts, we expected a decreasing trend for similarity metrics. There are two exceptions (AUC and VOLSMTY), that presented an oscillatory behavior (increasing and then decreasing) in S_9^0. This disagreement shows that the bias field had more effect than noise on these similarities with respect to the other metrics. Decreasing trend for similarity metrics when the artifacts were added, averagely, was only 0.05% which shows the robustness of Mqe.

Next, we performed MMRF on the label maps produced by Mqe on the

Improvement for the simulations (%)								
Metrics	S_0^0	S_3^0	S_0^{20}	S_3^{20}	S_0^{40}	S_9^0	S_9^{40}	Mean
DICE	0.42	0.25	0.42	0.26	0.71	0.61	0.69	0.48
JACRD	0.83	0.48	0.80	0.50	1.32	1.14	1.25	0.90
AUC	0.63	0.02	0.43	0.20	0.71	0.30	0.24	0.36
KAPPA	0.52	0.31	0.53	0.33	0.89	0.76	0.84	0.60
RNDIND	0.33	0.19	0.34	0.21	0.58	0.46	0.49	0.37
ADJRIND	0.76	0.44	0.78	0.47	1.30	1.07	1.17	0.85
ICCORR	0.18	0.11	0.19	0.11	0.30	0.27	0.32	0.21
VOLSMTY	0.55	0.35	0.48	0.11	0.83	0.23	0.48	0.43
MUTINF	1.36	0.63	0.78	0.59	1.09	1.42	1.48	1.05

Table 4.1: *Improvements of similarity metrics after using MMRF on the label map produced by Mqe segmentation method.*

simulations. The results in Tab. 4.1 show the improvements for the similarity metrics, averagely, from 0.21 to 1.05%.

4.3 Training subjects evaluation

The five training MRIs, namely C1 to C5 were used to be segmented by all the pipelines. In Fig. 4.3, the same parts of an axial slice of each produced label map by all pipelines are represented. The initial segmentation, refined by MRF and MMRF, are shown in the first, second and third row, respectively. The manual segmentation, which is used as a gold standard reference, and the real image can be seen on the right side of the figure. Visually, each segmentation method was capable of achieving an acceptable result, but there are still some small differences in the borders between CSF and GM and WM as well as the connectivity of CSF. These differences may affect the interpretation of the final result, especially, if the final result in question involves a low rate of volume change quantification.

For the five training subjects, by considering each brain compartment as well as their combinations, we compared the similarity metrics calculated for Mqe-MMRF

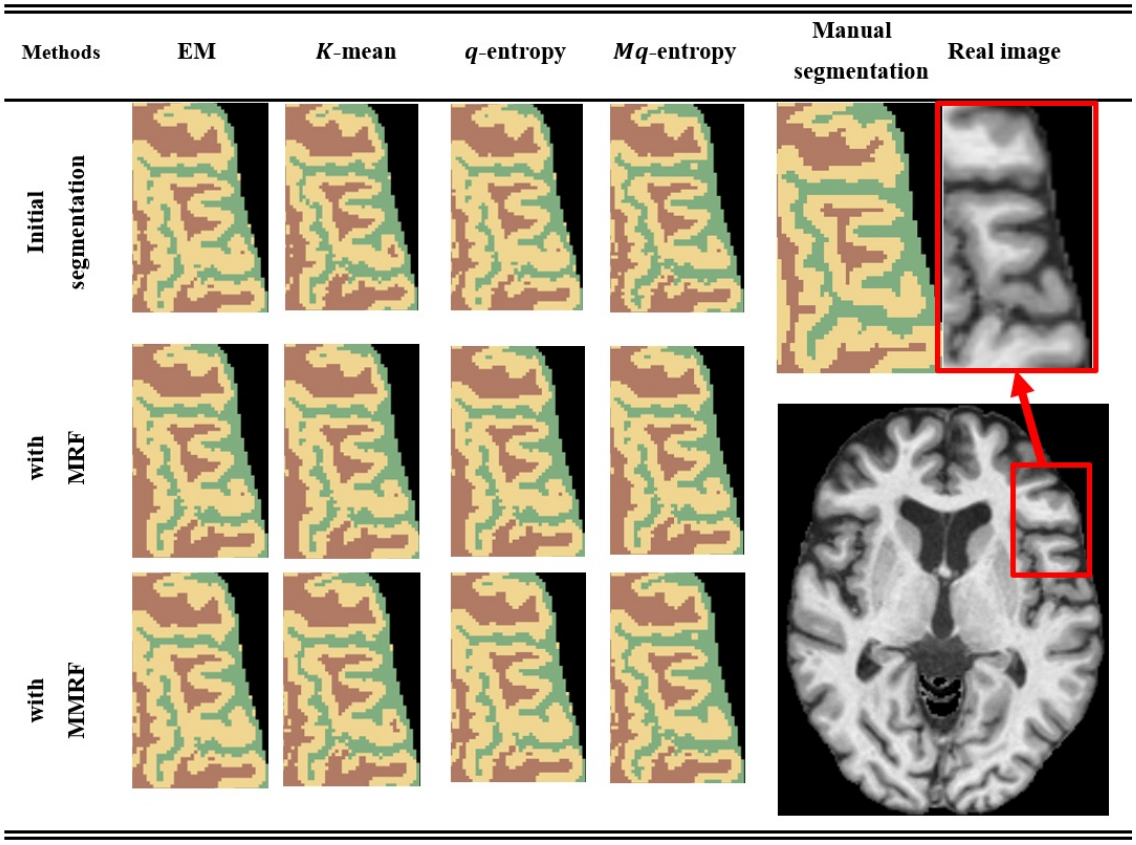


Figure 4.3: *The same part of the label maps extracted by four labeling techniques in the first row, post processed by MRF in the second row and by MMRF in the third row. Manual segmentation is shown in the right. The green color is CSF, yellow is GM and brown is WM. The borders of CSF in label maps and connectivity of CSF are two examples of mislabeling which can be used to compare these methods visually.*

with the other 11 methods by the Improvement definition (see Eq. 3.1). Fig. 4.4, presents the box chart including the minimum, maximum and mean Improvements of the similarity metrics, averagely, for these training subjects.

The mean Improvements of the nine similarities considering each compartment and together on C1 to C5, for Mqe-MMRF, are presented in Tab. 4.2. As it can be seen in this table, for 41 out of 45 of the means, Improvements are positive, i.e., Mqe-MMRF had a higher rate than the average of other 11 pipelines. Only for 3 situations in AUC and one in MUINF, Improvements were not positive, i.e., Mqe-MMRF could not increase these similarity metrics in those tissues.

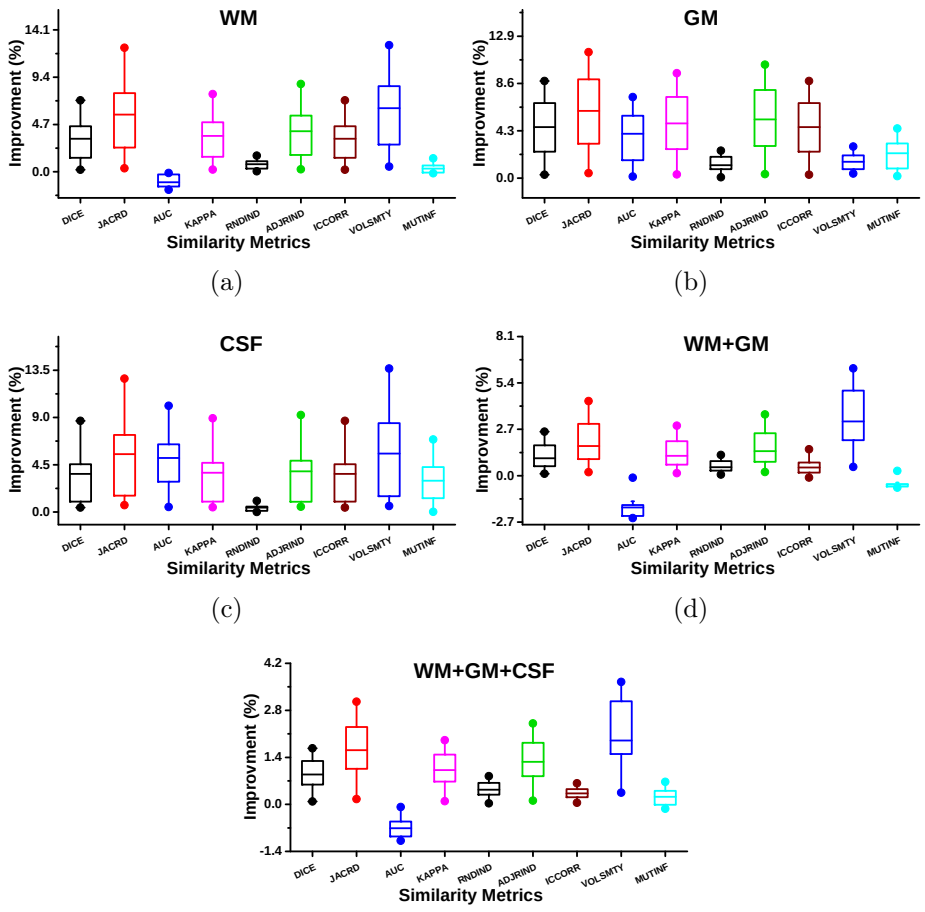


Figure 4.4: *Improvement in similarity metrics in brain tissues for Mqe-MMRF. (a) White Matter, (b) Gray Matter, (c) Cerebrospinal Fluid (d) Brain and (e) all three tissues. Error bars show the minimum and the maximum improvements of each similarity metric averagely for five training subjects.*

Mean Improvements of Similarity Metrics for C1 to C5 (%)									
Brain tissues	DICE	JACRD	AUC	KAPPA	RNDIND	ADJRIND	ICCORR	VOLSMTY	MUTINF
GM	4.64	6.11	4.02	4.97	1.17	5.33	4.64	1.49	2.26
WM	3.28	5.68	-1.04	3.56	0.75	4.03	3.28	6.32	0.32
CSF	3.64	5.52	5.14	3.74	0.43	3.87	3.64	5.57	2.98
WM+GM	1.01	1.73	-1.85	1.15	0.50	1.43	0.47	3.16	-0.50
WM+GM+CSF	0.89	1.61	-0.71	1.02	0.44	1.27	0.33	1.90	0.22

Table 4.2: *Average improvement of similarity metrics for Mqe-MMRF in comparison to mean of other methods for each tissue and tissue combination for five training subjects C1 to C5.*

Similarity Metrics						
	DC (%)		HD (mm)		AVD (%)	
Structure	Mean	STD	Mean	STD	Mean	STD
GM	80.09	2.25	3.93	1.32	4.87	4.82
WM	86.76	2.43	3.33	1.08	6.92	5.46
CSF	68.03	3.24	4.24	0.86	15.97	8.97
WM+GM	93.66	0.85	4.26	1.47	3.35	2.69
WM+GM+CSF	94.03	1.18	6.09	1.17	6.47	3.12

Table 4.3: *The means and standard deviations (STD) of three similarity metrics DC=DICE, HD=Hausdorff Distance and AVD= Absolute Volume Difference for Mqe-MMRF on 15 test images received from MRBrainS13 challenge.*

4.4 Test subjects evaluation

After performing the pipelines on five training images, Mqe-MMRF was chosen to segment 15 test images, and the produced label maps were submitted to MRBrainS13 challenge. In Tab. 4.3, three similarity metrics (averagely for 15 test images), i.e., DICE, HD and AVD measured by the automatic tool used in MRBrainS13 are presented. For AVD similarity, Mqe-MMRF ranked as the first among 61 methods in the submission date.

The results on the first attempt for global challenge (Tab. 4.3), revealed the ROBEX brain extraction method had a lower precision in CSF detection. Then, the outputs of the brain extraction, i.e., the brains were refined manually on the borders and the new label maps were submitted as the second attempt. The results showed that the similarities for CSF increased, considerably, as can be seen in Tab. 4.4.

As the final evaluation of Mqe-MMRF, it was compared with three references method, i.e., SPM, FSL and FreeSurfer by the scores and ranking provided by MRBrainS13 where the results tabulated in Tab. 4.5 .

The results showed the mean similarity metrics of Mqe-MMRF was significantly more than each pipeline using the t-test at the significance level (p-value < 0.02), individually, as well as the mean similarity metrics of all pipelines together. Also, we used the same t-test for the variances of similarity metrics results showed that

Similarity Metrics						
	DC (%)		HD (mm)		AVD (%)	
Structure	Mean	STD	Mean	STD	Mean	STD
GM	80.18	2.24	3.87	1.22	5.01	4.89
WM	86.80	2.38	3.20	0.71	6.82	5.33
CSF	75.11	3.27	3.00	0.11	9.05	6.17
WM+GM	93.73	0.76	4.10	1.06	3.22	2.77
WM+GM+CSF	95.87	0.32	5.24	1.16	3.21	1.81

Table 4.4: *The means and standard deviations (STD) of three similarity metrics DC=DICE, HD=Hausdorff Distance and AVD= Absolute Volume Difference for Mqe-MMRF on 15 test images received from the MRBrainS13 challenge after CSF border correction.*

	Scores				
Tools	GM	WM	CSF	GM+WM	Total Rank
Mqe-MMRF	105	122	151	132	1
SPM12 T1	130	107	161	131	2
FSL Seg	159	154	131	131	3
FreeSurfer	148	134	172	153	4

Table 4.5: *Score = Rank DC+Rank HD+Rank AVD, DC=DICE, HD=Hausdorff Distance and AVD= Absolute Volume Difference calculated for Mqe-MMRF and referenced methods on 15 test images.*

they are not significantly different.

4.5 Discussion

Averagely, Mqe-MMRF significantly improved the similarity metrics calculated for its label maps when compared to other pipelines. However, it should be noted that, in a few cases, the difference between the precision of Mqe-MMRF is negligible in comparison to the mean of others because the average of similarity metrics for all 11 pipelines still contains either Mqe or MMRF, i.e., our modifications are used in. Therefore, if one removes those modifications in the pipelines, the improvements will increase even more.

In a few cases, either MRF or MMRF reduced the similarity metrics. These

undesired consequences can be explained by how MRF works. For relabeling a voxel, it is assumed that the neighbors of that voxel are already labeled correctly. In practical applications, however, this condition is not necessarily met, and the mistakes of the initial labeling can also be propagated to the results. Meanwhile, in addition to K-mean mistakes in MRF, the local Gaussian PDFs used in MMRF may not also fit the histograms perfectly, resulting in another source of decrease in similarity metrics.

Although the results showed that Mqe-MMRF for five training subjects has no better result than the average of other methods for WM, WM+GM and total in AUC and MUTINF (Tab. 4.2, negative improvements), it should be kept in mind that these metrics are commonly used for segmentation method in brain tumor [68], diffusion tensor imaging [69] and also breast lesions [70] in dynamic contrast-enhanced MR images. Generally, common validation metrics used in brain segmentation methods [38, 71–74] are DICE, JACCARD and VOLSMTY where Mqe-MMRF had satisfactory results in them. More details are in supplementary data in Appendix B.2.

Performing q-entropy labeling on one of the training subjects (C3) was not successful due to the wrong extracted threshold which could not separate tissues correctly. The first threshold t_1 was between two values where the picks of CSF and GM were in the histogram, but t_2 was after WM peak. Therefore, WM and GM labeled as one tissue. While the suggested strategy to modify q-entropy segmentation not only increased the accuracy of segmentation, but its performance on C3 was successful.

In some specific clinical analysis of the brain, not all tissues may be under investigation. For example, there are many studies which are only interested in examining the changes in GM [75–77], because GM includes regions of the brain involved in muscle control, and sensory perception such as seeing and hearing, memory, emotions, speech, decision making, and self-control [78]. Considering the improvement of similarity metrics in GM, it confirms the significant difference between the proposed Mqe-MMRF pipeline with the other methods. In addition, Mqe-MMRF was applied on MRBrainS13 challenge for 15 real subjects without using any pre-processing technique except brain extraction. Since CSF is the outer

layer, it may not be entirely detected by brain extraction techniques and some parts are lost. As it can be seen in Tab. 3, the minimum score is associated with CSF. There are three chances for resubmission the new label maps for MRBranS13 in which we can change brain extraction to improve the ranking in the other tissues.

CPU times for the implemented pipelines in this study were roughly the same. These CPU usage times were averagely 3 to 6 minutes by using a Core i7 (4 cores, 8 threads, 3.20 GHz) and 8 Gigabyte RAM. By using stronger specifications, such as dedicated processing servers, computation time will decrease considerably. Therefore, the CPU time comparison will not show a significant difference between pipelines.

T1-weighted images are generally used for volumetric analysis, although, in some special cases such as lesion detection, T2-weighted images might also be used. The intensity distribution for GM, WM, and CSF in T2 weighted images is very different from T1 weighted images. For instance, the mean intensity order for each tissue, and the distribution of WM and GM intensity are different between image modalities. Testing the performance of our proposed methods in other modalities, such as T2 or other modalities is still necessary. Another application of our suggestion could be in lesion detection. These two aims will be investigated as the next step.

4.6 Conclusion

MRI noise simulations showed a slow decreasing for segmentation scores metrics showing the robustness of Mqe-MMRF. For five training subjects, we found significant improvements in similarity metrics, i.e., for the whole brain, White Matter, Gray Matter and Cerebrospinal Fluid (p-values < 0.02) in quality metrics when Mqe-MMRF is compared to the other state of the art methods. The Mqe-MMRF was performed on 15 other real subjects in MRBrainS13 online challenge, and the results held the first rank among more than 60 methods for volume measurement precision similarity in Gray matter.

4.7 Future steps

Our propose for brain image segmentation, i.e., Mqe-MMRF showed a high precision in parceling the brain compartments, specifically, in GM. Since several diseases, e.g., Multiple Sclerosis and Alzheimer, demyelinate nerve cells mostly in GM regions, a longitudinal study to discover the correlation between changes in GM and medical symptoms and treatment plans will be investigated.

Bibliography*

[1] KONISHI, S.; KITAGAWA, G. *Information criteria and statistical modeling*. [S.l.]: Springer Science & Business Media, 2008.

[2] BULMER, M. G. *Principles of statistics*. [S.l.]: Courier Corporation, 1979.

[3] MENDENHALL, W.; WACKERLY, D.; SCHEAFFER, R. *Mathematical Statistics with Applications PWS*. [S.l.]: KENT Publishing Company, USA, 1990.

[4] MILLER, I.; MILLER, M. *John E. Freund's mathematical statistics with applications*. [S.l.]: Pearson, 2015.

[5] MCCULLAGH, P. What is a statistical model? *Annals of statistics*, JSTOR, p. 1225–1267, 2002.

[6] HAZELTON, M. L. Methods of moments estimation. In: *International Encyclopedia of Statistical Science*. [S.l.]: Springer, 2011. p. 816–817.

[7] SCHOLZ, F. Maximum likelihood estimation. *Encyclopedia of statistical sciences*, Wiley Online Library, 1985.

[8] SHEIKHRABORI, R.; AMINNAYERI, M.; AYOUBI, M. Maximum likelihood estimation of change point from stationary to nonstationary in autoregressive models using dynamic linear model. *Quality and Reliability Engineering International*, Wiley Online Library.

[9] WANG, S. Two-stage maximum likelihood estimation in the misspecified restricted latent class model. *British Journal of Mathematical and Statistical Psychology*, Wiley Online Library, 2017.

*De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

[10] SNOKE, J. et al. Providing accurate models across private partitioned data: Secure maximum likelihood estimation. *arXiv preprint arXiv:1710.06933*, 2017.

[11] BERTL, J. et al. Approximate maximum likelihood estimation for population genetic inference. *Statistical Applications in Genetics and Molecular Biology*, De Gruyter, 2017.

[12] LIMA, V. M.; CRIBARI-NETO, F. Penalized maximum likelihood estimation in the modified extended weibull distribution. *Communications in Statistics-Simulation and Computation*, Taylor & Francis, p. 1–16, 2017.

[13] FERGUSSON, K. Explicit formulae for parameters of stochastic models of a discounted equity index using maximum likelihood estimation with applications. *Annals of Financial Economics*, World Scientific, v. 12, n. 02, p. 1750010, 2017.

[14] WOODBURY, M. A. *A missing information principle: theory and applications*. [S.l.], 1970.

[15] DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, JSTOR, p. 1–38, 1977.

[16] MCLACHLAN, G.; KRISHNAN, T. *The EM algorithm and extensions*. [S.l.]: John Wiley & Sons, 2007.

[17] LLOYD, S. Least square quantization in pcm. bell telephone laboratories paper. published in journal much later: Lloyd, sp: Least squares quantization in pcm. *IEEE Trans. Inform. Theor.(1957/1982) Google Scholar*, 1957.

[18] FORGEY, E. Cluster analysis of multivariate data: Efficiency vs. interpretability of classification. *Biometrics*, Oxford, v. 21, n. 3, p. 768–769, 1965.

[19] FRIEDMAN, H. P.; RUBIN, J. On some invariant criteria for grouping data. *Journal of the American Statistical Association*, Taylor & Francis, v. 62, n. 320, p. 1159–1178, 1967.

[20] JAIN, A. K.; DUBES, R. C. Algorithms for clustering data. Prentice-Hall, Inc., 1988.

[21] KAPUR, J. N.; SAHOO, P. K.; WONG, A. K. A new method for gray-level picture thresholding using the entropy of the histogram. *Computer vision, graphics, and image processing*, Elsevier, v. 29, n. 3, p. 273–285, 1985.

[22] ALBUQUERQUE, M. P. D.; ESQUEF, I. A.; MELLO, A. G. Image thresholding using tsallis entropy. *Pattern Recognition Letters*, Elsevier, v. 25, n. 9, p. 1059–1065, 2004.

[23] TSALLIS, C. et al. Nonextensive statistical mechanics and economics. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 324, n. 1-2, p. 89–100, 2003.

[24] HELD, K. et al. Markov random field segmentation of brain mr images. *IEEE transactions on medical imaging*, IEEE, v. 16, n. 6, p. 878–886, 1997.

[25] DENG, H.; CLAUSI, D. A. Unsupervised image segmentation using a simple mrf model with a new implementation scheme. *Pattern recognition*, Elsevier, v. 37, n. 12, p. 2323–2335, 2004.

[26] POGGIO, T.; GAMBLE, E. B.; LITTLE, J. Parallel integration of vision modules. *Science*, American Association for the Advancement of Science, v. 242, n. 4877, p. 436–440, 1988.

[27] Sridevi, M. Image segmentation based on multilevel thresholding using firefly algorithm. In: *2017 International Conference on Inventive Computing and Informatics (ICICI)*. [S.l.: s.n.], 2017. p. 750–753.

[28] LINDA, G.; SHAPIRO, C. G. *Stockman, Computer vision*. [S.l.]: Prentice Hall, Upper Saddle River, NJ, 2001.

[29] DASS, R.; DEVI, S. Image segmentation techniques 1. Citeseer, 2012.

[30] DISORDERS, N. I. of N.; (NINDS), S. *Cerebral Atrophy Information Page*. 2018. `https://www.ninds.nih.gov/Disorders/All-Disorders/Cerebral-Atrophy-Information-Page/`. [Online; accessed 10-April-2018].

[31] KIKINIS, R. et al. Routine quantitative analysis of brain and cerebrospinal fluid spaces with mr imaging. *Journal of Magnetic Resonance Imaging*, Wiley Online Library, v. 2, n. 6, p. 619–629, 1992.

[32] LIM, K. O.; PFEFFERBAUM, A. Segmentation of mr brain images into cerebrospinal fluid spaces, white and gray matter. *Journal of Computer Assisted Tomography*, v. 13, n. 4, p. 588–593, 1989.

[33] SMITH, S. M. et al. Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage*, Elsevier, v. 23, p. S208–S219, 2004.

[34] AVANTS, B. B. et al. An open source multivariate framework for n-tissue segmentation with evaluation on public data. *Neuroinformatics*, Springer, v. 9, n. 4, p. 381–400, 2011.

[35] PIEPER, S.; HALLE, M.; KIKINIS, R. 3d slicer. In: IEEE. *Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on*. [S.l.], 2004. p. 632–635.

[36] RASBAND, W. Imagej: Image processing and analysis in java. *Astrophysics Source Code Library*, 2012.

[37] FISCHL, B. Freesurfer. *Neuroimage*, Elsevier, v. 62, n. 2, p. 774–781, 2012.

[38] POHL, K. M. et al. A hierarchical algorithm for mr brain image parcellation. *IEEE transactions on medical imaging*, IEEE, v. 26, n. 9, p. 1201–1212, 2007.

[39] MILLER, M. I. et al. Mathematical textbook of deformable neuroanatomies. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 90, n. 24, p. 11944–11948, 1993.

[40] SRIVASTAVA, S. et al. Effects of anatomical asymmetry in spatial priors on model-based segmentation of the brain mri: A validation study. In: SPRINGER. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. [S.l.], 2004. p. 327–334.

[41] COLLINS, D. L. et al. Animal+ insect: improved cortical structure segmentation. In: SPRINGER. *Biennial International Conference on Information Processing in Medical Imaging*. [S.l.], 1999. p. 210–223.

[42] WOODS, R. P. et al. Automated image registration: I. general methods and intrasubject, intramodality validation. *Journal of computer assisted tomography*, LWW, v. 22, n. 1, p. 139–152, 1998.

[43] IGLESIAS, J. E. et al. Robust brain extraction across datasets and comparison with publicly available methods. *IEEE transactions on medical imaging*, IEEE, v. 30, n. 9, p. 1617–1634, 2011.

[44] GERIG, G. et al. Nonlinear anisotropic filtering of mri data. *IEEE Transactions on medical imaging*, IEEE, v. 11, n. 2, p. 221–232, 1992.

[45] TUSTISON, N. J. et al. N4itk: improved n3 bias correction. *IEEE transactions on medical imaging*, NIH Public Access, v. 29, n. 6, p. 1310, 2010.

[46] LEEMPUT, K. V. et al. Automated model-based tissue classification of mr images of the brain. *IEEE transactions on medical imaging*, IEEE, v. 18, n. 10, p. 897–908, 1999.

[47] WELLS, W. M. et al. Adaptive segmentation of mri data. *IEEE transactions on medical imaging*, IEEE, v. 15, n. 4, p. 429–442, 1996.

[48] POHL, K. M. et al. A bayesian model for joint segmentation and registration. *NeuroImage*, Elsevier, v. 31, n. 1, p. 228–239, 2006.

[49] SMITH, S. M. Fast robust automated brain extraction. *Human brain mapping*, Wiley Online Library, v. 17, n. 3, p. 143–155, 2002.

[50] ESKILDSEN, S. F. et al. Beast: brain extraction based on nonlocal segmentation technique. *NeuroImage*, Elsevier, v. 59, n. 3, p. 2362–2373, 2012.

[51] HARTIGAN, J. A. Clustering algorithms, new york: John willey and sons. *Inc. Pages113129*, 1975.

[52] DINIZ, P. et al. Brain tissue segmentation using q-entropy in multiple sclerosis magnetic resonance images. *Brazilian Journal of Medical and Biological Research*, SciELO Brasil, v. 43, n. 1, p. 77–84, 2010.

[53] GEMAN, S.; GEMAN, D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. In: *Readings in Computer Vision*. [S.l.]: Elsevier, 1987. p. 564–584.

[54] MANJUNATH, B.; CHELLAPPA, R. Unsupervised texture segmentation using markov random field models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 13, n. 5, p. 478–482, 1991.

[55] KRISHNAMACHARI, S.; CHELLAPPA, R. Multiresolution gauss-markov random field models for texture segmentation. *IEEE Transactions on image processing*, IEEE, v. 6, n. 2, p. 251–267, 1997.

[56] ZHANG, Y.; BRADY, M.; SMITH, S. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE transactions on medical imaging*, IEEE, v. 20, n. 1, p. 45–57, 2001.

[57] WELLS, W. et al. Adaptive segmentation of mri data. In: SPRINGER. *Computer Vision, Virtual Reality and Robotics in Medicine*. [S.l.], 1995. p. 59–69.

[58] BESAG, J. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society: Series B (Methodological)*, Wiley Online Library, v. 48, n. 3, p. 259–279, 1986.

[59] MENDRIK, A. M. et al. Mrbrains challenge: online evaluation framework for brain image segmentation in 3t mri scans. *Computational intelligence and neuroscience*, Hindawi Publishing Corp., v. 2015, p. 1, 2015.

[60] KUMAR, R.; MATHAI, K. J. Brain tumor segmentation by modified k-mean with morphological operations. *Brain*, v. 6, n. 8, 2017.

[61] KUBICEK, J. et al. Detection and dynamical tracking of temperature facial distribution caused by alcohol intoxication with using of modified otsu regional

segmentation. In: SPRINGER. *Asian Conference on Intelligent Information and Database Systems*. [S.l.], 2018. p. 357–366.

[62] PARE, S. et al. A new technique for multilevel color image thresholding based on modified fuzzy entropy and lévy flight firefly algorithm. *Computers & Electrical Engineering*, Elsevier, v. 70, p. 476–495, 2018.

[63] NEMCEV, N.; GILMUTDINOV, M. Modified em-algorithm for motion field refinement in motion compensated frame interpoliation. In: *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*. [S.l.]: Springer, 2017. p. 662–670.

[64] GONG, M. et al. Fuzzy clustering with a modified mrf energy function for change detection in synthetic aperture radar images. *IEEE Transactions on Fuzzy Systems*, IEEE, v. 22, n. 1, p. 98–109, 2014.

[65] COCOSCO, C. A. et al. Brainweb: Online interface to a 3d mri simulated brain database. In: CITESEER. *NeuroImage*. [S.l.], 1997.

[66] TAHA, A. A.; HANBURY, A. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC medical imaging*, BioMed Central, v. 15, n. 1, p. 29, 2015.

[67] AZIMBAGIRAD, M. et al. Partial volume transfer (pvt) conversion of cerebral tissue volumes between different magnetic fields mri. *Research on Biomedical Engineering*, v. 35, n. 1, p. 11–20, Mar 2019. ISSN 2446-4740. Disponível em: <https://doi.org/10.1007/s42600-019-00003-4>.

[68] ZOU, K. H. et al. Three validation metrics for automated probabilistic image segmentation of brain tumours. *Statistics in medicine*, Wiley Online Library, v. 23, n. 8, p. 1259–1282, 2004.

[69] MOBERTS, B.; VILANOVA, A.; WIJK, J. J. V. Evaluation of fiber clustering methods for diffusion tensor imaging. In: IEEE. *VIS 05. IEEE Visualization, 2005.* [S.l.], 2005. p. 65–72.

[70] PANG, Y. et al. Computerized segmentation and characterization of breast lesions in dynamic contrast-enhanced mr images using fuzzy c-means clustering and snake algorithm. *Computational and mathematical methods in medicine*, Hindawi, v. 2012, 2012.

[71] STORELLI, L. et al. A semiautomatic method for multiple sclerosis lesion segmentation on dual-echo mr imaging: Application in a multicenter context. *American Journal of Neuroradiology*, Am Soc Neuroradiology, v. 37, n. 11, p. 2043–2049, 2016.

[72] EGGER, C. et al. Mri flair lesion segmentation in multiple sclerosis: Does automated segmentation hold up with manual annotation? *NeuroImage: Clinical*, Elsevier, v. 13, p. 264–270, 2017.

[73] BOESEN, K. et al. Quantitative comparison of four brain extraction algorithms. *NeuroImage*, Elsevier, v. 22, n. 3, p. 1255–1261, 2004.

[74] SCHMIDT, P. et al. An automated tool for detection of flair-hyperintense white-matter lesions in multiple sclerosis. *Neuroimage*, Elsevier, v. 59, n. 4, p. 3774–3783, 2012.

[75] FISNIKU, L. K. et al. Gray matter atrophy is related to long-term disability in multiple sclerosis. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, Wiley Online Library, v. 64, n. 3, p. 247–254, 2008.

[76] RUDICK, R. A. et al. Gray matter atrophy correlates with ms disability progression measured with msfc but not edss. *Journal of the neurological sciences*, Elsevier, v. 282, n. 1-2, p. 106–111, 2009.

[77] FRANKLIN, T. R. et al. Decreased gray matter concentration in the insular, orbitofrontal, cingulate, and temporal cortices of cocaine patients. *Biological psychiatry*, Elsevier, v. 51, n. 2, p. 134–142, 2002.

[78] MILLER, A.; ALSTON, R.; CORSELLIS, J. Variation with age in the volumes of grey and white matter in the cerebral hemispheres of man: measurements

with an image analyser. *Neuropathology and applied neurobiology*, Wiley Online Library, v. 6, n. 2, p. 119–132, 1980.

Appendix A

Algorithms

A.1 EM Brain Segmentation

Algorithm 1 :EM Brain Segmentation

Input: *A Brain Image* and Parameters:(μ_1, μ_2, μ_3), $(\sigma_1^2, \sigma_2^2, \sigma_3^3)$, (π_1, π_2, π_3)
  while *error* > *epsilon* **do**
    E-Step: Calculate W_{ij} according to eq.(1.28)
    M-Step: Update Parameters:μ_i, σ_i, π_i according to eq.(1.31) , eq.(1.32), eq.(1.33)
  end while
  for all voxels in Image **do**
    Chose a label (1 or 2 or 3) for current voxel based on its weights
  end for
Output: *Label Map*

A.2 K-mean Brain Segmentation

Algorithm 2 :K-mean Brain Segmentation

Input: *A Brain Image* and Parameters:(μ_1, μ_2, μ_3)
  while *error* > *epsilon* **do**
    for all voxels in Image **do**
      Chose a label (1 or 2 or 3) for current voxel based on minimum Euclidean distance
    end for
    Update $\mu_i,\ i = 1, 2, 3$ according to new label map
  end while
Output: *Label Map*

A.3 Entropy Brain Segmentation

Algorithm 3 :entropy Brain Segmentation

Input: *A Brain Image*
Calculate the Normalized Histogram of the image
Solve subproblem (2.4) and find $\tilde{t}$ and $\tilde{k}$
for all voxels in Image **do**
Chose a label (1 or 2 or 3) for current voxel based on $\tilde{t}$ and $\tilde{k}$
end for
Output: *Label Map*

A.4 q-Entropy Brain Segmentation

Algorithm 4 :q-entropy Brain Segmentation

Input: *A Brain Image*
Calculate the Normalized Histogram of the image
Solve subproblem (2.6) and find $\tilde{t}$, $\tilde{k}$ and q
for all voxels in Image **do**
Chose a label (1 or 2 or 3) for current voxel based on $\tilde{t}$ and $\tilde{k}$
end for
Output: *Label Map*

A.5 MRF label refining

Algorithm 5 :MRF using ICM

Input: *Brain Image*(I), *labelmap*, parameter α
Using *labelmap* Calculate the parameters of PDFs:(μ_i, σ_i), i=1,2,3
Solve subproblem (2.6) and find $\tilde{t}$, $\tilde{k}$ and q
for all voxels in Image **do**
Chose a label (1 or 2 or 3) for current voxel based on $\tilde{t}$ and $\tilde{k}$
end for
Output: *Label Map*

A.6 Modified q-entropy segmentation

Algorithm 6 :Modified q-entropy Brain Segmentation

Input: *Brain Image*(I) , q
  Calculate the Normalized Histogram of the image
  Solve subproblem (2.6) and find $\tilde{t}$, $\tilde{k}$ and q
  for all voxels in Image **do**
    Chose a label (1 or 2 or 3) for current voxel based on $\tilde{t}$ and $\tilde{k}$
  end for
Output: *Label Map*

A.7 Modified MRF label refining

Algorithm 7 :Modified MRF

Input: *Brain Image*(I), *labelmap*, parameter α
  Using *labelmap* Calculate the parameters of PDFs:(μ_i, σ_i), i=1,2,3
  Solve subproblem (2.6) and find $\tilde{t}$, $\tilde{k}$ and q
  for all voxels in Image **do**
    Chose a label (1 or 2 or 3) for current voxel based on $\tilde{t}$ and $\tilde{k}$
  end for
Output: *Label Map*

Appendix B

Supplementary Data

B.1 Similarity Metrics calculated for simulation dataset

	Simulations with different percentage of Noise and RF (%)						
Metrics	N=0 RF=0	N=3 RF=0	N=0 RF=20	N=3 RF=20	N=0 RF=40	N=9 RF=0	N=9 RF=40
DICE	99.2	98.4	97.2	97.1	96.1	95.9	94.9
JACRD	98.4	96.8	94.5	94.4	92.6	92.2	90.3
AUC	99.2	99.6	97.7	97.8	96.5	97.7	97.9
KAPPA	99.0	98.0	96.5	96.4	95.2	95.0	93.7
RNDIND	99.4	98.7	97.8	97.7	96.9	96.9	96.2
ADJRIND	98.6	97.1	94.9	94.8	93.0	92.8	91.1
ICCORR	99.7	99.3	98.8	98.8	98.4	98.3	97.8
VOLSMTY	99.2	98.4	98.4	98.9	96.9	99.3	97.1
MUTINF	68.7	66.2	63.9	63.5	62.5	60.1	57.8

Table B.6: *Calculated metrics for the label maps produced by Modified q-entropy (Mqe) on the simulation dataset. N=noise, RF= intensity non-uniformity*

	Simulations with different percentage of Noise and RF (%)						
Metrics	N=0 RF=0	N=3 RF=0	N=0 RF=20	N=3 RF=20	N=0 RF=40	N=9 RF=0	N=9 RF=40
DICE	99.6	98.6	97.6	97.4	96.9	96.6	95.6
JACRD	99.3	97.3	95.3	94.9	93.9	93.3	91.6
AUC	99.9	99.6	98.1	98.0	97.2	98.0	98.2
KAPPA	99.5	98.3	97.0	96.7	96.1	95.7	94.6
RNDIND	99.7	98.9	98.1	97.9	97.5	97.3	96.7
ADJRIND	99.3	97.5	95.6	95.3	94.3	93.8	92.3
ICCORR	99.8	99.4	99.0	98.9	98.7	98.5	98.1
VOLSMTY	99.8	98.8	98.9	99.0	97.7	99.5	97.6
MUTINF	70.0	66.8	64.7	64.1	63.6	61.6	59.3

Table B.7: *Calculated metrics for the label maps produced by Modified q-entropy (Mqe) and refined by MMRF on the simulation dataset. N=noise, RF= intensity non-uniformity*

B.2 Similarity Metrics calculated for 5 training dataset

In all the following tables, P1=EM, P2=K-mean, P3=q-entropy, P4= Modified q-entropy, P5= EM-MRF, P6=K-mean-MRF, P7=q-entropy-MRF, P8=Modified q-entropy-MRF, P9=EM-Modified MRF, P10=K-mean-Modified MRF, P11= q-entropy-Modified MRF and P12=Modified q-entropy-Modified MRF. Calculated metrics shows the precision of the pipelines for WM, GM, CSF, WM+GM, WM+GM+CSF detection using the label maps produced by the pipelines on training subjects C1 to C5 as follows

	The precision of all pipelines in WM detection (%)											
	Metrics	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
	DICE	86.5	83.5	86.1	87.4	84.2	83.6	83.8	84.8	86.5	84.4	86.3	87.4
	JACRD	76.3	71.6	75.6	77.6	72.8	71.9	72.1	73.6	76.3	73.0	75.9	77.6
	AUC	96.1	97.4	96.5	91.9	97.1	97.3	97.3	96.9	96.2	97.2	96.3	91.9
	KAPPA	85.8	82.6	85.4	86.7	83.4	82.8	82.9	84.0	85.8	83.6	85.6	86.7
C1	RNDIND	97.4	96.6	97.3	97.7	96.8	96.7	96.7	97.0	97.4	96.9	97.4	97.7
	ADJRIND	84.6	81.0	84.1	85.6	81.9	81.2	81.4	82.6	84.6	82.1	84.4	85.6
	ICCORR	86.5	83.5	86.1	87.4	84.2	83.6	83.8	84.8	86.5	84.4	86.3	87.4
	VOLSMTY	92.7	86.5	91.5	96.4	88.0	86.9	87.2	89.0	92.7	88.1	92.1	96.4
	MUTINF	19.9	19.9	19.9	19.6	19.9	19.9	19.9	19.9	19.9	19.9	19.9	19.6
	DICE	74.2	77.0	72.6	79.8	74.6	77.2	72.9	79.4	75.3	77.7	73.9	80.3
	JACRD	59.0	62.6	56.9	66.4	59.5	62.9	57.4	65.9	60.4	63.5	58.6	67.1
	AUC	94.5	93.8	94.7	90.8	94.7	93.8	94.9	92.3	94.3	93.5	94.5	88.7
	KAPPA	72.7	75.7	70.9	78.9	73.1	76.0	71.3	78.3	73.8	76.4	72.3	79.3
C2	RNDIND	94.2	95.1	93.6	96.4	94.3	95.1	93.7	95.9	94.5	95.3	94.1	96.3
	ADJRIND	70.2	73.6	68.3	77.3	70.7	73.9	68.7	76.5	71.5	74.4	69.8	77.7
	ICCORR	74.2	77.0	72.5	79.8	74.6	77.2	72.9	79.4	75.3	77.7	73.9	80.3
	VOLSMTY	80.9	85.9	78.4	96.4	81.0	86.1	78.5	92.3	82.7	87.3	80.5	97.4
	MUTINF	17.1	17.3	17.0	15.9	17.3	17.4	17.1	17.2	17.2	17.3	17.1	16.8
	DICE	87.7	85.0	f	87.4	86.8	85.4	f	86.5	87.7	85.7	f	87.6
	JACRD	78.1	73.9	f	77.7	76.7	74.5	f	76.2	78.0	75.0	f	78.0
	AUC	93.3	96.5	f	92.1	95.6	96.5	f	95.9	93.4	96.1	f	92.9
	KAPPA	87.0	84.1	f	86.7	86.1	84.5	f	85.7	87.0	84.9	f	87.0
C3	RNDIND	97.6	96.7	f	97.4	97.2	96.8	f	97.1	97.6	96.9	f	97.5
	ADJRIND	85.9	82.6	f	85.5	84.7	83.0	f	84.3	85.8	83.5	f	85.8

	ICCORR	87.7	85.0	f	87.4	86.8	85.4	f	86.5	87.6	85.7	f	87.6
	VOLSMTY	99.6	90.0	f	96.8	94.0	90.5	f	93.0	99.7	91.7	f	98.6
	MUTINF	20.5	21.1	f	20.9	21.2	21.2	f	21.2	20.5	21.1	f	20.7
	DICE	85.6	80.6	81.4	86.7	82.0	80.9	79.1	84.1	85.6	81.6	82.3	86.7
	JACRD	74.8	67.6	68.7	76.6	69.5	67.9	65.4	72.6	74.8	69.0	69.9	76.6
	AUC	96.9	98.0	97.9	90.4	97.7	97.9	98.1	97.0	97.0	97.9	97.9	90.4
	KAPPA	84.9	79.7	80.6	86.2	81.2	80.0	78.1	83.4	84.9	80.8	81.4	86.2
C4	RNDIND	97.6	96.5	96.7	97.9	96.8	96.6	96.1	97.3	97.6	96.7	96.9	97.9
	ADJRIND	83.8	78.2	79.1	85.2	79.7	78.4	76.4	82.2	83.8	79.3	80.0	85.2
	ICCORR	85.6	80.6	81.4	86.7	82.0	80.9	79.1	84.1	85.6	81.6	82.3	86.7
	VOLSMTY	90.1	82.5	83.5	93.0	84.5	82.9	80.6	88.3	90.0	83.7	84.6	93.0
	MUTINF	17.8	17.5	17.6	17.7	17.5	17.4	17.3	17.5	17.8	17.6	17.6	17.7
	DICE	89.5	85.4	83.2	90.3	87.8	86.0	83.4	88.1	89.5	86.6	84.7	90.0
	JACRD	81.0	74.5	71.3	82.3	78.2	75.4	71.5	78.7	81.0	76.4	73.5	81.9
	AUC	96.8	98.0	98.2	93.9	97.6	98.0	98.2	97.5	96.8	97.9	98.1	93.3
	KAPPA	89.0	84.6	82.2	89.8	87.1	85.2	82.4	87.4	89.0	85.9	83.8	89.5
C5	RNDIND	97.9	96.9	96.3	98.1	97.5	97.0	96.4	97.6	97.9	97.2	96.7	98.1
	ADJRIND	87.9	83.1	80.5	88.8	85.9	83.8	80.7	86.2	87.9	84.5	82.3	88.5
	ICCORR	89.5	85.4	83.2	90.3	87.8	86.0	83.4	88.1	89.5	86.6	84.7	90.0
	VOLSMTY	94.8	87.5	84.7	97.7	91.1	88.3	84.8	91.7	94.7	89.2	86.6	96.5
	MUTINF	21.7	21.4	21.1	21.5	21.7	21.5	21.1	21.7	21.8	21.6	21.3	21.6

Table B.8: *Calculated metrics for WM detection of the label maps produced by all pipelines on training subjects C1 to C5. f=fail in true detection*

	The precision of all pipelines in GM detection (%)												
	Metrics	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
	DICE	76.2	73.9	74.9	80.1	74.4	74.1	73.5	76.3	76.8	75.3	76.1	80.7
	JACRD	61.5	58.6	59.9	66.8	59.2	58.9	58.1	61.7	62.3	60.4	61.4	67.6
	AUC	90.0	85.0	89.4	90.7	86.6	85.4	86.0	86.8	89.9	86.1	89.6	89.9
	KAPPA	74.5	72.2	73.2	78.9	72.7	72.5	71.7	74.8	75.2	73.7	74.4	79.4
C1	RNDIND	94.0	94.1	93.6	95.4	93.9	94.1	93.7	94.5	94.2	94.3	94.0	95.4
	ADJRIND	71.9	69.8	70.5	76.8	70.2	70.0	69.1	72.5	72.6	71.3	71.8	77.3
	ICCORR	76.2	73.9	74.9	80.1	74.4	74.1	73.4	76.3	76.8	75.3	76.1	80.7
	VOLSMTY	92.6	96.5	92.5	96.8	99.5	97.5	99.6	98.1	93.8	97.8	93.5	99.6
	MUTINF	18.4	16.1	17.8	19.0	16.7	16.3	16.2	17.3	18.5	16.8	18.2	19.6
	DICE	68.4	73.2	65.0	77.4	69.3	73.7	66.3	77.0	70.3	74.1	67.7	76.8
	JACRD	52.0	57.7	48.2	63.1	53.0	58.3	49.6	62.6	54.2	58.9	51.2	62.4
	AUC	81.0	82.9	79.3	89.6	80.7	83.3	78.9	85.4	82.0	83.5	80.8	90.5
	KAPPA	65.9	71.1	62.3	75.6	66.9	71.6	63.7	75.2	68.0	72.1	65.1	75.0
C2	RNDIND	91.1	92.5	90.2	93.4	91.6	92.6	90.9	93.4	91.6	92.8	90.9	93.4
	ADJRIND	62.3	67.9	58.5	72.6	63.5	68.5	60.2	72.2	64.5	69.0	61.5	72.1
	ICCORR	68.4	73.2	65.0	77.4	69.2	73.7	66.3	77.0	70.3	74.1	67.7	76.8
	VOLSMTY	93.1	91.2	93.3	95.1	90.0	91.5	88.8	93.3	93.2	91.6	93.6	92.2
	MUTINF	15.8	18.0	14.2	20.6	16.0	18.3	14.6	20.2	16.7	18.6	15.5	20.1
	DICE	78.7	73.7	f	78.2	76.8	74.2	f	76.4	78.9	75.1	f	78.8
	JACRD	64.8	58.3	f	64.2	62.3	59.0	f	61.8	65.2	60.1	f	65.1
	AUC	92.0	85.2	f	88.7	88.4	86.0	f	87.6	91.7	86.2	f	88.4
	KAPPA	77.2	72.1	f	76.8	75.3	72.6	f	74.9	77.5	73.5	f	77.5
C3	RNDIND	94.6	94.1	f	95.0	94.5	94.1	f	94.5	94.7	94.4	f	95.1
	ADJRIND	74.8	69.6	f	74.6	72.9	70.2	f	72.5	75.1	71.1	f	75.3

	ICCORR	78.7	73.7	f	78.2	76.8	74.2	f	76.3	78.9	75.0	f	78.8
	VOLSMTY	91.3	97.4	f	99.3	97.9	99.1	f	99.6	92.5	98.3	f	99.1
	MUTINF	19.7	15.8	f	18.0	17.7	16.2	f	17.3	19.6	16.5	f	18.5
	DICE	78.8	74.3	71.4	81.5	75.4	74.2	70.4	79.3	79.4	75.7	73.3	81.9
	JACRD	65.1	59.1	55.6	68.8	60.5	59.0	54.4	65.7	65.8	60.8	57.9	69.4
	AUC	90.0	84.0	85.9	89.3	86.0	84.6	83.3	87.3	89.8	84.9	86.7	90.1
	KAPPA	77.4	72.7	69.5	80.2	73.8	72.6	68.5	78.0	78.0	74.1	71.5	80.7
C4	RNDIND	94.7	94.2	92.9	95.4	94.2	94.1	93.1	95.2	94.9	94.5	93.4	95.6
	ADJRIND	75.0	70.3	66.5	78.1	71.3	70.1	65.8	75.8	75.7	71.8	68.7	78.6
	ICCORR	78.8	74.3	71.4	81.5	75.4	74.2	70.4	79.3	79.4	75.6	73.3	81.9
	VOLSMTY	96.4	92.4	96.5	97.7	97.4	94.8	96.9	95.0	97.7	93.2	97.3	99.4
	MUTINF	19.5	16.3	15.9	20.7	17.1	16.4	14.9	18.8	19.6	16.9	16.7	20.7
	DICE	84.7	80.2	76.4	83.2	83.1	81.2	77.9	83.2	84.5	81.4	78.7	83.7
	JACRD	73.4	66.9	61.8	71.2	71.1	68.4	63.8	71.3	73.2	68.6	64.9	71.9
	AUC	91.2	86.9	86.7	93.5	89.3	87.8	86.3	89.0	90.8	87.7	87.9	93.5
	KAPPA	83.4	78.7	74.5	81.9	81.8	79.8	76.2	81.9	83.3	79.9	77.0	82.4
C5	RNDIND	95.5	94.5	93.2	95.3	95.2	94.8	93.8	95.3	95.5	94.8	93.9	95.4
	ADJRIND	81.2	76.1	71.5	79.6	79.5	77.3	73.4	79.7	81.1	77.5	74.2	80.2
	ICCORR	84.7	80.2	76.4	83.2	83.1	81.2	77.9	83.2	84.5	81.4	78.7	83.7
	VOLSMTY	98.8	92.8	98.3	93.7	95.6	93.9	94.5	94.5	97.8	93.4	98.1	94.3
	MUTINF	24.9	21.5	19.9	23.4	23.5	22.2	20.3	23.5	24.7	22.3	21.1	23.8

Table B.9: *Calculated metrics for GM detection of the label maps produced by all pipelines on training subjects C1 to C5. f=fail in true detection*

	The precision of all pipelines in CSF detection (%)											
	Metrics	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
	DICE	81.8	87.0	80.4	89.9	83.5	82.7	83.1	83.4	83.2	87.1	82.3	89.6
	JACRD	69.2	76.9	67.2	81.6	71.7	70.5	71.1	71.6	71.2	77.2	69.9	81.2
	AUC	85.1	89.7	84.0	94.5	86.6	86.2	86.2	87.0	86.2	89.9	85.5	95.6
	KAPPA	80.8	86.2	79.4	89.2	82.6	81.7	82.2	82.5	82.3	86.4	81.4	89.0
C1	RNDIND	96.4	97.2	96.1	97.6	96.6	96.5	96.6	96.6	96.6	97.2	96.4	97.6
	ADJRIND	79.2	84.8	77.7	88.0	81.0	80.1	80.6	80.9	80.7	85.0	79.7	87.8
	ICCORR	81.8	86.9	80.4	89.9	83.5	82.7	83.1	83.4	83.1	87.1	82.3	89.6
	VOLSMTY	83.6	90.9	81.9	99.7	86.1	86.2	85.5	87.7	85.3	91.2	84.3	97.5
	MUTINF	19.1	21.9	18.5	24.5	20.0	19.5	19.8	19.9	19.8	22.0	19.4	23.9
	DICE	81.9	86.0	79.0	85.3	84.5	86.0	83.4	86.4	83.3	86.2	81.0	85.8
	JACRD	69.4	75.4	65.3	74.3	73.2	75.4	71.6	76.0	71.4	75.7	68.1	75.1
	AUC	86.4	91.6	83.7	89.0	89.1	91.5	87.8	93.3	87.8	92.1	85.5	89.7
	KAPPA	81.0	85.2	78.0	84.3	83.7	85.2	82.5	85.6	82.4	85.4	80.1	84.9
C2	RNDIND	96.5	97.0	96.1	96.5	96.8	97.0	96.7	97.0	96.7	97.0	96.4	96.6
	ADJRIND	79.4	83.7	76.3	82.6	82.2	83.8	81.0	84.1	80.9	83.9	78.4	83.3
	ICCORR	81.9	86.0	79.0	85.3	84.5	86.0	83.4	86.4	83.3	86.2	81.0	85.8
	VOLSMTY	87.8	97.5	83.2	91.2	92.5	97.1	90.1	98.7	90.5	98.6	86.4	92.6
	MUTINF	18.3	20.9	16.9	22.6	19.7	20.9	19.1	21.6	19.0	21.1	17.9	22.5
	DICE	83.3	84.9	f	86.9	81.5	80.5	f	81.1	84.2	85.4	f	86.9
	JACRD	71.3	73.8	f	76.9	68.8	67.3	f	68.2	72.7	74.5	f	76.9
	AUC	87.3	89.1	f	94.1	86.2	85.3	f	86.1	88.3	89.7	f	94.1
	KAPPA	82.4	84.1	f	86.2	80.6	79.5	f	80.2	83.4	84.6	f	86.2
C3	RNDIND	96.8	97.1	f	97.3	96.5	96.4	f	96.4	97.0	97.1	f	97.3
	ADJRIND	81.0	82.7	f	84.9	79.1	77.9	f	78.6	82.0	83.2	f	84.9

	ICCORR	83.3	84.9	f	86.9	81.5	80.5	f	81.1	84.2	85.3	f	86.9
	VOLSMTY	88.7	91.8	f	97.6	88.0	86.6	f	88.1	90.5	93.0	f	97.6
	MUTINF	18.5	19.4	f	20.9	17.7	17.1	f	17.5	19.0	19.7	f	20.9
	DICE	83.7	87.6	76.8	86.9	83.2	81.0	81.7	80.8	85.1	87.9	79.4	88.1
	JACRD	71.9	78.0	62.4	76.8	71.3	68.1	69.1	67.7	74.1	78.4	65.9	78.7
	AUC	86.7	90.7	81.4	97.1	86.7	85.1	85.1	85.9	88.0	91.0	83.3	96.6
	KAPPA	82.9	87.0	75.8	86.2	82.4	80.1	80.9	79.8	84.4	87.3	78.5	87.5
C4	RNDIND	97.0	97.6	96.0	97.4	96.9	96.5	96.7	96.4	97.2	97.6	96.4	97.6
	ADJRIND	81.5	85.8	74.1	85.0	81.0	78.5	79.4	78.2	83.0	86.1	76.9	86.3
	ICCORR	83.7	87.6	76.8	86.9	83.2	81.0	81.7	80.8	85.1	87.9	79.4	88.1
	VOLSMTY	86.2	92.6	77.8	91.2	86.8	85.1	83.9	88.2	88.2	93.2	80.8	93.6
	MUTINF	18.7	20.9	15.8	20.4	18.5	17.4	17.8	17.3	19.4	21.0	16.9	21.1
	DICE	85.5	85.5	78.0	84.9	86.1	85.8	83.3	86.2	85.8	85.7	80.7	85.5
	JACRD	74.7	74.7	64.0	73.7	75.6	75.1	71.4	75.8	75.1	75.0	67.6	74.6
	AUC	91.0	91.0	82.8	89.5	91.4	90.8	87.5	92.4	91.7	91.6	85.1	90.5
	KAPPA	84.8	84.8	77.2	84.1	85.4	85.1	82.6	85.6	85.1	85.1	79.9	84.7
C5	RNDIND	97.4	97.4	96.6	97.0	97.5	97.5	97.2	97.5	97.4	97.4	96.9	97.2
	ADJRIND	83.6	83.6	75.7	82.7	84.3	83.9	81.3	84.4	83.9	83.8	78.5	83.4
	ICCORR	85.5	85.5	78.0	84.8	86.1	85.8	83.3	86.2	85.7	85.7	80.7	85.5
	VOLSMTY	96.2	96.2	81.4	93.2	96.8	95.5	89.3	99.0	97.8	97.7	85.5	95.2
	MUTINF	18.2	18.2	14.7	19.4	18.5	18.3	16.8	18.8	18.5	18.4	15.7	19.3

Table B.10: *Calculated metrics for CSF detection of the label maps produced by all pipelines on training subjects C1 to C5. f=fail in true detection*

	The precision of all pipelines in WM+GM detection (%)											
	Metrics	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
	DICE	90.9	91.7	90.5	93.0	91.4	91.7	91.1	92.2	91.3	91.9	91.0	93.1
	JACRD	83.4	84.6	82.6	87.0	84.1	84.6	83.7	85.5	84.0	85.0	83.5	87.2
	AUC	98.2	98.1	98.3	96.0	98.3	98.2	98.4	98.1	98.2	98.1	98.2	95.5
	KAPPA	90.0	90.8	89.5	92.4	90.5	90.8	90.2	91.4	90.4	91.1	90.1	92.5
C1	RNDIND	96.7	97.0	96.5	97.6	96.9	97.0	96.8	97.2	96.9	97.1	96.7	97.6
	ADJRIND	88.2	89.2	87.6	91.1	88.8	89.2	88.4	89.9	88.7	89.5	88.3	91.2
	ICCORR	93.7	94.6	93.4	95.3	94.3	94.6	94.1	94.9	94.0	94.7	93.8	95.4
	VOLSMTY	92.7	93.8	92.0	99.5	93.0	93.6	92.6	94.7	93.2	94.2	92.8	98.3
	MUTINF	34.3	34.6	34.2	33.9	34.6	34.6	34.5	34.7	34.5	34.6	34.4	34.4
	DICE	89.8	91.0	88.9	90.5	90.3	91.1	89.7	91.6	90.2	91.1	89.5	90.9
	JACRD	81.5	83.5	80.0	82.7	82.3	83.7	81.4	84.5	82.2	83.7	81.1	83.2
	AUC	97.3	96.4	97.5	96.6	97.0	96.5	97.3	95.6	97.1	96.2	97.4	96.0
	KAPPA	88.6	90.0	87.5	89.5	89.2	90.1	88.5	90.6	89.1	90.1	88.3	89.9
C2	RNDIND	95.7	96.3	95.3	96.4	96.0	96.4	95.7	96.7	95.9	96.4	95.6	96.5
	ADJRIND	86.1	87.9	84.9	87.5	86.9	88.1	86.1	88.8	86.8	88.1	85.8	87.9
	ICCORR	93.6	94.3	93.1	93.3	94.0	94.4	93.8	94.4	93.9	94.3	93.5	93.7
	VOLSMTY	92.8	96.7	91.2	95.7	94.2	96.6	92.9	99.4	93.9	97.3	92.4	97.4
	MUTINF	36.3	36.1	36.1	34.1	36.3	36.3	36.3	35.9	36.4	36.0	36.3	34.6
	DICE	92.0	91.7	f	92.6	92.1	91.8	f	92.2	92.2	92.0	f	92.6
	JACRD	85.2	84.7	f	86.2	85.4	84.8	f	85.5	85.5	85.1	f	86.3
	AUC	97.2	97.5	f	95.3	97.5	97.7	f	97.5	97.1	97.3	f	95.5
	KAPPA	91.2	90.9	f	91.8	91.3	90.9	f	91.4	91.4	91.1	f	91.9
C3	RNDIND	97.1	96.9	f	97.3	97.1	96.9	f	97.1	97.2	97.1	f	97.4
	ADJRIND	89.6	89.2	f	90.4	89.7	89.2	f	89.8	89.8	89.5	f	90.4

	ICCORR	94.5	94.6	f	95.0	94.7	94.6	f	94.8	94.6	94.7	f	95.0
	VOLSMTY	96.2	95.2	f	98.5	95.7	94.8	f	95.8	96.7	95.9	f	98.8
	MUTINF	34.9	35.0	f	34.7	35.2	35.1	f	35.2	34.9	35.0	f	34.7
	DICE	91.6	91.6	89.2	92.7	91.4	91.4	90.0	92.7	92.0	91.9	89.9	93.0
	JACRD	84.5	84.5	80.4	86.3	84.1	84.1	81.8	86.4	85.1	85.0	81.7	86.9
	AUC	98.4	98.1	98.6	94.2	98.3	98.3	98.5	97.6	98.3	98.1	98.6	94.6
	KAPPA	90.8	90.9	88.1	92.0	90.5	90.5	89.0	92.1	91.2	91.2	89.0	92.3
C4	RNDIND	97.2	97.2	96.2	97.6	97.1	97.1	96.5	97.6	97.3	97.3	96.5	97.7
	ADJRIND	89.3	89.3	86.1	90.7	89.0	89.0	87.1	90.7	89.8	89.7	87.1	91.1
	ICCORR	94.2	94.7	92.6	94.9	94.3	94.5	93.5	95.3	94.5	94.9	93.1	95.2
	VOLSMTY	93.3	93.9	89.8	95.4	93.1	93.2	91.0	96.3	93.9	94.4	90.8	96.2
	MUTINF	33.2	32.9	32.5	33.3	33.0	32.9	32.7	33.0	33.3	33.0	32.7	33.3
	DICE	93.9	93.1	91.6	93.1	93.7	93.4	92.5	93.7	93.8	93.3	92.3	93.4
	JACRD	88.5	87.1	84.4	87.0	88.2	87.5	86.0	88.2	88.4	87.4	85.7	87.6
	AUC	97.4	97.6	98.4	97.0	97.5	97.7	98.2	97.2	97.2	97.3	98.3	96.8
	KAPPA	93.2	92.3	90.6	92.3	93.0	92.6	91.6	93.1	93.2	92.5	91.4	92.7
C5	RNDIND	97.6	97.2	96.5	97.4	97.5	97.3	96.9	97.5	97.6	97.3	96.8	97.5
	ADJRIND	91.8	90.7	88.6	90.8	91.6	91.1	89.8	91.6	91.8	91.0	89.6	91.2
	ICCORR	95.7	95.5	94.4	95.0	95.8	95.6	95.1	95.8	95.7	95.5	94.9	95.3
	VOLSMTY	98.0	96.7	92.7	98.0	97.6	96.7	94.4	98.4	98.5	97.5	93.9	98.9
	MUTINF	38.4	38.0	38.2	36.6	38.4	38.3	38.3	38.2	38.2	38.0	38.3	37.1

Table B.11: *Calculated metrics for WM+GM detection of the label maps produced by all pipelines on training subjects C1 to C5. f=fail in true detection*

	The precision of all pipelines in WM+GM+CSF detection (%)											
	Metrics	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
	DICE	94.9	94.9	94.6	95.9	94.6	94.3	94.5	94.6	95.0	95.1	94.9	95.9
	JACRD	90.3	90.3	89.8	92.1	89.8	89.2	89.5	89.8	90.5	90.7	90.3	92.2
	AUC	98.5	98.5	98.6	97.7	98.4	98.0	98.4	97.8	98.5	98.5	98.5	97.5
	KAPPA	94.2	94.3	94.0	95.4	93.9	93.6	93.8	93.9	94.4	94.5	94.3	95.4
C1	RNDIND	97.8	97.8	97.7	98.3	97.7	97.6	97.6	97.7	97.9	97.9	97.8	98.3
	ADJRIND	92.9	93.0	92.6	94.4	92.6	92.2	92.4	92.6	93.2	93.2	93.0	94.4
	ICCORR	97.6	97.6	97.5	98.0	97.5	97.4	97.4	97.5	97.7	97.7	97.6	98.0
	VOLSMTY	96.7	96.7	96.4	99.9	96.7	97.2	96.5	97.9	97.0	97.1	96.7	99.4
	MUTINF	42.1	42.2	42.0	42.2	41.8	41.2	41.7	41.3	42.2	42.3	42.2	42.4
	DICE	93.4	94.2	92.8	94.5	93.6	94.2	93.2	94.6	93.7	94.3	93.2	94.6
	JACRD	87.5	89.0	86.5	89.5	88.0	89.1	87.2	89.8	88.1	89.3	87.3	89.7
	AUC	98.0	97.6	98.1	97.9	97.9	97.6	98.0	97.1	97.9	97.5	98.0	97.4
	KAPPA	92.4	93.4	91.7	93.8	92.7	93.4	92.2	93.9	92.8	93.5	92.2	93.9
C2	RNDIND	96.7	97.2	96.4	97.5	96.9	97.2	96.7	97.5	96.9	97.3	96.7	97.5
	ADJRIND	90.5	91.7	89.6	92.2	90.8	91.8	90.2	92.4	90.9	91.9	90.3	92.3
	ICCORR	97.0	97.3	96.8	97.3	97.1	97.3	96.9	97.4	97.1	97.4	97.0	97.4
	VOLSMTY	95.7	97.8	94.7	97.7	96.3	97.9	95.5	99.6	96.3	98.2	95.5	98.8
	MUTINF	43.5	43.7	43.3	43.0	43.6	43.8	43.4	43.7	43.7	43.8	43.5	43.2
	DICE	95.5	95.1	f	95.7	95.0	94.6	f	94.8	95.6	95.2	f	95.8
	JACRD	91.4	90.6	f	91.8	90.5	89.7	f	90.2	91.5	90.9	f	91.9
	AUC	98.1	98.4	f	97.3	97.9	97.9	f	97.8	98.0	98.3	f	97.5
	KAPPA	94.9	94.4	f	95.2	94.4	93.9	f	94.2	95.0	94.7	f	95.2
C3	RNDIND	98.0	97.8	f	98.2	97.8	97.6	f	97.8	98.1	97.9	f	98.2
	ADJRIND	93.8	93.2	f	94.1	93.1	92.5	f	92.9	93.9	93.4	f	94.2

	ICCORR	97.9	97.8	f	98.0	97.7	97.6	f	97.7	98.0	97.8	f	98.0
	VOLSMTY	98.6	97.3	f	99.3	98.3	97.7	f	98.4	98.8	97.8	f	99.6
	MUTINF	42.5	42.5	f	42.5	42.0	41.7	f	41.7	42.5	42.5	f	42.5
	DICE	95.1	94.7	93.7	95.7	94.4	93.8	93.8	94.2	95.3	94.9	94.1	95.8
	JACRD	90.7	89.9	88.1	91.7	89.4	88.4	88.3	89.1	91.0	90.3	88.9	92.0
	AUC	98.6	98.6	98.7	96.8	98.3	97.9	98.6	97.1	98.6	98.5	98.7	97.0
	KAPPA	94.6	94.1	93.0	95.2	93.8	93.1	93.0	93.6	94.8	94.3	93.4	95.4
C4	RNDIND	98.0	97.9	97.4	98.3	97.7	97.5	97.5	97.7	98.1	98.0	97.6	98.3
	ADJRIND	93.4	92.9	91.5	94.2	92.5	91.7	91.6	92.3	93.7	93.2	92.0	94.4
	ICCORR	97.7	97.5	97.1	97.9	97.4	97.1	97.1	97.2	97.8	97.6	97.3	98.0
	VOLSMTY	96.9	96.5	94.8	98.0	96.6	97.0	95.1	99.2	97.1	96.8	95.3	98.3
	MUTINF	40.4	40.0	39.6	40.6	39.6	38.8	39.5	38.5	40.5	40.1	39.9	40.6
	DICE	96.5	95.8	94.9	96.2	96.2	95.9	95.3	96.2	96.5	96.0	95.3	96.3
	JACRD	93.2	91.9	90.2	92.7	92.8	92.1	91.0	92.8	93.2	92.2	91.1	92.9
	AUC	98.5	98.6	98.9	98.2	98.5	98.6	98.8	98.4	98.4	98.5	98.9	98.1
	KAPPA	96.0	95.2	94.2	95.8	95.7	95.4	94.7	95.8	96.0	95.4	94.7	95.9
C5	RNDIND	98.4	98.0	97.6	98.3	98.3	98.1	97.8	98.3	98.4	98.1	97.8	98.4
	ADJRIND	95.0	94.0	92.7	94.7	94.7	94.2	93.3	94.7	95.0	94.3	93.4	94.9
	ICCORR	98.5	98.2	97.8	98.3	98.4	98.2	98.0	98.4	98.5	98.2	98.0	98.4
	VOLSMTY	98.9	97.7	95.7	99.2	98.4	97.8	96.5	98.9	99.1	98.2	96.4	99.7
	MUTINF	45.4	44.8	44.6	44.4	45.2	45.0	44.8	45.0	45.3	44.9	44.9	44.7

Table B.12: *Calculated metrics for WM+GM+CSF detection of the label maps produced by all pipelines on training subjects C1 to C5. f=fail in true detection*

Index