UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO
DEPARTAMENTO DE CIÊNCIAS DAS IMAGENS E FÍSICA MÉDICA

Aplicação de técnicas de reconhecimento de padrões em dados quantitativos de neuroimagens por ressonância magnética em pacientes de Esclerose Múltipla

RODRIGO ANTONIO PESSINI

Ribeirão Preto
2016

Rodrigo Antonio Pessini

Aplicação de técnicas de reconhecimento de padrões em dados quantitativos de neuroimagens por ressonância magnética em pacientes de Esclerose Múltipla

Dissertação apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para obtenção do título de Mestre Profissional em Física Médica.

Área: Física Médica

Orientador: Prof. Dr. Carlos Ernesto Garrido Salmon

(“Versão corrigida. A versão original encontra-se disponível tanto na Biblioteca da Unidade que aloja o Programa, quanto na Biblioteca Digital de Teses e Dissertações da USP (BDTD)”).

Ribeirão Preto
2016

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

FICHA CATALOGRÁFICA

Pessini, Rodrigo Antonio.
Aplicação de técnicas de reconhecimento de padrões em dados quantitativos de neuroimagens por ressonância magnética em pacientes de esclerose múltipla / Rodrigo Antonio Pessini; Orientador: Carlos Ernesto Garrido Salmon – Ribeirão Preto, 2016.
103 págs, 36 ilustrações, 30 cm.

Dissertação (Programa de Mestrado Profissional em Ciências das Imagens e Física Médica – Área de Concentração: Física Médica) – Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo.

1) Reconhecimento de Padrões; 2) Esclerose Múltipla; Ressonância Magnética Quantitativa.

FOLHA DE APROVAÇÃO

Rodrigo Antonio Pessini

Aplicação de técnicas de reconhecimento de padrões em dados quantitativos de neuroimagens por ressonância magnética em pacientes de esclerose múltipla.

Dissertação apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para obtenção do título de Mestre Profissional em Física Médica.

Área de Concentração: Física Médica.

Aprovado em: 17/06/2016.

Banca Examinadora:

Prof. Dr. Paulo Mazzoncini de Azevedo Marques.

Instituição: Universidade de São Paulo Assinatura:_____

Prof. Dr. Renato Tinós.

Instituição: Universidade de São Paulo Assinatura:_____

Prof. Dr. Edilson Carlos Caritá.

Instituição: Universidade de Ribeirão Preto Assinatura:_____

DEDICATÓRIA

Dedico a todas as pessoas que direcionam suas vidas para o caminho da iluminação.

AGRADECIMENTOS

Agradeço primeiramente a Deus pela vida e pela inspiração de ideias para realizar essa dissertação.

A toda minha família em especial a minha mãe Irany, minha irmã Elisabete e minha tia Nadir pela ajuda incondicional em todos os sentidos.

A Fabiola pelos ensinamentos de vida e carinho que foram importantes para a realização dessa dissertação.

Ao Prof. Dr. Antônio Carlos dos Santos pela oportunidade de trabalho, experiência, conhecimento, ajuda e autorização na realização dessa dissertação.

A Profª. Drª. Paula Diniz pela criação da base de conhecimento no processamento de imagens e pelo seu conhecimento que foram fundamentais para a criação e desenvolvimento dessa dissertação.

Ao meu orientador Prof. Dr. Carlos Ernesto Garrido Salmon pelo conhecimento, experiência, orientação e ajuda fundamental na criação, desenvolvimento e conclusão dessa dissertação.

Ao Prof. Dr. Paulo Mazzoncini de Azevedo Marques pela ideia base deste trabalho, pela orientação e apoio no desenvolvimento dessa dissertação.

A Profª. Drª. Juliana Souza pelo incentivo, apoio e ajuda na criação dessa dissertação.

A Lucas Frighetto Pereira pela colaboração fundamental com seus conhecimentos em técnicas de reconhecimento de padrões, pela ajuda com o inglês e apoio nas diversas tarefas deste programa de pós-graduação.

A Rafael Reis pela sua cooperação em diversos momentos deste trabalho, pela ajuda na criação de figuras e edição e com suas experiências nos programas de pós-graduação.

A Jaqueline Macarios pela ajuda e conhecimento nos processamentos de imagem.

A Carlo Rondinoni pela ajuda em física médica e nos procedimentos de processamento de imagens.

A Jeam Barbosa pela ajuda em física médica, na apresentação de resultados e nas colocações de referências bibliográficas.

A todos os pesquisadores que realizaram pesquisas junto ao CCIFM e disponibilizaram os dados quantitativos de pacientes que utilizei para essa pesquisa, sendo fundamentais para a realização deste trabalho.

A todos os pacientes diagnosticados que realizaram exames de ressonância magnética em busca de tratamento no HCRP e disponibilizaram as suas imagens do cérebro para que fossem gerados os dados quantitativos fundamentais para a realização dessa dissertação.

A todos os controles normais que realizaram exames de ressonância magnética e disponibilizaram suas imagens para que fossem gerados os dados quantitativos fundamentais para a realização dessa dissertação.

A todos os docentes do CCIFM do Hospital das Clínicas de Ribeirão Preto pelo conhecimento, experiência e orientação nas disciplinas do programa e na realização dessa dissertação.

A todos os alunos e professores do INBRAIN pelos conhecimentos em física médica e pela ajuda em vários momentos da pesquisa.

A todos os funcionários do CCIFM do Hospital das Clínicas de Ribeirão Preto em especial a Luciana, Dona Elza e Bete pelo apoio na realização dessa dissertação.

A todos os funcionários da Pós Graduação da Faculdade de Medicina pela orientação na execução dos procedimentos do programa de mestrado profissional.

A todos os alunos que passaram pelo CCIFM que deixaram seus conhecimentos e experiências dos cursos de graduação e pós-graduação.

Ao departamento de Unidade de Pesquisa Clínica do Hospital das Clínicas de Ribeirão Preto pela aprovação do projeto inicial e devidas orientações com a pesquisa.

Ao Comitê de Ética do Hospital das Clínicas de Ribeirão Preto pelo apoio e autorização para a realização desta pesquisa.

A todos os meus amigos que contribuíram direta ou indiretamente para a realização deste trabalho.

RESUMO

Pessini, R. A. **Aplicação de técnicas de reconhecimento de padrões em dados quantitativos de neuroimagens por ressonância magnética em pacientes de Esclerose Múltipla**. 2016. 103 páginas. Dissertação (Mestrado Profissional). Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2016.

Na última década diferentes modalidades e técnicas quantitativas de neuroimagens aplicadas ao estudo de doenças neuro-degenerativas vêm fornecendo um volume cada vez maior de dados, tornando sua utilização uma tarefa complexa. Paralelamente, técnicas computacionais de reconhecimento de padrões vêm sendo desenvolvidas para apoiar a tomada de decisão humana. O propósito geral do presente estudo é aplicar técnicas de reconhecimento de padrões em dados quantitativos de neuroimagens adquiridas por ressonância magnética (RM) em pacientes com Esclerose Múltipla (EM). Especificamente foram avaliados dados retrospectivos de um grupo de 203 sujeitos controle sem doenças neurológicas e um grupo de 144 pacientes portadores de Esclerose Múltipla. Os dados usados foram provenientes da combinação de ferramentas computacionais de processamento de imagens e neuroimagens adquiridas em um aparelho de RM de 3 Tesla usando diferentes técnicas quantitativas: Difusão, Relaxometria, Taxa de Transferência de Magnetização (MTR) e Volumetria. Os dados das diferentes técnicas quantitativas em 126 regiões cerebrais não excludentes foram processados no programa de mineração de dados WEKA. Os algoritmos: *k-nearest-neighbor* (KNN) com diferentes números de vizinhos e *Support vector machine* (SVM) foram utilizados para a classificação e agrupamento desses dados. As regiões com maior contribuição na separação de ambos os grupos foram encontradas na substância branca (SB) nas seguintes estruturas: corpo caloso, precúneo, cerebelo e fusiforme esquerdos. Outro atributo significante foi o de hipo-intensidades que pode ser associado à presença de lesões, também na substância branca. Dentre as técnicas, a mais relevante foi a MTR com 92,9% de valor médio de área sob a curva ROC (*Receiver Operating Characteristic*), considerando os diferentes algoritmos de classificação e métodos de seleção de atributos, porém uma análise global incluindo os dados de todas as técnicas elevou esta área para 96,2%. O algoritmo KNN com 5 vizinhos foi considerado o melhor classificador geral para o conjunto de dados e tarefas aqui explorados resultando em 91,7% de valor médio de área sob a curva ROC, seguido por KNN3 com 90,8%, KNN1 com 87,1% e SVM com 85,8%. Uma classificação restrita, com áreas reconhecidamente afetadas pela EM e com KNN5 trouxe resultados de classificação 2,1% de valor médio de área sob a curva ROC inferiores à classificação principal sem restrições. O uso de técnicas de reconhecimento de padrões a partir dos dados de técnicas quantitativas de neuroimagem aplicadas à amostra estudada, demonstrou que a substância branca do cérebro é a mais afetada pela EM seguindo um padrão global com maior envolvimento no hemisfério esquerdo. A estratégia sugerida neste problema de classificação seria o uso dos dados de todas as técnicas quantitativas aqui discutidas provenientes das regiões envolvidas na classificação restrita conjuntamente com o *precuneus* e o *fusiform*, aplicando o KNN5 com seleção de melhores atributos.

Palavras-chave: Reconhecimento de Padrões, Esclerose Múltipla, Ressonância Magnética Quantitativa.

ABSTRACT

Pessini, R. A. **Application of pattern recognition techniques for quantitative data from neuroimaging MRI in multiple sclerosis patients.** 2016. 103 pages. Dissertation (Professional Master). Faculty of Medicine of Ribeirão Preto, University of São Paulo, Ribeirão Preto, 2016.

In the last decade different modalities and quantitative neuroimaging techniques applied to the study of neurodegenerative diseases have been providing an increasing volume of data, making use of such a complex task. At the same time, computational techniques of pattern recognition have been developed to favor the human performance. The purpose of this study is to apply techniques of pattern recognition in quantitative data of neuroimaging acquired by magnetic resonance (MR) in patients of multiple sclerosis (MS). There were evaluated retrospective data from a control group of 203 people without neurological diseases and a group of 144 patients with MS. The data used were from the combination of computational tools of image processing and neuroimaging acquired in an MR apparatus of 3 Tesla using different quantitative techniques: diffusion, relaxometry, magnetization transfer rate (MTR) and volumetry. The data of the different quantitative techniques in 126 regions of the brain not mutually excludent were processed in the software WEKA of data mining. The algorithms: *k*-nearest-neighbor (KNN) with different numbers of neighbors and support vector machine (SVM) were used for the classification and grouping of the data. The regions with the highest contribution to the separation of both groups were found in the white matter of the following structures: corpus callosum, precuneus, cerebellum and left spindle. Another significant attribute was the hypo-intensities that can be associated to the presence of lesions in the white matter also. The most relevant technique was the MTR with 92.9% of average area under ROC, considering the different algorithms of classification and methods of feature selection, however a global analysis including data from all techniques increased this area under ROC to 96.2%. The KNN algorithm for $k = 5$ is considered the best overall classifier for the dataset resulting in 91.7% average area under ROC, followed by KNN for $k = 3$ with 90.8%, KNN for $k = 1$ with 87.1%, and SVM with 85.8%. A restricted classification with known affected areas by MS and KNN for $k = 5$ brought results of classification of average area under ROC of 2.1% inferior to the main classification without restrictions. The use of techniques of pattern recognition from the data of quantitative techniques of neuroimaging showed that the white matter of the brain is the most affected by MS following a global pattern with higher involvement in the left hemisphere. The strategy suggested in this classification issue would be the use of data from all quantitative techniques discussed from the regions involved in restricted classification jointly with precuneus and fusiform, applying KNN for $k = 5$ with selection of best features.

Keywords: Pattern Recognition, Multiple Sclerosis, MRI Quantitative.

LISTA DE FIGURAS

Figura 1.1 – Diagrama de Fluxo de Trabalho do CCIFM.....20

Figura 1.2 – Esquema Padrão de Processamento de Imagens.....21

Figura 2.1 – Grau de destrutibilidade das lesões dos pacientes de EM.....32

Figura 2.2 – Representação de dois conjuntos de classes saudáveis (azuis) e pacientes (brancos) sobre os eixos de duas características (V.H.) e (V.C.).....40

Figura 2.3 – Fluxograma do modelo de classificação baseado em KNN e SVM.....41

Figura 2.4 – Esquema representativo das tarefas de classificação executadas 45

Figura 3.1 – Valores médios dos melhores atributos obtidos na tarefa de classificação usando o vetor global nos grupos de controles e pacientes.....48

Figura 3.2 – Valores médios dos melhores atributos obtidos na tarefa de classificação usando o vetor técnica Anisotropia Fracionada nos grupos de controles e pacientes.....49

Figura 3.3 – Valores médios dos melhores atributos obtidos na tarefa de classificação usando o vetor técnica Difusibilidade Média nos grupos de controles e pacientes.....50

Figura 3.4 – Valores médios dos melhores atributos obtidos na tarefa de classificação usando o vetor técnica Difusibilidade Paralela nos grupos de controles e pacientes.....51

Figura 3.5 – Valores médios dos melhores atributos obtidos na tarefa de classificação usando o vetor técnica Difusibilidade Perpendicular nos grupos de controles e pacientes.....51

Figura 3.6 – Valores médios dos melhores atributos obtidos na tarefa de classificação usando o vetor técnica Espessura Cortical nos grupos de controles e pacientes.....52

Figura 3.7 – Valores médios dos melhores atributos obtidos na tarefa de classificação usando o vetor técnica Transferência de Magnetização nos grupos de controles e pacientes.....53

Figura 3.8 – Valores médios dos melhores atributos obtidos na tarefa de classificação usando o vetor técnica Relaxometria nos grupos de controles e pacientes.....54

Figura 3.9 – Valores médios dos melhores atributos obtidos na tarefa de classificação usando o vetor técnica Volumetria nos grupos de controles e pacientes.....55

Figura 3.10 – Regiões cerebrais com maior número indicações de melhores atributos na classificação usando os vetores técnicas quantitativas.....56

Figura 3.11 – Regiões cerebrais com maior número indicações de melhores atributos na classificação restrita usando os vetores técnicas quantitativas.....56

Figura 3.12 – Valores de AUROC por vetor de atributos usando KNN1 e considerando diferentes esquemas de pré-seleção de atributos.....58

Figura 3.13 – Valores de AUROC por vetor de atributos usando KNN3 e considerando diferentes esquemas de pré-seleção de atributos.....59

Figura 3.14 – Valores de AUROC por vetor de atributos usando KNN5 e considerando diferentes esquemas de pré-seleção de atributos.....59

Figura 3.15 – Áreas sob a curva ROC por vetor e por algoritmo de classificação, considerando a melhor estratégia de pré-seleção de atributos em cada caso.....60

Figura 3.16 – Valores de sensibilidade e especificidade nas melhores classificações usando os vetores Global ou MTR.....61

Figura 3.17 – Modelo em 3D do cérebro indicando a localização das 3 regiões com os melhores atributos na classificação usando os diferentes vetores das técnicas.....62

Figura 3.18 – Modelo em 3D do cérebro indicando a localização de outras regiões envolvidas nos melhores atributos na classificação usando os diferentes vetores das técnicas.....62

Figura 4.1 – Exemplos de mapas de neuroimagens obtidos a partir de dados de DTI de uma paciente de EM (feminino, 51 anos): (A) Anisotropia Fracionada ou Fraccional. (B) Difusibilidade média.....70

Figura 4.2 – Exemplo do resultado de uma segmentação de estruturas corticais e subcorticais realizada pelo FreeSurfer a partir de uma imagem por RM ponderada em T1 de uma paciente de EM (feminino, 47 anos).....71

Figura 4.3 – Exemplos de mapas de neuroimagens obtidos a partir de dados de IRMQ de uma paciente de EM (feminino, 51 anos): (A) MTR (B) Relaxometria.....73

LISTA DE ABREVIAÇÕES E SIGLAS

ADC – do inglês, "*Apparent Diffusion Coefficient*", Coeficiente de Difusão Aparente.

ANF – Anisotropia Fracionada.

AUROC – do inglês, "*Área Under ROC*", Área sob a curva ROC.

CEP – Comitê de Ética em Pesquisa.

CID – Classificação Internacional de Doenças.

CCIFM – Centro de Ciências da Imagem e Física Médica.

CN – Controles Normais.

CNS – Conselho Nacional de Saúde.

DICOM – do inglês, "*Digital Imaging and Communications in Medicine*", Imagem Digital e Comunicações em Medicina.

DTI – do inglês, "*Diffusion Tensor Imaging*", Imagem por tensor de difusão.

DME – Difusibilidade Média.

DPA – Difusibilidade Paralela.

DPE – Difusibilidade Perpendicular.

ECT – Espessura Cortical.

EM – Esclerose Múltipla.

EPI – do inglês, "*echo planar image*", Imagem eco-planar.

FOV – do inglês, "*field of view*", Campo de visão.

GLO – Vetor Global.

GraSE – do inglês, "*Gradient Spin Echo*", Combinação das técnicas Eco de gradiente de eco e eco de spins.

HCFMRP – Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto.

IRMQ – Imagem por Ressonância Magnética Quantitativa.

KNN – do inglês, "*K-Nearest-Neighbors*", Vizinhos mais próximos.

LAPIR – Laboratório de Processamento de Imagens e Informática Radiológica.

LH – do inglês, "*Left Hemisphere*", Hemisfério Esquerdo.

MA – Melhores Atributos.

MINC – do inglês, "*Medical Image NetCDF*", Imagem Médica NetCDF.

MTR – do inglês, "*Magnetization Transfer Ratio*", Taxa de Transferência de Magnetização.

NAA – N-Acetil-Aspartato.

PIXEL – Combinação de palavras do inglês, "*Picture*" e "*Element*".

REL – Relaxometria.

RH – do inglês, "*Right Hemisphere*", Hemisfério Direito.

RM – Ressonância Magnética.

ROC – do inglês, "*Receiver Operating Characteristic"*, Característica de Operação do Receptor.

ROI – do inglês, "*region-of-interest*", Região de interesse.

SI – do inglês, "*shot interval*", Intervalo de disparo.

T – Tesla.

TA – Todos os Atributos.

TE – Tempo de Eco.

TFE – do inglês, "*turbo field echo*", Eco de campo acelerado.

TR – Tempo de Repetição.

TSE – do inglês, "*Turbo Spin Echo*", Eco de spins acelerado.

SVM – do inglês, "*Support Vector Machines*", Máquinas de vetor de suporte.

SNC – Sistema Nervoso Central.

USP – Universidade de São Paulo.

VOL – Volumetria.

VOXEL – Combinação de palavras do inglês, "*Volume*" e "*Pixel*".

WEKA – do inglês, "*Waikato Environment for Knowledge Analysis*", Software que indica: Ambiente de Análise de Dados de Conhecimento da Universidade de Waikato.

LISTA DE QUADROS

Quadro 1.1 – Nomenclatura das imagens, usada no CCIFM, e as respectivas medidas geradas pelas sequências quantitativas.....22

Quadro 2.1 – Dados demográficos dos casos analisados.....32

Quadro 2.2 – Unidades de medida das técnicas quantitativas.....34

Quadro 2.3 – Motivos de eliminação de registros por técnica quantitativa.....35

Quadro 2.4 – Motivos de eliminação de registros pela checagem por técnica quantitativa.....37

Quadro 2.5 – Atributos eliminados com volumes médios inferiores a 0,2%.....38

Quadro 2.6 – Atributos unidos para alcançar volumes médios superiores a 0,2%.....38

Quadro 2.7 – Quantidade de instâncias por vetor.....43

Quadro 3.1 – Comparativo de porcentagem dos valores de AUROC para as classificações geral e restrita.....63

SUMÁRIO

LISTA DE FIGURAS 8
LISTA DE ABREVIAÇÕES E SIGLAS 10
LISTA DE QUADROS 12
SUMÁRIO 13
1. INTRODUÇÃO 14
1.1 Esclerose Múltipla 14
1.2 Ressonância Magnética 17
1.3 Centro de Imagens 19
1.4 Técnicas Quantitativas 22
1.5 Técnicas de Reconhecimento de Padrões 25
1.5.1 *K-Nearest-Neighbors (*KNN*)* 26
1.5.2 *Support Vector Machines (*SVM*)* 28
1.6 Justificativa e Objetivos 29
2. MATERIAIS E MÉTODOS 30
2.1 Casuística 30
2.2 Dados Quantitativos 33
2.3 Critérios de Inclusão e Exclusão de Dados 34
2.3.1 Atributos e Registros 35
2.3.2 Checagem de Valores 36
2.3.3 Redução das Áreas de Interesse 37
2.4 Algoritmos de Classificação 39
2.5 Classificações de Dados 42
2.6 Tarefas de Classificação 44
3. RESULTADOS 46
3.1 Melhores Atributos 46
3.2 Classificação 57
4. DISCUSSÃO 64
4.1 Melhores Atributos 64
4.2 Classificação 67
5. CONCLUSÕES 75
6. REFERÊNCIAS BIBLIOGRÁFICAS 78
APÊNDICES 83
ANEXOS 102

1. INTRODUÇÃO

Na literatura existem muitos trabalhos de aplicação de algoritmos de reconhecimento de padrões, principalmente na área de negócios. A aplicação destes no estudo de doenças neurodegenerativas usando dados de neuroimagens é relativamente recente (QUDDUS; FIEGUTH; BASIR, 2005; DAMANGIR *et al*., 2012; BENDFELDT *et al*., 2012). No tema específico de Esclerose Múltipla existem alguns trabalhos que utilizam técnicas de reconhecimento de padrões para análise das lesões (KAMBER *et al*., 1995; LAO *et al*., 2008; GEREMIA *et al*., 2011), melhorando o desempenho das classificações. Este trabalho se restringe somente ao uso de técnicas quantitativas e focadas em regiões específicas do cérebro. Também relacionado à Esclerose Múltipla existem alguns trabalhos de aplicação destes algoritmos usando dados de origem genética (BLONDA *et al*., 2002; WEYGANDT *et al*., 2011). A utilização de reconhecimento de padrões com técnicas quantitativas de neuroimagens em pacientes com esclerose múltipla não foi encontrada em pesquisa na literatura atual, sendo o assunto que pretendemos abordar neste trabalho.

1.1 Esclerose Múltipla

As doenças cerebrais são um grande problema para a saúde e bem estar das pessoas. Prejudicam atividades voluntárias e involuntárias do cérebro, causando uma desordem no equilíbrio das funções cerebrais. Com a maior longevidade das pessoas é cada vez maior o aparecimento dessas doenças devido à degeneração do organismo.

Dentre essas doenças a Esclerose Múltipla (EM) é uma doença neurodegenerativa caracterizada pela desmielinização inflamatória no sistema nervoso central (SNC) (KRITAS, *et al*., 2014). Seu código de classificação é G35 na Classificação Internacional de Doenças (CID-10). Ela é também considerada como autoimune crônica (HAUSER; GOODWIN, 2008).

Os órgãos de sentidos enviam informações ao cérebro e as ações e reações do corpo a estes estímulos envolvem um processo rápido e complexo. Esse processo depende da comunicação adequada entre as células nervosas. As sinapses que são conectadas pelas fibras nervosas ocorrem através de impulsos nervosos. Essas fibras nervosas são revestidas por uma camada gordurosa chamada bainha de mielina que torna possível a adequada condução dos impulsos nervosos.

Por ser considerada autoimune crônica, o próprio sistema imunológico do organismo ataca um tecido ou órgão normal. Na EM, o ataque imunológico dirige-se contra a mielina do SNC. O cérebro e a medula espinhal são os principais componentes do SNC, sendo o cerebelo e o tronco cerebral algumas das regiões afetadas pela EM (KALB, 2000).

A EM ataca os axônios mielinizados no SNC, destruindo a bainha de mielina que envolve os axônios (WEINSHENKER, 1996; OLEK, 2011). O curso da EM é altamente variável e imprevisível. Na maioria dos pacientes a doença é caracterizada por episódios de déficits neurológicos reversíveis que são muitas vezes seguidas por deterioração neurológica progressiva ao longo do tempo (GOLDENBERG, 2012). Com o passar do tempo ela causa a deterioração do encéfalo e medula espinhal pela repetição de surtos de desmielinização inflamatória multifocal, que leva a distúrbios de condução, destruição axonal e neuronal, com perda tecidual (SANTOS, 2007).

Os sintomas, decorrentes da desmielinização, podem incluir perda de visão, visão dupla, rigidez, fraqueza, falta de equilíbrio, dormência, dor, problemas no controle da bexiga e intestinos, fadiga, mudanças emocionais e comprometimento intelectual. O tipo e número de sintomas variam enormemente de um indivíduo para outro, dependendo do local, no SNC, onde ocorre dano à bainha de mielina (KALB, 2000). Estes sintomas estão frequentemente relacionados com o comprometimento de tratos longos, como as vias ópticas, vias sensitivas espino-corticais e motoras cortiço-espinais cerebelares (SANTOS, 2007).

Qualquer dano à bainha de mielina resulta em uma interferência na transmissão de impulsos nervosos do cérebro. Na EM as células imunológicas destroem algumas áreas da mielina e forma-se uma placa (lesão) com acúmulo de tecido enrijecido (esclerose) (KALB, 2000). Como esta situação ocorre em diversos locais do SNC origina-se o nome esclerose múltipla. Algumas dessas cicatrizes não causam sintomas visíveis, porém outras interferem em qualquer sensação ou função controlada naquela área específica do cérebro ou da medula espinhal. Além disso, a desmielinização pode interferir na comunicação entre diferentes partes do cérebro, mesmo que essas permaneçam intactas.

As lesões são compostas por domínios de mielina e oligodendrócitos juntamente com a perda de infiltrados inflamatórios (HAUSER; GOODWIN, 2008). O aspecto patológico típico das lesões é caracterizado por infiltrado inflamatório, contendo linfócitos T, macrófagos, linfócitos B, e ainda, por axônios desmielinizados circunjacentes às vênulas, particularmente, das regiões periventriculares (TRAUGOTT; REINHERZ; RAINE, 1983; LUCCHINETTI, 2008). As lesões da EM são caracterizadas por pequenos focos de desmielinização, semelhantes a placas, com distribuição perivenular, cuja composição varia

com a fase aguda ou crônica da placa em questão. Nas placas agudas encontram-se polimorfos nucleares e macrófagos, sendo os últimos responsáveis pela fagocitação dos restos de mielina. Nas placas crônicas encontra-se cavitação e proliferação astroglial. Macroscopicamente as placas agudas são lesões rosadas e amolecidas e as crônicas acinzentadas (LASSMANN, 1999).

A perda de função parece estar relacionada à interrupção ou lentificação da transmissão do impulso nervoso ligada à destruição da mielina, porém também ocorre a destruição de oligodendrócitos e de axônios (WAXMAN, 1998).

A EM é uma doença neurológica crônica, que aparece mais frequentemente em adultos jovens, sendo 90% de casos diagnosticados entre 16 a 60 anos. É mais comum em indivíduos de cor branca do que em hispânicos ou negros, e é relativamente rara entre os asiáticos e alguns outros grupos (KALB, 2000). As mulheres são afetadas em dobro em relação aos homens e as pessoas descendentes de europeus do norte parecem ter mais risco para desenvolver a doença (CREE, 2007; HAUSER; GOODWIN, 2008). A EM é mais comum em áreas de clima temperado, e é relativamente pouco comum nos trópicos. Embora não seja herdada diretamente, a doença aparece em indivíduos geneticamente pré-dispostos, os quais são aparentemente mais reativos ou suscetíveis a qualquer estímulo ou agente que possa desencadear a doença. A conclusão geral é de que descendentes de pessoas do norte e centro da Europa, bem como aqueles cujos ancestrais se misturaram com esse grupo, parecem ter alguma pré-disposição genética para a doença (KALB, 2000).

A EM tem uma aparente distribuição familiar, com uma frequência dez vezes maior em parentes de primeiro grau de pacientes com EM do que na população em geral. Este dado poderia ser explicado por algum fator genético, mas também por algum fator ambiental, eventualmente relacionado com um agente infeccioso (SANTOS, 2007).

Os casos de EM são muito diferentes entre si. O curso da doença é específico para cada paciente. A inexistência de um padrão clínico específico leva ao desconhecimento por parte do paciente sobre a evolução da doença. Este fator pode gerar um descontrole emocional, aumentando ainda os problemas físicos já causados pela doença (KALB, 2000). A doença produz dano muito precoce, o que é mais bem avaliado pela mensuração do N-Acetil-Aspartato (NAA), na fase com pequena desabilidade clínica. Na fase progressiva são mais eficientes as modalidades sensíveis ao dano tecidual, como a medida da atrofia e cálculo da taxa de transferência de magnetização (MTR, do inglês *Magnetization Transfer Ratio*), tanto nas lesões, quanto na substância branca (SB) e substância cinzenta (SC) aparentemente normal (SANTOS, 2007).

A EM é diagnosticada com base em resultados clínicos e provas de exames auxiliares de imagem como a ressonância magnética e análise do líquido cefalorraquidiano ou líquor. A doença apresenta-se mais tipicamente em adultos de 20 a 45 anos de idade e ocasionalmente apresenta-se na infância ou mais tarde (CREE, 2007). A causa é desconhecida, mas parece envolver uma combinação de susceptibilidade genética e um gatilho não genético tal como um vírus, o metabolismo ou fatores ambientais que em conjunto resultam em uma desordem autoimune e autossustentada que leva a ataques inumes recorrentes no SNC (CREE, 2007). Não há teste diagnóstico específico para EM (HAUSER; GOODWIN, 2008). O diagnóstico é com base na evidência de: (1) pelo menos duas diferentes lesões (placas ou cicatrizes) na SB do SNC; (2) pelo menos duas manifestações clínicas da doença; e (3) crônica inflamação do SNC, tal como determinado por análise do líquor. A presença de um ou mais destes critérios permite um diagnóstico geral de EM (GOLDENBERG, 2012).

A EM é uma doença progressiva e sem cura até o momento. Existem tratamentos disponíveis para gerenciar o curso da doença que são parcialmente eficazes no tratamento. Uma combinação de drogas, fala e terapias ocupacionais, exercícios físicos, descanso e nutrição saudável podem aliviar os sintomas e promover uma qualidade de vida satisfatória (GOLDENBERG, 2012).

1.2 Ressonância Magnética

Diagnosticar com precisão doenças neurológicas apenas com exames manuais ou testes psicológicos fica muito difícil. Torna-se necessário a visualização e interpretação de um exame de imagem de qualidade para um estudo adequado das patologias do SNC. Nos últimos anos, as tecnologias de aquisição de imagens radiológicas têm se desenvolvido cada vez mais com a geração de imagens de excelente definição que têm sido cada vez mais utilizadas pelos profissionais da saúde. Exames radiológicos do cérebro podem identificar alterações na anatomia para que seja direcionado o tratamento médico. Dentre as modalidades de exames do cérebro destaca-se a ressonância magnética (RM).

As imagens por RM são muito utilizadas para o diagnóstico e tratamento de doenças cerebrais por proporcionar imagens bem detalhadas das estruturas cerebrais. Além disso, o exame não utiliza radiação ionizante e por isso não prejudica o organismo. A construção da imagem de ressonância baseia-se na interação magnética dos campos magnéticos gerados pelo tomógrafo de RM e os spins nucleares presentes nos tecidos. Nas imagens convencionais são utilizados os prótons (^{1}H) contidos nos núcleos de hidrogênio do organismo.

A imagem por RM apresenta uma sensibilidade muito boa para mostrar lesões desmielinizantes do cérebro, porém apresenta baixa especificidade, não tendo capacidade de diferenciar edema, desmielinização, remielinização, gliose e lesão axonal, permanecendo restrita a uma análise qualitativa do sinal. Porém, o potencial enorme de informação intrínseca no sinal de RM a partir de diferentes sequências de pulsos, levou ao desenvolvimento de diferentes metodologias quantitativas, cada uma contemplando aos menos um aspecto biológico envolvido no processo fisiopatológico da produção da lesão tecidual. Aos poucos foram se somando à volumetria de estruturas, a maneira mais elementar de quantificação, a mensuração do tempo de relaxação (relaxometria), da quantificação de metabólitos (espectroscopia), da transferência de magnetização, da difusão das moléculas de água, etc. Com estas técnicas tornou-se possível a quantificação de fenômenos biológicos *in vivo* (SANTOS, 2007).

O uso intensivo de técnicas quantitativas em RM produziu uma mudança de paradigma na radiologia e elevou a importância científica da imagem médica, gerando uma grande quantidade de artigos científicos (SANTOS, 2007). Com as técnicas quantitativas podem ser realizadas imagens específicas para melhor visualização e estudo das doenças. Com estas técnicas é possível estudar a doença por vários ângulos, permitindo o melhor tratamento de doenças complexas. Essas imagens podem ser utilizadas para tratamento de doenças e por pesquisadores para o desenvolvimento de suas pesquisas. Mas apenas visualizar as imagens pode não ser suficiente para investigar as doenças. Muitas vezes torna-se necessário quantificar os achados médicos de imagens para obter um estudo mais completo das patologias, o que nem sempre é simples de se realizar.

A ressonância magnética convencional, devido à sua grande sensibilidade para detectar lesões na substância branca, tornou-se poderosa aliada do diagnóstico e seguimento dos pacientes com EM. Porém, na esclerose múltipla, as lesões visíveis nas imagens ditas convencionais, ponderadas nos tempos de relaxação (T1 e T2) e densidade de prótons, mostram somente as lesões maiores, praticamente aquelas macroscópicas, com alguns milímetros de tamanho, deixando ocultas as lesões microscópicas (SOUZA, 2010). Na RM a placa na fase aguda, com atividade inflamatória, é vista como uma área oval de aumento de sinal em sequências ponderadas em T2, com halo de edema ao redor e com realce após a administração de agente de contraste com gadolíneo. Devido à distribuição periventricular do infiltrado inflamatório, as lesões se distribuem preferencialmente de maneira perpendicular aos ventrículos, ou na transição entre a SB e SC (SANTOS, 2007).

Desde o início na década de 90 têm sido utilizadas sequências de pulsos de RM mais sofisticadas, associadas a métodos de pós-processamento que fornecem dados quantitativos que permitem mensurar as alterações microscópicas descritas anteriormente, não visíveis nas imagens convencionais. Dentre essas técnicas, as mais conhecidas e difundidas são: a volumetria cerebral, relaxometria, a taxa ou razão de transferência de magnetização, a espectroscopia de prótons e a difusão. As alterações detectadas nos pacientes com EM, tanto em estudos microscópicos quanto por técnicas quantitativas de RM são influenciadas pela duração da doença, forma clínica e idade (SOUZA, 2010). Tomada de maneira isolada, nenhuma modalidade de imagem RM qualitativa ou quantitativa é suficiente para providenciar uma compreensão ampla no processo de instalação da lesão tecidual da EM. Por outro lado, quando combinamos as modalidades de RM qualitativas e quantitativas juntas, se tornam uma poderosa ferramenta de investigação (SANTOS, 2007). Outro problema importante no acompanhamento da EM é quantificar os danos causados no cérebro pela doença e não somente na área das lesões. O cérebro por ser uma rede complexa interconectada sofre danos em regiões não necessariamente próximas às lesões.

1.3 Centro de Imagens

O Centro de Ciências da Imagem e Física Médica (CCIFM) através do Laboratório de Processamento de Imagens e Informática Radiológica (LAPIR) no Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto (HCFMRP) da Universidade de São Paulo (USP) disponibiliza um ambiente de pesquisa para o processamento e análise de imagens médicas de ressonância magnética. Uma das atividades do processamento de imagens é a geração de dados quantitativos a partir dessas técnicas de imagens. Esses dados são gerados por programas e *scripts* desenvolvidos através de uma rotina de trabalho. As imagens e os dados resultantes são enviados para um servidor de pesquisa. Os dados representam medidas das estruturas cerebrais e ficam armazenados em um banco de dados divididos por projetos. A base de dados de pesquisa é constituída de informações de indivíduos saudáveis e pacientes. A base de saudáveis é do HCFMRP, sendo utilizada como referência em comparativos com dados de indivíduos doentes de diferentes projetos de pesquisa. Na figura 1.1 é mostrado um diagrama do fluxo de trabalho do CCIFM.

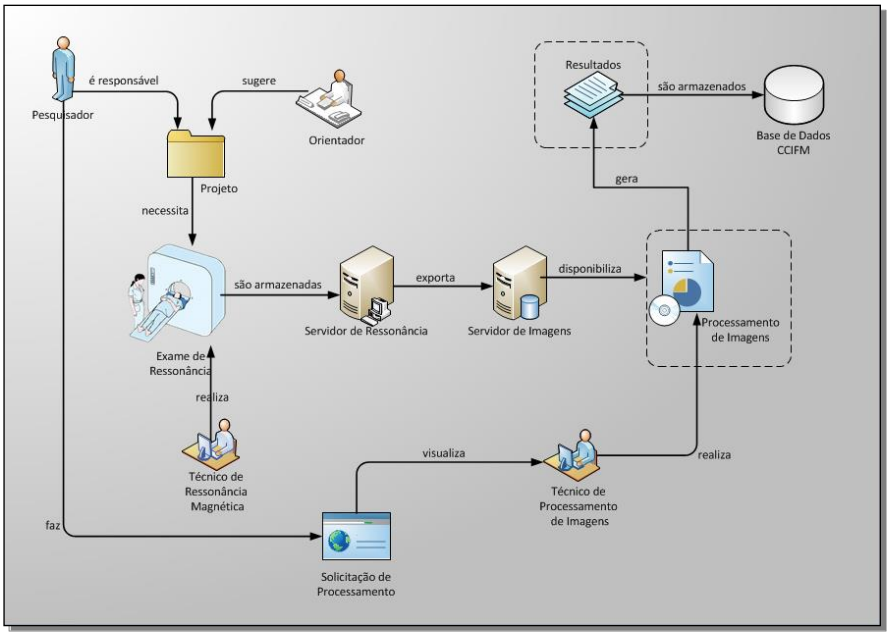

Figura 1.1 – Diagrama de Fluxo de Trabalho do CCIFM.
Fonte: Elaborada pelo autor.

Inicialmente o pesquisador e o orientador criam o projeto com todos os protocolos definidos. Depois são feitos exames de ressonância magnética nos participantes do projeto por um técnico de ressonância. As imagens são armazenadas em um servidor de ressonância e posteriormente exportadas para um servidor de imagens. Em seguida o pesquisador faz a solicitação de processamento das imagens que é visualizada pelo técnico de processamento de imagens. Este realiza o processamento de imagens e gera os resultados que são enviados para o pesquisador e armazenados no banco de dados do CCIFM. Os técnicos também realizam os processamentos de imagens para auxílio a tratamentos de pacientes do HCFMRP de acordo com as solicitações os profissionais da saúde.

Para a extração de dados quantitativos a partir das imagens por RM são utilizados programas de segmentação automática de imagens, de segmentação manual, de geração de gráficos de espectroscopia e *scripts* específicos para geração de dados. O principal programa utilizado na segmentação automática de estruturas corticais e subcorticais é o FreeSurfer desenvolvido pela Athinoula A. Martinos Centro de Imagem Biomédica no Hospital Geral em Massachusetts (FreeSurfer Software Suite - versão atual 4.5) (SCHMANSKY, 2015). Ele contém ferramentas para construção de modelos topologicamente corretos e geometricamente precisos das superfícies de substância cinzenta, branca e *pial*, para medir a espessura cortical e área da superfície. Como resultado desta segmentação são gerados arquivos (.mnc) na forma de *labels* que podem ser facilmente usados por outros programas.

O programa Display 1.5.1 das ferramentas MINC 2.2 é utilizado para visualização e segmentação manual de imagens. Este programa contém ferramentas de desenho apropriadas para a manipulação e segmentação da imagem com a posterior geração de arquivos (.lbl) de *labels*.

Por fim existem *scripts* específicos criados para conversão de imagens e geração de dados a partir de segmentações automáticas ou manuais. Estes scripts são modificados de acordo com a necessidade de novos estudos. As imagens resultantes do processamento ficam armazenadas em um servidor de pesquisa juntamente com os dados que ficam armazenados em um banco de dados separados por projeto. Na figura 1.2 é mostrado um esquema padrão para o processamento de imagens.

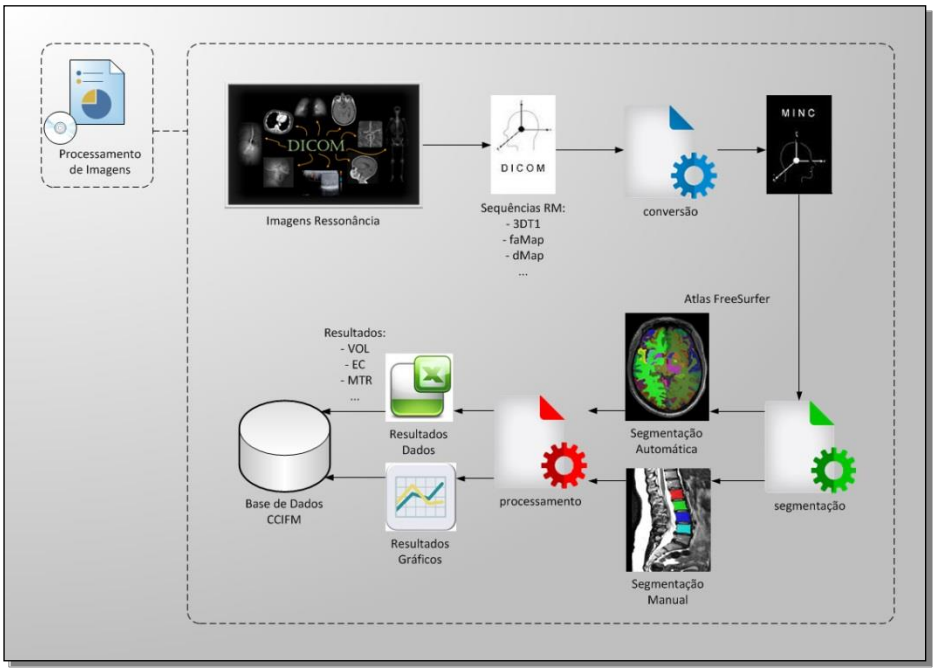

Figura 1.2 – Esquema Padrão de Processamento de Imagens.
Fonte: Elaborada pelo autor.

As imagens adquiridas pela ressonância magnética são exportadas do aparelho no formato *"Digital Imaging and Communications in Medicine"* (DICOM). Atualmente, uma conversão para o formato *Medical "Image NetCDF"* (MINC) é realizada para que seja possível o seu processamento. Depois é feita a segmentação das regiões de interesse que pode ser feita automaticamente ou manual, de acordo com a pesquisa. Em seguida são realizados processamentos específicos nas imagens e gerados os dados quantitativos ou gráficos como a espectroscopia. Por fim os resultados são armazenados na base de dados do CCIFM para consultas e pesquisas.

Essa base de dados armazena informações de projetos e controles normais (CN). São quase dez anos de rotina de envio de dados e nesse período houve mudanças tanto na aquisição quanto no processamento de imagens, o que pode ter acarretado problemas na integridade desses dados. Analisando melhor esses dados, foram encontradas divergências que serão mais bem detalhadas nos critérios de inclusão e exclusão de dados vistos mais adiante.

1.4 Técnicas Quantitativas

A ressonância magnética nos últimos vinte anos tem incrementado o número de técnicas quantitativas disponíveis. As sequências de pulsos usadas na máquina de ressonância já são feitas especificamente para cada técnica quantitativa. No quadro 1.1 são mostradas algumas medidas quantitativas utilizadas e a nomenclatura dos dados no formato DICOM e MINC gerados pelas respectivas sequências ou técnicas quantitativas usada no CCIFM.

Quadro 1.1 – Nomenclatura das imagens, usada no CCIFM, e as respectivas medidas geradas pelas sequências quantitativas.

Sequências Quantitativas		
Formato DICOM	**Formato MINC**	**Medida**
3DT1	MPR	Volumetria Espessura Cortical
3DMTR	D1 e D2	MTR
Relax	E1...E5	Relaxometria
faMap e dMap	faMap e dMap	Anisotropia Fracionada e Difusibilidades
Flairvista	Flair	Volumetria de Lesões

Fonte: Elaborada pelo autor.

A partir de cálculos baseados em informações de cada ponto da imagem (*voxel*) é possível realizar processamentos que permitem extrair característica do dado (técnica quantitativa adquirida na máquina de RM) e fornecer novas informações a respeito da estrutura e função cerebral. Muitas vezes esta informação não é acessível pela análise convencional do radiologista (OLIVEIRA JUNIOR, 2008). A seguir são descritos os parâmetros a serem analisados neste estudo:

- **Anisotropia Fracionada** – é um valor escalar entre zero e um que descreve o grau de anisotropia da difusão das moléculas de água. Um valor zero significa que a difusão é isotrópica, isto é, sem restrições ou igualmente restrito em todas as direções. Um valor

unitário indica que a difusão ocorre apenas ao longo de um eixo e é totalmente restrito ao longo de todas as outras direções. Ela é uma medida frequentemente usada em imagens ponderadas por difusão, representando: densidade de fibras, diâmetro axonal, e mielinização na substância branca (BASSER; PIERPAOLI, 1996);

- **Coeficientes de Difusão** – no caso específico da RM deve ser considerada a autodifusão das moléculas de água pelo movimento *browniano*. A difusão é espontânea e depende das propriedades do meio; pela complexidade do meio biológico o parâmetro característico da difusão é chamado de coeficiente de difusão aparente (ADC das siglas em inglês). Este coeficiente é normalmente representado pelas suas componentes, que no caso de simetria axial, como a aqui considerada, são duas: paralela (maior intensidade) e perpendicular. A relação entre ambas as componentes definem o grau de anisotropia do *voxel* em questão. A difusão é medida através de um ajuste exponencial;
- **Espessura Cortical** – é uma medida morfométrica utilizada para descrever a espessura combinada das camadas do córtex cerebral (THORSTENSEN, 2006). Dado que a espessura cortical aproximadamente se correlaciona com o número de neurônios dentro de uma coluna ortogenética, é muitas vezes tomado como indicativo da capacidade cognitiva de um indivíduo, embora este último seja conhecido por terem múltiplos determinantes;
- **Taxa de Transferência de Magnetização** – reflete um fenômeno de troca de energia entre moléculas do tecido. A imagem de RM convencional é baseada apenas na porção de tecido com tempo de relaxação relativamente longo, na ordem de 10 a 100 milissegundos, devido à maneira como as sequências são desenhadas. Assim ficam virtualmente invisíveis as proteínas e moléculas complexas que possuem tempo de relaxação em torno de microssegundos, explicando a insensibilidade para lesões microscópicas. Para contornar essa deficiência, o fenômeno da transferência de magnetização é utilizado para viabilizar lesões nas macromoléculas. Um pulso de radiofrequência direcionado para estas macromoléculas é utilizado para saturá-las e técnicas de pós-processamento permitem a quantificação da energia transferida. Quando as macromoléculas estão integras, a taxa de transferência é alta. Quando existem lesões e desestruturações destas macromoléculas, a taxa de transferência de magnetização cai de uma forma relativamente proporcional. Assim, o método denominado de razão de transferência de magnetização é utilizado para uma semiquantificação da taxa de lesão macromolecular, que, no SNC significa

desmielinização e lesão axonal na SB e perda de membrana citoplasmática na SC (KUCHARCZYK *et al*., 1994). Fornece uma estimativa da gravidade de algumas doenças do cérebro em função do conteúdo macromolecular existente em cada *voxel*. Trata-se de uma importante técnica para avaliar a integridade mielínica na SB cerebral. Gera-se um contraste na imagem baseado na troca de energia entre um componente macromolecular e a água livre (RORIZ, 2010);

- **Relaxometria** – consiste na medida de tempo de relaxação longitudinal T1 ou transversal T2 de uma imagem, baseado na utilização de uma série de sequências com variação progressiva do tempo de repetição (TR), para relaxometria T1, ou do tempo de eco (TE) para relaxometria T2. Após as aquisições, métodos de pós-processamento permitem a geração de mapas de T1 ou T2 quantitativo que, numericamente, permitem a quantificação da concentração de água tecidual e a detecção de alterações que se correlacionam com desmielinização, necrose, edema, gliose ou lesão axonal. Portanto, a relaxometria permite mensurar aumentos focais de água, muitas vezes não visíveis em sequências T1 e T2, mas que, apesar disto, fornecem informações relativamente pouco específicas (SOUZA, 2010). Quantifica tempo de relaxação dos spins nucleares, fenômeno físico em que ocorre troca de energia, T1 e T2 e outras de uma imagem. Nas imagens por RM, um aumento no tempo de relaxação reflete um aumento da concentração de água nos tecidos, que pode indicar: edema, desmielinização, gliose ou disfunção axional. Quando a lesão apresenta prolongamento apenas de T2, em geral, isto indica que há a ocorrência apenas de desmielinização (GIRALDI *et al*., 2008) (RORIZ, 2010). Por outro lado, o encurtamento de T2 para lesões indica depósito de hemossiderina (HAACKE *et al*., 2005). Também em Relaxometria, um *pixel* mais brilhante indica uma maior intensidade de sinal na imagem RM, ou seja, um ponto brilhante numa imagem ponderada em T2 significa tempo de relaxação T2 prolongado e um ponto brilhante numa sequência ponderada em T1 significa tempo de relaxação T1 curto;
- **Volumetria** – é a técnica quantitativa de imagens mais simples, que consiste, basicamente, da contagem de pontos da imagem (*pixels*) relacionados com a espessura do corte, que gera um elemento de volume (*voxel*). A partir deste elemento de volume, faz-se uma estimativa do volume da estrutura ou do órgão em estudo. Na maioria das vezes, a volumetria é utilizada para estudos de atrofia e fornece informações acerca da perda parenquimatosa com uma boa correlação com a perda celular neuronal e glial (RORIZ, 2010). A volumetria pode ser usada para estudos globais de atrofia, para

atrofia diferenciada de SC e SB, para atrofias regionais em lobos ou parte deles, ou para estudos isolados de estruturas como o corpo caloso, tracto piramidal, etc. Em estudos de atrofia, é fundamental levar em conta o processo normal de amadurecimento e envelhecimento do encéfalo, bem como as variações normais entre os sexos e entre indivíduos. Na esclerose múltipla, a volumetria tem sido utilizada para mensurar a carga da doença visível sob a forma de lesões detectáveis em sequências ponderadas em T2, lesões mais destrutivas visíveis em sequências ponderadas em T1 ou lesões com realce pelo contraste (SANTOS, 2007). A segmentação automática para volumetria se torna essencial devido a grande complexidade das regiões cerebrais, no entanto é passível de erros como segmentação de pequenas regiões que podem ser misturar SB e SC.

As técnicas quantitativas possibilitam a caracterização das diferentes regiões do cérebro, mas geram uma grande quantidade de resultados numéricos. São mais de mil atributos de regiões do cérebro que podem ser quantificados, tornando o estudo complexo e ineficiente com uma grande quantidade de dados. São muitos dados que precisam ser analisados e otimizados para que tenham algum significado importante na condução do estudo. O mais comum nos estudos de neuroimagem é a utilização de estatística paramétrica para caracterizar grupos de indivíduos, assim como a utilização de testes de hipóteses para realizar comparações. Porém, ante esta grande quantidade de dados que podem estar ou não fortemente correlacionados entre si, as técnicas de reconhecimento de padrões são uma excelente alternativa não somente na classificação de grupos, senão em sugerir o conjunto de variáveis que mais influenciam na identificação de um sujeito individual.

1.5 Técnicas de Reconhecimento de Padrões

O reconhecimento de padrões é um subtópico da aprendizagem de máquina que por sua vez é um subcampo da inteligência artificial. As técnicas de reconhecimento de padrões têm sido utilizadas nos últimos quinze anos em diferentes áreas. Principalmente na área de negócios são feitos estudos de mineração de dados para descobrir estratégias de melhoria em vendas e gestão de negócios. Essas técnicas permitem o agrupamento de uma grande quantidade de resultados numéricos, permitindo ao gestor tomar melhores decisões nos seus negócios. Na área médica, essas técnicas têm ganhado espaço principalmente para analisar dados de medidas quantitativas de imagens médicas. Especificamente em imagens médicas existe uma nova ferramenta que permite e auxilia os médicos na tomada de decisões de

diagnósticos e tratamentos que são as classificações de dados. Existem diversos tipos de técnicas de reconhecimento de padrões ou classificadores, mas neste projeto abordamos especificamente as técnicas KNN e SVM que serão abordadas brevemente nos itens 1.5.1 e 1.5.2. Utilizamos o KNN pela sua simplicidade e praticidade e o SVM porque vendo sendo muito utilizado atualmente em trabalhos parecidos.

1.5.1 *K-Nearest-Neighbors (*KNN*)*

O *K-Nearest-Neighbors* (KNN) é um modelo de indutor muito popular e frequentemente utilizado na literatura (SHIMO, 2009). Trata-se de um algoritmo para geração de classificadores partindo de uma coleção de exemplos, que segue o paradigma de Aprendizado Baseado em Instâncias. Nesta classe de indutores, inicialmente é armazenado o todo ou parte (utilizando alguma heurística de exclusão) de um conjunto de exemplos conhecidos, chamado conjunto de treinamento. Para classificação, um exemplo desconhecido é comparado com o conjunto de treinamento e sua classe será definida pelos exemplos que apresentem maior similaridade de acordo com uma determinada função. A métrica mais comumente utilizada é a distância de Minkowski:

$$dist_p(A,B) = \sqrt[p]{\sum_i |a_i - b_i|^p} \qquad (1)$$

Onde a_i e são os i-ésimos atributos dos exemplos A e B respectivamente, e p é um parâmetro de forma da distância. Intuitivamente, no caso de funções de distância, apresenta maior similaridade o par que possuiu menor valor de distância.

A classificação no KNN considera uma vizinhança tal que a classe de um exemplo desconhecido é baseada nas classes das amostras do conjunto de treinamento contidas nesta vizinhança. Esta é definida por um parâmetro k do algoritmo que representa o número de exemplos dentro da vizinhança. A política de classificação considera a classe do novo exemplo como a classe majoritária dos k vizinhos.

Idealmente o teste de classificadores é feito com dois conjuntos isolados. Um utilizado para o treinamento do classificador e um conjunto separado para validação. Casos em que a amostra não é suficientemente grande são comuns o emprego de técnicas de amostragem. O *stratified r fold cross-validation* é uma técnica de amostragem que divide aleatoriamente o conjunto em *r* partes (ou *folds*) independentes, isto é, sem intersecções, com um mesmo tamanho e mantendo a proporção entre o número de exemplos de cada classe (estratificação).

O treinamento é realizado em *r-1 folds* utilizando o remanescente para teste. Este processo é realizado *r* vezes utilizando em cada uma um *fold* diferente como conjunto de teste. A taxa de erros é obtida através de um erro médio entre os *r folds r*=10, aplicado neste trabalho, sendo o valor padrão do programa WEKA e é frequentemente utilizado na literatura (PESSOTI, 2008; SHIMO, 2009). No final da geração dos classificadores para os *r folds* da validação cruzada é feita uma análise da matriz de confusão, resultado do cruzamento das classes verdadeiras com as classes preditas pelo classificador.

Independente do classificador usado, a partir da matriz de confusão são extraídas duas medidas importantes para a avaliação de classificadores:

$$\text{Sensibilidade} = \frac{v_p}{v_p + f_n} \qquad (2)$$

$$\text{Especificidade} = \frac{v_n}{v_n + f_p} \qquad (3)$$

Onde v_p é a quantidade absoluta de verdadeiros positivos, isto é, exemplos classificados como positivos (classe Normal), e de fato serem positivos; f_p a quantidade de falsos positivos, exemplos classificados como positivos, quando, na verdade, são negativos (classe Paciente); f_n é a quantidade de falsos negativos, exemplos classificados como negativos, enquanto são positivos; v_n é a quantidade de verdadeiros negativos.

Com a sensibilidade e a especificidade, pode-se utilizar outro método de avaliação dos classificadores chamado de curva "*Receiver Operating Characteristic"* (ROC). A curva ROC é um gráfico da variação da sensibilidade versus (1-especificidade) para diferentes pontos de corte ou limiares. Cada ponto do gráfico (Sensibilidade,1-Especificidade) é obtido a partir de uma matriz de confusão. A curva é obtida através da variação de um limiar de separação da classe entre positivo ou negativo de modo que com um limiar mínimo todos os exemplos seriam classificados como negativos (ponto (0,0) da curva), e em um limiar máximo todos os exemplos seriam classificados como positivo (ponto (1,1) da curva). Uma medida retirada deste gráfico e muito utilizada como indicador de desempenho em algoritmos de classificação é a área abaixo sob a curva ROC (Área *Under* ROC – AUROC A_z). Desse modo, o classificador ideal, para um determinado conjunto de exemplos, é aquele que possui especificidade e sensibilidade unitárias para qualquer ponto de corte, e, portanto uma área unitária.

1.5.2 *Support Vector Machines (*SVM*)*

O *Support Vector Machines* (SVM) é essencialmente um classificador de duas classes que maximiza a largura da margem entre as classes, isto é, a área vazia em torno do hiperplano de separação definido da distância para as amostras de formação mais próximas (GIRALDI *et al*., 2008). Pode ser alargado a problemas de múltiplas classes, resolvendo essencialmente vários problemas de duas classes.

Dado um conjunto de treinamento que consiste em N pares de (x_1, y_1), (x_2, y_2)... (x_n, y_n), onde x_1 denotam as observações de formação n-dimensional e $y_1 \in \{-1,1\}$ são os rótulos de classificação correspondentes. O método *SVM* procura encontrar o hiperplano definido por

$$f(x) = (x \cdot w) + b = 0 \qquad (4)$$

que separa observações positivas e negativas com a margem máxima. O vetor w e o valor limite escalar b determinam a orientação e a posição do hiperplano de separação. Pode ser mostrado que o vetor solução W_{svm} é definido em termos de uma combinação linear das observações de formação, isto é,

$$W_{svm} = \sum_{i=1}^{g} \alpha_i \, y_i x_i \qquad (5)$$

onde α_i são coeficientes de Lagrange não negativos obtidos através da resolução de um problema de otimização quadrática com restrições de desigualdades lineares. Essas observações de treinamento x_i com α_i diferente de zero encontram-se no limite da margem e são chamados vetores de suporte.

Caso os padrões não sejam linearmente separáveis, uma das vantagens do SVM é que possibilita uma mudança da dimensionalidade através da aplicação de *kernels* não lineares.

1.6 Justificativa e Objetivos

Na extração de dados através da segmentação automática utilizando o programa FreeSurfer são gerados muitas medidas de diferentes estruturas do cérebro. Atualmente no LAPIR utiliza-se um total de 158 a 1000 medidas por caso estudado. O número exato em cada caso depende das técnicas de aquisição aplicadas e o número de regiões cerebrais segmentadas. Muitos valores dessas medidas são próximos tanto para indivíduos saudáveis quanto para pacientes. A definição a "olho nu" dos melhores atributos para caracterização de cada doença cerebral é quase impossível.

Na rotina do CCIFM utilizamos técnicas de estatística focadas na comparação entre grupos de indivíduos das médias e dos desvios padrões destes atributos, muitas vezes a interação entre eles é desconsiderada. Adicionalmente, o excesso de atributos desnecessários tende a diminuir o desempenho dos resultados das análises dos dados e atrasar a entrega dos dados pós-processados pelos técnicos responsáveis para os pesquisadores. Portanto há a necessidade de otimização dessas análises e uma possível solução para isso seria o agrupamento dos melhores atributos de cada doença cerebral através da comparação entre os classificadores.

Como foi comentado, a Esclerose Múltipla é uma doença com alterações difusas no cérebro do paciente e sem uma causa definida que tem sido bastante estudada em diversos projetos associados ao CCIFM.

O propósito geral deste estudo foi aplicar técnicas de reconhecimento de padrões em dados quantitativos de neuroimagens adquiridas por ressonância magnética, em especial de pacientes de Esclerose Múltipla. Duas foram as perguntas básicas a serem respondidas: Quais são as principais variáveis quantitativas para a separação dos grupos controle saudável e paciente? É possível definir uma estratégia de classificação que separe estes grupos? Ainda esperamos sugerir qual dos algoritmos de classificação aqui avaliados é mais conveniente para o problema em questão.

2. MATERIAIS E MÉTODOS

2.1 Casuística

Este projeto segue a resolução 466/2012 CNS/MS e foi aprovado em 21/10/2015 pelo Comitê de Ética em Pesquisa (CEP) do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto (HCFMRP-USP) sob Ofício nº 3778/2015 e sob processo HCRP nº 14124/2015 (Anexo 1) e posterior versão 2 (Anexo 2).

A aquisição de imagens de RM foi realizada em equipamento supercondutor de alto campo, Achieva 3T da Philips, tendo por base o protocolo para os controles normais do HCFMRP. Quatro tipos de sequências foram definidos, denominados 3DT1, DTI, MTR e RELAX. O protocolo abrange modalidades de neuroimagens com potencial para atender a estudos de diferentes interesses morfológicos e fisiológicos da esfera celular-molecular ao anatômico-estrutural (CARMO, 2014).

A sequência 3DT1 é uma neuroimagem anatômica tridimensional ponderada no tempo de relaxação longitudinal (T1). Esta sequência é ponderada em T1, com tempo de repetição (TR) = 7,0 ms e um tempo de eco (TE) = 3,2 ms, além de uma sequência pulsada de gradiente-eco, *turbo field echo* (TFE) com intervalo de tiro *shot interval* (SI) = 2000 ms, e um ângulo de giro de 8º além do pulso de inversão com atraso de *inversion delay* = 900 ms. Para otimizar a matriz de aquisição escolhe-se uma orientação sagital e um campo de visão, *field of view* (FOV) de 240x240x180mm suficiente para cobrir toda a cabeça. Deste modo, define-se a dimensão da matriz em 240x240x180 para se assegurar uma resolução isotrópica de 1 mm^3, gerando-se 180 imagens de fatias sagitais.

A sequência do tipo *Diffusion Tensor Imaging* (DTI) foi usada para quantificação do tensor de difusão necessário para os mapas de difusibilidades ponderados em coeficientes como a anisotropia fracionada e a difusibilidade média. Esta sequência é fundamental para a análise de conectividade cerebral por índices anatômicos e da difusibilidade anatômica comparada em análise de grupos pela estatística espacial baseada nos tratos. Trata-se de uma sequência spin eco combinada com um aquisição echo planar imaging (EPI) com um EPI factor=67 e com 32 gradientes de difusão b = 1000 s/mm^2. Sobre o tamanho do *voxel*, temos um cubo de 2 mm de aresta, haja vista que a aquisição é de 72 (setenta e dois) cortes axiais de 2 mm de espessura cada, com um FOV de 256x256 mm em cada fatia em uma matriz de 128x128 *pixels*, ou seja, temos um *voxel* isotrópico de 8 mm^3. Os valores TR / TE = 9300 ms/

54 ms e o tempo total de aquisição está determinado em 06 minutos e 15 segundos no aparelho.

A sequência de transferência de magnetização foi definida com um TR igual a 7,3 ms e um TE igual a 3,5 ms. A cobertura completa tem 96 aquisições axiais, sendo 2 aquisições por fatia e cada fatia possui uma resolução espacial de 1 mm² de área e 3 mm de espessura. A sequência usa cerca de 2 minutos e 50 segundos para a primeira cobertura e mais 2 minutos e 50 segundos para a última cobertura, já com o sinal de magnetização transferida. Esta técnica requer, tal como a imagem para relaxometria e a imagem por tensor de difusão, um pós-processamento, neste caso sobre as aquisições de mesma fatia captadas pela transferência de magnetização para gerar os mapas com os valores das taxas de transferência, usa-se $MTR = (M_0 - M_t)/M_0$. Sendo $M_{0,t}$ as intensidades do pixel na aquisição sem e com pulso de transferência de magnetização, respectivamente.

A sequência de relaxometria trata-se de uma modalidade de aquisição para a quantificação do tempo de relaxação transversal (T2). No aparelho esta sequência é chamada *Gradient Spin-Echo* (GraSE) no modo multi-cortes resultado da combinação de uma sequência *Turbo Spin Eco* (TSE, fator = 5) com uma sequência *Echo Planar Imaging* (EPI, fator = 5). Nesta sequência, ajustam-se os 05 (cinco) valores do tempo de eco que são eles: 20, 40, 60, 80 e 100 ms. Todos otimizados para se assegurar uma boa reprodução do decaimento exponencial de T2 na faixa de 40 ms a 100 ms, onde estão contidos os valores esperados para a substância branca e cinzenta. O tempo de repetição necessário é de 3000 ms, para se cobrir todo o encéfalo, ajusta-se um FOV de 240x180x144 com 48 (quarenta e oito) cortes axiais de 3 mm de espessura cada e sem separação entre eles. Visando uma maior precisão espacial na quantificação mantém-se a resolução espacial no plano de 1 mm.

Foram utilizados os dados de 144 pacientes EM e 203 indivíduos saudáveis que já realizaram exames de ressonância magnética do encéfalo em um equipamento de 3T (Achieva, Philips) instalado no HCFMRP e dos quais foram gerados os dados quantitativos entre os anos de 2008 e 2014, totalizando 347 participantes. Em ambos os grupos a faixa etária foi de 7 a 65 anos. Cada paciente foi classificado seguindo a confirmação clínica existente. Neste trabalho foi escolhida a doença neurológica como uma das que possuem maior número de indivíduos na base de dados de projetos de pesquisa do CCIFM que é a EM. Além disso, é reconhecido na literatura que essa doença atinge áreas conhecidas no cérebro como a substância branca, sendo importante para a classificação às regiões lesionadas estarem previstas. Esta condição faz-se necessária para a correta aplicação dos algoritmos de

reconhecimento de padrões a serem avaliados. A quantidade de indivíduos em cada grupo de estudo está exibida no quadro 2.1.

Quadro 2.1 – Dados demográficos dos casos analisados.

Grupo	N	Idade (anos)		Sexo (M/F)
		Faixa	Média (DP)	
Pacientes EM	144	7-65	41,5 (11,6)	44/100
Controles	203	7-65	29,3 (16,2)	108/95

Fonte: Elaborada pelo autor.

Em relação a sexo as maiores diferenças são nos volumes. Nos valores quantitativos não existem diferenças de variações devido ao sexo, somente na volumetria. Porém a volumetria foi normalizada em relação ao volume total, para atenuar ou reduzir este efeito. Em relação à idade existe a diferença entre os grupos, pois quisemos utilizar a maior quantidade de dados possíveis, no entanto incluímos a idade como um dos atributos na classificação.

Dentre os 144 pacientes de EM foi realizada uma pesquisa nos resultados dos laudos dos exames de RM no módulo Radiologia 2 do HCFMRP para descobrir o grau de severidade da doença em cada caso. Os resultados foram divididos de acordo com o grau de destrutibilidade da lesão EM. Foram considerados nos laudos os graus de destrutibilidade "baixa", "moderada" e "alta". De acordo com a figura 2.1 dentre os pacientes EM considerados na amostra 12% possuem lesões de baixa destrutibilidade, 32% possuem grau de destrutibilidade moderada e 56% possuem alto grau de destrutibilidade, configurando como sendo mais da metade dos casos analisados. Desta forma podemos afirmar que a população de pacientes foi maioritariamente constituída por indivíduos com a doença em estado moderado ou avançado.

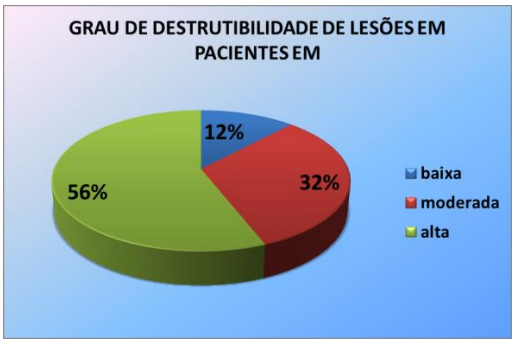

Figura 2.1 – Grau de destrutibilidade das lesões nos pacientes de EM.
Fonte: Elaborada pelo autor.

2.2 Dados Quantitativos

Para geração dos dados quantitativos foram utilizados diferentes programas de processamento de imagens. Para a segmentação automática foi utilizado o programa FreeSurfer versão 4.5 utilizando a função *recom-all*. Foram segmentadas automaticamente em torno de 200 regiões do cérebro, mas usamos 126 regiões para este trabalho (Apêndice H). Ao longo do trabalho as regiões estão escritas na forma original como aparecem no FreeSurfer por maior simplicidade.

Para a geração de alguns mapas quantitativos foram utilizados *scripts* desenvolvidos no CCIFM na linguagem Perl e utilizando as ferramentas MINC. Para a geração dos dados de Volumetria e Espessura Cortical foi utilizado o script FreeSurfer_DataN.pl. Para a geração dos dados das outras técnicas restantes utiliza-se as *"regions of interest"* (ROIS) gerados pelo FreeSurfer: aparc+aseg.mnc, aseg.mnc e wmparc.mnc. sendo aparc as estruturas subcorticais da SC e SB e aseg a subdivisão do córtex. O mapa de MTR foi gerado com os scripts DEMMT e MAPS.

Para as estruturas cerebrais contendo majoritariamente SB foram analisadas as medidas de Volume (VOL), Relaxometria (REL), Taxa de Transferência de Magnetização (MTR, do inglês *Magnetization Transfer Ratio*), Anisotropia Fracionada (ANF) e Difusibilidades (Média (DME), Paralela (DPA) e Perpendicular (DPE)). Nas regiões com predominância de SC é adicionada a Espessura Cortical (ECT). Todas essas medidas formaram os atributos/características que foram analisados. Com esses dados foi gerado um vetor característica para cada imagem.

A mineração de dados é o processo de explorar uma grande quantidade de dados à procura de padrões consistentes para detectar relacionamentos entre variáveis extraindo novos conjuntos de dados. Os dados quantitativos foram organizados em atributos e introduzidos em um programa de mineração de dados para classificação e agrupamento, com e sem seleção dos melhores atributos. Os dados foram ordenados por idade seguidos por nome, para garantir maior facilidade no momento da classificação.

Cada técnica quantitativa possui uma unidade de valor diferente do outro. As unidades utilizadas são mostradas no quadro 2.2.

Quadro 2.2 – Unidades de medida das técnicas quantitativas.

Técnica Quantitativa	Unidade
Anisotropia Fracionada	Adimensional
Difusibilidades (Média, Paralalela e Perpendicular)	mm^2/s
Espessura Cortical	Mm
Transferência de Magnetização	%
Relaxometria	Ms
Volumetria	%

Fonte: Elaborada pelo autor.

Os dados que estavam formatados para geral foram formatados para numéricos. Os valores de espessura cortical foram formatados de texto para numérico. Foi determinado por convenção que os dados terão oito casas decimais após a vírgula com exceção da técnica espessura cortical que possui três casas após a vírgula. No caso especifico da volumetria os dados foram normalizados pelo volume de parênquima (SB+SC) de cada sujeito para reduzir o efeito do tamanho da cabeça, e assim do sexo. Desta forma os atributos envolvendo esta técnica a unidade apresentada nesta dissertação é porcentagem.

2.3 Critérios de Inclusão e Exclusão de Dados

Neste trabalho foram padronizados alguns critérios de inclusão e exclusão de dados para definição da população a ser estudada. As tabelas das bases de dados do CCIFM possuem muitos atributos que são as regiões segmentadas pelo FreeSurfer. Existem cerca de 170 atributos para cada técnica quantitativa com valores que representam cada região do cérebro. Ao longo desses anos de processamento de imagens foram feitos modificações que acarretaram problemas em alguns dados. Os valores desses atributos precisaram ser observados quanto a sua integridade, pois alguns apresentaram divergências. Esses erros podem ter origem de problemas no processamento de imagens que podem ser manuais ou devido a erros nos programas. Dentre os problemas foram encontrados resultados em branco, zerados, duplicados, entre outros, que inviabilizam sua utilização na classificação de dados. Foi feito uma organização de dados para garantir a qualidade e integridade da informação a ser classificada. Também foi realizada uma checagem de resultados para garantir a maior integridade e qualidade desses valores. No item 2.3.1 são apresentados alguns critérios de inclusão e exclusão de dados que foram feitos antes da classificação.

2.3.1 Atributos e Registros

Foram observados alguns valores com problemas que precisaram ser excluídos da classificação e realizado um controle de qualidade nos dados. Foram excluídos tanto atributos quanto registros que diminuíram a quantidade de instâncias a serem estudadas, mas foram necessários para garantir uma melhor qualidade da classificação dos dados. Optou-se por excluir alguns atributos por apresentarem alguns valores zerados ou em branco do que excluir o registro. Esses valores zerados ou em branco foram possivelmente causados por erros na segmentação ou erros no processamento. Isto foi feito para garantir o maior número de instâncias possíveis para a classificação.

Foram excluídos os seguintes atributos por apresentarem alguns dados em branco ou zerados, representativos de diferentes áreas do cérebro: *lh-insula, rh-insula, wm_lh_insula, wm_rh_insula, lh-transversetemporal, rh-transversetemporal, wm_lh_transversetemporal, wm_rh_transversetemporal*. O termo "lh" do inglês (*left hemisphere*) significa hemisfério esquerdo e "rh" do inglês (*right hemisphere*) significa hemisfério direito.

Também foram excluídos alguns registros que apresentaram algum tipo de problema. No quadro 2.3 são mostrados os motivos da exclusão de registros bem como a técnica quantitativa envolvida e suas respectivas quantidades:

Quadro 2.3 – Motivos de eliminação de registros por técnica quantitativa.

Motivos da Eliminação de registros e suas quantidades por Técnica Quantitativa									
Motivo	AF	DM	DPa	DPe	EC	MT	Rel	Vol	Tot
Registro Duplicado	12	12	14	17	12	14	04	13	98
Registro Triplicado					02	02		02	06
Registro Inconsistente		05					01	03	09
Reg. com atrib. zerado ou em branco	03	03	02	01			01	10	20
Registro com instância desigual	04	04	01	03				05	17
Total									**150**

Fonte: Elaborada pelo autor.

Os registros duplicados são aqueles onde o mesmo exame do participante apareceu mais uma vez na base de dados. Precisaram ser excluídos para garantir que dados repetidos ou próximos não fossem considerados na classificação. Em alguns casos onde o mesmo paciente fez exame em outro dia o registro foi considerado por se tratar de uma situação diferente. Os registros triplicados são aqueles onde o mesmo exame do participante apareceu mais duas vezes na base de dados. Esses outros dois registros também foram excluídos pelo mesmo motivo anterior.

Os registros inconsistentes são aqueles onde os valores encontrados estão fora do padrão aceitável. Valores muito altos ou muito baixos ou mesmo um conjunto de valores muito divergente dos demais foram considerados inaceitáveis para a classificação e descartados. Foi feita uma checagem de valores inconsistentes que será descrita mais adiante.

Foram excluídos também alguns registros que apresentaram alguns atributos zerados ou em branco. Foi feita esta opção ao invés de excluir os atributos para que fossem preservados em outras instâncias.

Na sequência são mostrados os atributos que se apresentaram zerados ou em branco em sua respectiva técnica quantitativa:

- AF: *WM-hypointensities, lh_paracentral* e *Left-Accumbens-area.*
- DM: *Optic-Chiasm, Left-Accumbens-area* e *wm_rh_transversetemporal.*
- DPer: *Optic-Chiasm.*
- Relax: *rh-entorhinal.*
- Vol: *wm_rh_transversetemporal, rh-entorhinal e Optic-Chiasm, lh-superiorparietal.*

Os registros com instância desigual foram aqueles que apresentaram valores que existiam em uma técnica e não existiam em outra. Foram excluídos para equiparar a quantidade de registros e garantir que existissem em todas as técnicas da doença. Foi excluído um total de 150 registros de participantes com alguma divergência nos dados.

2.3.2 Checagem de Valores

Após a exclusão de atributos e registros, os dados passaram por uma checagem para garantir sua integridade e poderem ser utilizados na classificação, sendo considerado como um filtro prévio. Inicialmente foi calculada a média e o desvio padrão de cada valor de registro por atributo. Depois foi realizada uma condição para cada valor do atributo. Se o valor do atributo fosse menor ou igual a media menos duas vezes o desvio padrão era colocado o valor “um”. Senão se o valor do atributo fosse maior ou igual à média mais duas vezes o desvio padrão era colocado o valor “zero”.

Essa condição deu origem a uma matriz de zero e uns nas mesmas posições dos valores. Depois foi feita uma média das linhas dos sujeitos e das colunas dos parâmetros para verificar se os dados estavam dentro do limite aceitável para a classificação.

Para a linha dos sujeitos foi estabelecida a condição de que se caso o valor da média fosse maior que 20% deveria-se revisar os dados da coluna. Senão se o valor da média fosse

menor ou igual a 20% os dados da coluna estariam corretos. Para cada coluna os valores não passaram de 20%, sendo considerados todos os dados dentro do padrão e aceitáveis para a classificação.

Para a coluna de sujeitos foi feita a condição de que se caso o valor da média fosse maior que 30% deveria-se revisar os dados da linha. Devido a esta condição muitos registros ficaram de fora, então resolvemos aumentar a porcentagem de média de erros para incluir mais dados. Então os registros que apresentaram média até 30% de erros foram considerados dentro da média e aceitos para a classificação. Os registros que ficaram entre 30% a 50% foram aceitos com ressalvas, ou seja, onde aparecesse "um" na coluna de checagem ele deveria ser substituído por "em branco" para diminuir a porcentagem de erros. Agora, os registros que apresentaram porcentagem acima de 50% de erros foram eliminados da classificação. O quadro 2.4 mostra a quantidade de registros considerados e eliminados por técnica quantitativa.

Quadro 2.4 – Motivos de eliminação de registros pela checagem por técnica quantitativa.

Motivo	AF	DM	DPa	DPe	EC	MT	Rel	Vol	Tot
Regs. considerados com ressalvas	03	02	02	06	08	05	13	03	42
Regs. eliminados da classificação	01	02	06	02	07	05	06	01	30

Fonte: Elaborada pelo autor.

De acordo com o quadro 2.4 foram considerados 42 registros com ressalvas e excluídos 30 registros com grande porcentagem de erros.

2.3.3 Redução das Áreas de Interesse

Ficou decidido que regiões do cérebro muito pequenas deveriam ser revisadas pelos possíveis erros da segmentação automática. Esses dados passaram por um procedimento para serem considerados na classificação. Os dados da técnica de volumetria de controles normais e pacientes EM foram a base para realizar o procedimento. Inicialmente foi feita a média de todos os valores de atributos. Os atributos que apresentaram valores de média abaixo de 0,2% (referido ao volume de parênquima) foram eliminados, com exceção de dados de atributos que tiveram a soma das médias dos hemisférios direito e esquerdo que ultrapassaram 0,2% e então foram considerados. Nesse caso foi feita uma concatenação dos hemisférios de cada registro para a técnica de volumetria. O quadro 2.5 mostra os atributos eliminados por

hemisfério com exceção do atributo *Optic-Chiasm* que não é dividida na segmentação nos dois hemisférios.

Quadro 2.5 – Atributos eliminados com volumes médios inferiores a 0,2%.

Hemisfério Esquerdo	Hemisfério Direito
Left-Accumbens-area	*Right-Accumbens-area*
lh-frontalpole	*rh-frontalpole*
wm_lh_entorhinal	*wm_rh_entorhinal*
wm_lh_parsorbitalis	*wm_rh_parsorbitalis*
wm_lh_frontalpole	*wm_rh_frontalpole*
wm_lh_temporalpole	*wm_rh_temporalpole*
Optic-Chiasm	

Fonte: Elaborada pelo autor.

Os atributos que formam o corpo caloso foram somados também num único atributo denominado "*wm_corpuscallosum*". O quadro 2.6 mostra os nomes dos atributos antes e depois da junção.

Quadro 2.6 – Atributos unidos para alcançar volumes médios superiores a 0,2%.

Atributos Anteriores		Atributo Posterior
lh-caudalanteriorcingulate	*rh-caudalanteriorcingulate*	*caudalanteriorcingulate*
lh-entorhinal	*rh-entorhinal*	*Entorhinal*
lh-isthmuscingulate	*rh-isthmuscingulate*	*isthmuscingulate*
lh-parahippocampal	*rh-parahippocampal*	*parahippocampal*
lh-parsorbitalis	*rh-parsorbitalis*	*Parsorbitalis*
lh-pericalcarine	*rh-pericalcarine*	*Pericalcarine*
lh-rostralanteriorcingulate	*rh-rostralanteriorcingulate*	*rostralanteriorcingulate*
lh-temporalpole	*rh-temporalpole*	*Temporalpole*
wm_lh_caudalanteriorcingulate	*wm_rh_caudalanteriorcingulate*	*wm_caudalanteriorcingulate*
wm_lh_corpuscallosum	*wm_rh_corpuscallosum*	*wm_corpuscallosum*
wm_lh_cuneus	*wm_rh_cuneus*	*wm_cuneus*
wm_lh_parahippocampal	*wm_rh_parahippocampal*	*wm_parahippocampal*
wm_lh_rostralanteriorcingulate	*wm_rh_rostralanteriorcingulate*	*wm_rostralanteriorcingulate*
Left-Pallidum	*Right-Pallidum*	*Pallidum*
Left-Amygdala	*Right-Amygdala*	*Amygdala*
CC_Posterior, Mid_Posterior, Anterior, Central e Mid_Anterior		*wm_corpuscallosum*

Fonte: Elaborada pelo autor.

Depois de tomar como base a técnica de volumetria para a junção de regiões, estes atributos também foram unidos nas outras técnicas, porém foi realizada a média entre os hemisférios e entre os atributos que formam o corpo caloso.

Também ficou definido que seria feita uma classificação restrita usando regiões reconhecidamente afetadas pela EM segundo a literatura (SANTOS, 2007). O intuito desta nova classificação foi de comparar com os resultados da classificação principal e verificar se seria mantido o *ranking* de regiões como melhores atributos. Um neuroradiologista com vários anos de experiência no estudo da doença foi responsável pela definição das seguintes áreas utilizadas nesta classificação restrita:

- Lobos Frontal, Parietal, Occipital e Temporal da SC e SB;
- Tronco cerebral;
- Cerebelo (SC e SB);
- Corpo Caloso;
- *Label* de Hipointensidade.

Foram utilizados os *labels* que possuíam no seu nome todas as regiões anteriormente mencionadas, como por exemplo, relacionados com a região occiptal: *lh-lateraloccipital*, *rh-lateraloccipital*, *wm_lh_lateraloccipital* e *wm_rh_lateraloccipital*.

2.4 Algoritmos de Classificação

O programa utilizado nas tarefas de classificação foi o *framework* WEKA (WEKA 3: Data Mining Software in Java.) (HALL, 2015), que é um programa desenvolvido pela Universidade de Waikato para mineração de dados através de algoritmos de aprendizado de máquina. Os algoritmos podem ser aplicados diretamente a um conjunto de dados ou chamados a partir do seu próprio código Java. Contém também ferramentas para pré-processamento de dados, classificação, regressão, clusterização, regras de associação e visualização. É um programa adequado para a mineração de dados e de seleção para agrupamentos dos melhores atributos dos dados quantitativos resultantes dos processamentos de imagem. O WEKA é um programa que implementa diversos indutores além de ferramentas para validação e análise de classificadores e suas bibliotecas são disponibilizadas livremente para desenvolvedores (HALL *et al*., 2009).

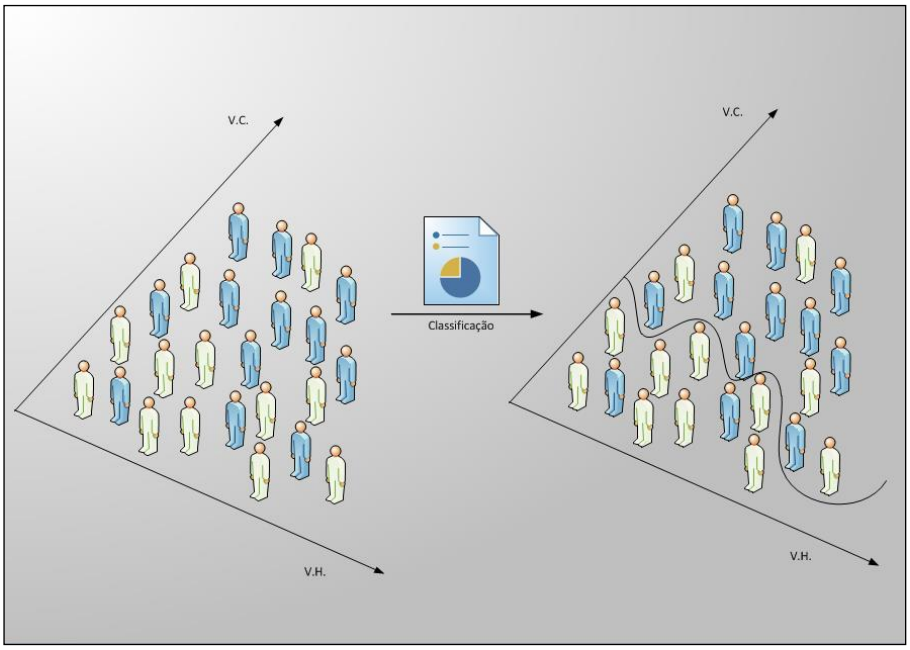

Figura 2.2 – Representação de dois conjuntos de classes (saudáveis em azul e pacientes em branco) sobre os eixos de duas características (V.H.) e (V.C.). Após a classificação, à direita, observa-se a definição de uma fronteira que define a melhor separação entre as classes do algoritmo usado.
Fonte: Elaborada pelo autor.

A classificação de dados foi entre indivíduos saudáveis e pacientes com esclerose múltipla. A figura 2.2 mostra uma representação da separação entre as classes feitas por uma classificação. Utilizaremos o classificador KNN e SVM implementados no programa WEKA versão 3.6.11 variando o fator próximo de 1, 3, 5 e opção de teste de validação cruzada (VC) com grupos de dez para KNN. O KNN foi utilizado por sua simplicidade e o SVM por sua grande utilização nos últimos anos em problemas similares. Foram realizadas três classificações considerando diferentes atributos:

1 – Com todos os atributos sem nenhum tipo de critério de seleção de atributos;

2 – Com seleção de melhores atributos sem grupo de teste usando todas as instâncias (TI);

3 – Com seleção de melhores atributos com grupo de teste usando validação cruzada (VC).

Para as 2ª e 3ª classificações serão utilizados dois métodos de seleção de atributos com o avaliador *WrapperSubsetEval* e o método de pesquisa de melhores resultados. O método *WrapperSubsetEval* avalia subconjuntos de variáveis, permitindo detectar possíveis variações entre variáveis. No avaliador *WrapperSubsetEval* para a seleção de atributos TI foram utilizados cinco grupos e para a seleção VC foram utilizados dez grupos.

O classificador SVM não foi utilizado como critério na seleção de atributos devido a sua grande demora na execução do procedimento. No final foram gerados arquivos texto de seleção de atributos e de classificações que foram analisados e os resultados colocados neste trabalho. A principal forma de análise dos resultados foi através das áreas sob a curva ROC elaborada com os valores de sensibilidade e especificidade originados da matriz de confusão. O classificador SVM foi utilizado na classificação dos melhores atributos realizados por KNN nos fatores 1, 3 e 5. Isto foi feito para comparar qual a melhor estratégia de classificação.

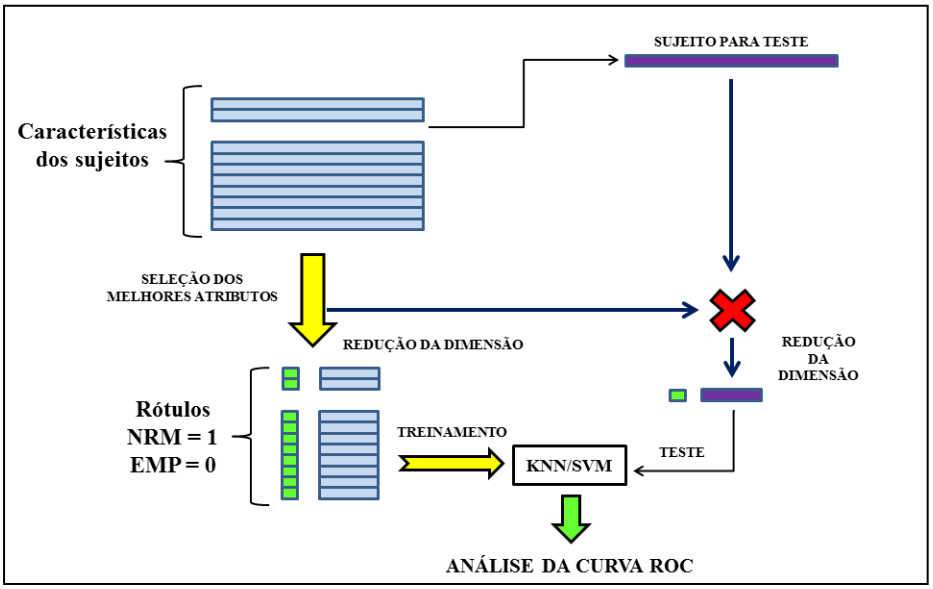

Figura 2.3 – Fluxograma do modelo de classificação baseado em KNN e SVM, incluindo os procedimentos treinamento e testes que usam a seleção dos melhores atributos.
Fonte: Elaborada pelo autor.

Os valores foram classificados de acordo com os parâmetros de cada algoritmo, ou seja, KNN e SVM, considerando para NRM = positivo (1) e para EMP = negativo (0). O propósito do SVM foi encontrar a margem de hiperplano máxima que representa a maior separação ou margem entre as duas classes (LI *et al.*,2014).

A figura 2.3 mostra o fluxograma de classificação para KNN e SVM utilizado neste trabalho. Conforme são feitas as seleções dos melhores atributos, a dimensão dos vetores diminui isto porque o vetor foi reduzido para somente com os melhores atributos. Os algoritmos realizaram o treinamento dos dados e fizeram testes com 10 registros do mesmo conjunto para validação da classificação. Foram feitas três classificações: com todos os atributos, com seleção de melhores atributos sem grupo de teste e com seleção de melhores atributos com grupo de teste. Foram feitos os testes e treinamentos utilizando os algoritmos KNN e SVM e por fim foram feitas as análises da curva ROC.

2.5 Classificações de Dados

A classificação de dados baseou-se a partir das sequências de RM realizadas previamente para a pesquisa de EM. Depois o atlas do FreeSurfer foi sobreposto às imagens e na sequência foi feita a segmentação das regiões de interesse. A partir dessa segmentação foi possível a geração de resultados que foi o objeto de interesse deste estudo.

A classificação de dados foi do tipo "Supervisionada", ou seja, já conhecemos previamente o padrão e o conjunto de classes que serão entre CN e pacientes EM. Para a classificação de dados foram utilizados duas classificações de vetores principais que são denominados: global e técnicas quantitativas. No vetor global foram colocados todos os atributos de todas as técnicas quantitativas em sequência e como vetores técnicas quantitativas ou simplesmente técnicas foram considerados oito vetores com a respectiva técnica em cada um deles e seus respectivos atributos. A seguir tem-se uma representação do vetor global:

$$GLOBAL[664] = \{idade, ANF[58], DME[58], DPA[58],$$

$$DPE[58], EC[52], MTR[126], RELAX[126], VOL[126], classe\}$$

onde ANF representa Anisotropia Fracionada, DME – Difusibilidade Média, DPA – Difusibilidade Paralela, DPE – Difusibilidade Perpendicular, ECT – Espessura Cortical, MTR – Transferência de Magnetização, REL – Relaxometria e VOL – Volumetria. E os valores entre "[ ]" representam o número total de atributos que são regiões, idade e classe avaliados em cada técnica quantitativa. Um exemplo de conjunto de atributos é mostrado a seguir:

$$ANF[60] = \{idade, left_{cerebral_{wm}}, left_{cerebral_{cortex}},$$

$$left_{cerebellum_{wm}}, left_{cerebellum_{cortex}}, \dots, classe\}$$

Foram feitos mais sete vetores para as técnicas quantitativas restantes, ou seja, DME[60], DPA[60], DPE[60], ECT[54], MTR[128], REL[128] e VOL[128]. Para cada vetor foram acrescentados os atributos de idade e classe totalizando mais dois atributos para cada vetor.

O vetor global representou as classes que foram estudadas sendo assim GLO representa os CN e pacientes com EM. O vetor global teve 664 posições que são todos os atributos de todas as técnicas acrescentados da idade e classe. Para os vetores técnica, os vetores de difusão que são ANF, DME, DPA e DPE tiveram 60 atributos cada um, que representam as regiões de SB. O vetor ECT foi o menor vetor de todos, pois contou com 54 atributos que representam apenas valores da região do córtex. Os demais vetores técnica MTR, REL e VOL contaram com 128 atributos que representaram as regiões de SB e SC do cérebro. No quadro 2.7 são mostrados todos os vetores utilizados nas classificações deste trabalho. No total foram nove vetores de classificação e esses vetores de atributos foram cruzados com os valores de registros de controles normais e pacientes EM, formando nove matrizes de características.

Os classificadores que foram utilizados são o KNN e o SVM usando como critério de seleção de atributos todo o conjunto de instâncias e a validação cruzada que dividiu em grupos. Lembrando que seleção de melhores atributos foi feita apenas pelo classificador KNN, devido a grande demora na seleção do SVM. Depois que o KNN trouxe os melhores atributos, foi utilizado o classificador SVM a partir dessa seleção do KNN, para a comparação entre os classificadores. O critério para a seleção dos melhores atributos foram os mais relevantes para o conjunto de instâncias e os que obtiveram participação maior ou igual a trinta por cento na classificação por validação cruzada.

O quadro 2.7 mostra a quantidade de instâncias para os nove vetores da classificação. Foi um vetor global contendo todas as técnicas e nove vetores técnica contendo cada um a técnica quantitativa específica.

Quadro 2.7 – Quantidade de instâncias por vetor.

TÉCNICA	Controles	Pacientes	MATRIZ
GLO	53	64	117 x 664
ANF	75	90	165 x 60
DME	75	90	165 x 60
DPA	76	90	166 x 60
DPE	76	90	166 x 60
ECT	167	144	311 x 54
MTR	91	109	200 x 128
REL	90	108	198 x 128
VOL	165	138	303 x 128

Fonte: Elaborada pelo autor.

A classificação foi feita no programa WEKA versão 3.6.11 utilizando os KNN de um três e cinco com a opção de teste de validação cruzada com dez grupos. Esse teste divide a quantidade total de instâncias em grupos de dez, realiza o treinamento em nove grupos e deixa um grupo para realização dos testes de classificação. Para a seleção de melhores características o avaliador a ser utilizado foi o *WrapperSubsetEval* com os classificadores KNN de um, três e cinco. O método de pesquisa foram os primeiros melhores resultados. O modo de seleção de características foi utilizando todo o conjunto de treinamento e por validação cruzada com dez grupos na seleção e dez grupos no avaliador de atributo.

Para cada classificação e seleção de atributos foi gerado um arquivo texto contendo todas as informações dos resultados. Dentre essas informações foram mostrados a matriz de confusão e a área sob a curva ROC. Foram realizados três tipos de classificação. A primeira com todos os atributos (TA), a segunda após a seleção das melhores características com todas as instâncias (TI) e a terceira após a seleção das melhores características por validação cruzada (VC). No final desta classificação espera-se encontrar as melhores características por vetor e discutir sobre as técnicas de classificação e seleção das melhores atributos das técnicas quantitativas e outras análises relevantes que poderão ocorrer.

Os apêndices A, B e C mostram respectivamente os parâmetros de classificação utilizados no programa WEKA sendo respectivamente a classificação de atributos KNN, a classificação de atributos SVM e seleção de atributos KNN.

O número máximo de indicações de um atributo como melhor atributo poderia ser 48 que corresponde com 3 algoritmos x 2 estratégias de seleção de atributos x 8 técnicas quantitativas. Aqui consideramos melhores atributos no máximo os 5 primeiros em cada situação.

2.6 Tarefas de Classificação

Para as tarefas de classificação foi escolhida a doença Esclerose Múltipla que afeta o SNC, pois as regiões afetadas são previamente conhecidas de acordo com a literatura. Adicionalmente aos grupos de pacientes contamos com os respetivos grupos controle formado por indivíduos saudáveis, segundo anteriormente descritos.

No geral, este trabalho teve uma tarefa principal de classificação que foi controles *versus* pacientes de esclerose múltipla. Foram utilizados 8 vetores que representaram as técnicas quantitativas e 1 vetor global contendo todas as técnicas, totalizando 9 vetores de

classificação. Os classificadores KNN e SVM foram utilizados, sendo que no caso do KNN foram utilizadas variantes de 1, 3 e 5 vizinhos. Os atributos considerados foram todos os atributos, melhores atributos sem grupo de teste e melhores atributos com grupo de teste, sendo que a seleção de melhores atributos somente foi feita por KNN. A figura 2.4 representa de forma resumida as tarefas de classificação que foram executadas.

A classificação foi separada por técnicas para verificar o comportamento dos dados sob diferentes aspectos, isto porque o princípio físico envolvido em cada técnica é diferente. As regiões mais envolvidas na classificação são diferentes entre as técnicas e a suas interpretações são mostradas nos resultados.

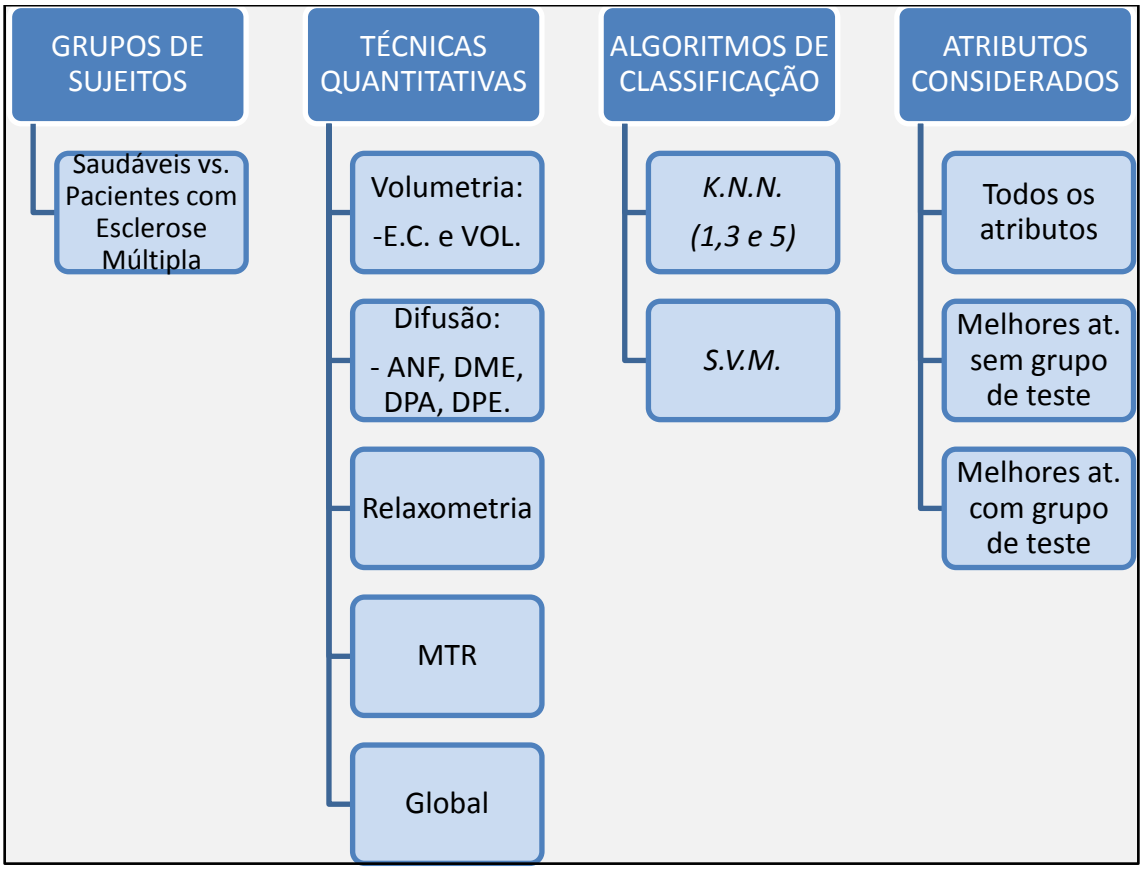

Figura 2.4 – Esquema representativo das tarefas de classificação a serem executadas indicando os dados e os algoritmos considerados em cada classificação.
Fonte: Elaborada pelo autor.

A classificação no caso dos dados com as regiões restritas foi realizada de forma semelhante à figura 2.4 com os mesmos dados de controles normais e pacientes EM e usando somente melhor classificador que foi encontrado na classificação principal: KNN5 (ver resultados, figura 3.11).

3. RESULTADOS

Após a realização da seleção das classificações foram gerados 279 arquivos texto contendo todas as informações sobre os resultados das classificações. As melhores características dos vetores foram selecionadas e colocadas em um *ranking* por vetor. Primeiramente foram observadas as melhores características por classificador e depois colocadas em um *ranking* por ordem de quantidade que a característica apareceu como critério de seleção usando todo o conjunto de instâncias e com quantidade maior ou igual a trinta por cento de aparecimento por validação cruzada. Foram consideradas as melhores características por vetor apresentadas sendo necessário esse critério de seleção devido a grande quantidade de características encontradas. O restante das características encontra-se no apêndice ao final deste trabalho. Para uma melhor compreensão dos resultados eles foram separados nas duas principais etapas do processamento dos dados: a obtenção dos melhores atributos e os resultados das classificações.

3.1 Melhores Atributos

Antes da tarefa ou processo de classificação, foi feita uma seleção dos melhores atributos a serem explorados. Isto foi feito para melhorar o desempenho de acertos da classificação bem como para descobrir quais são as melhores características que influenciam o processo de classificação. Esperamos desta forma sugerir aqueles atributos que diferenciam mais claramente entre indivíduos doentes e sadios. O que definem quais são os melhores atributos é a quantidade de valores diferentes entre as classes que são comparadas do mesmo atributo, ou seja, quanto maior for a diferença entre os valores do mesmo atributo, maior será as chances desses de serem considerados como melhores atributos.

Para descobrir os melhores atributos foi feito um *ranking* com os atributos que mais apareceram durante o processo de seleção. Foram divididas em planilhas por vetor a ser utilizado, ou seja, vetor global e vetores técnicas.

Para o vetor global todas as técnicas quantitativas e seus atributos foram colocados em um único vetor. Inicialmente as características foram ranqueadas por tipo de seleção de atributos, ou seja, por todas as instâncias e por validação cruzada. Nos dois casos todas as características que apareceram em todas as seleções de atributos foram colocadas em uma tabela divididas por número de KNN, ou seja, 1, 3 ou 5. Em seguida as características foram

colocadas em um *ranking* por ordem de quantidade em que apareceram na seleção de atributos, ainda divididos por tipo de seleção de atributos. Depois os dois *rankings* foram unidos em um *ranking* total colocando a característica e a quantidade de vezes que elas apareceram por tipo de seleção de atributos. Finalmente obteve-se uma tabela com as melhores características de todas as seleções de melhores atributos, ordenadas por quantidade de vezes que apareceram. As características que mais apareceram foram separadas para a análise de dados.

Para os vetores técnicas foram seguidas as mesmas regras estabelecidas para o vetor global. Obteve-se um número maior de características por se tratar de uma quantidade de vetores com menor quantidade de atributos. Optou-se por colocar o máximo de cinco melhores características por técnica, devido ao grande número de empates.

As análises das melhores características foram separadas pelos oito vetores técnica e pelo vetor global, totalizando nove análises das melhores características. Elas foram colocadas em uma tabela contendo nas linhas a média e o desvio padrão dos dados de saudáveis e pacientes. Nas colunas foram colocadas as melhores características por vetor. Esses dados foram colocados em nove gráficos de barras onde se pode observar a média e desvio padrão por saudável e controle. Cada gráfico por técnica teve a sua respectiva unidade por se tratar de uma medida única. Apenas o gráfico global precisou de adaptações, pois as unidades das características eram diferentes. Devemos salientar que esta forma de apresentação, através dos valores médios, não reflete diretamente a metodologia de classificação usada neste trabalho que não é baseada numa comparação de médias das populações; porém permite ilustrar a distribuição dos valores dos atributos mais relevantes no processo de classificação.

Os melhores atributos dos nove vetores foram separados e os dados originais foram analisados. Para se ter uma ideia da variação destes melhores atributos entre os grupos, foi feito a média e o desvio padrão de cada atributo separado pelas classes saudáveis e pacientes EM. Na sequência são mostrados os gráficos de média e desvio padrão dos dados de cada vetor.

Os dois melhores atributos do vetor global foram *VOL_WM-hypointensities* e *MTR_wm_cospuscallosum*. A figura 3.1 mostra o gráfico de média e desvio padrão dos dados de saudáveis e pacientes EM desses atributos.

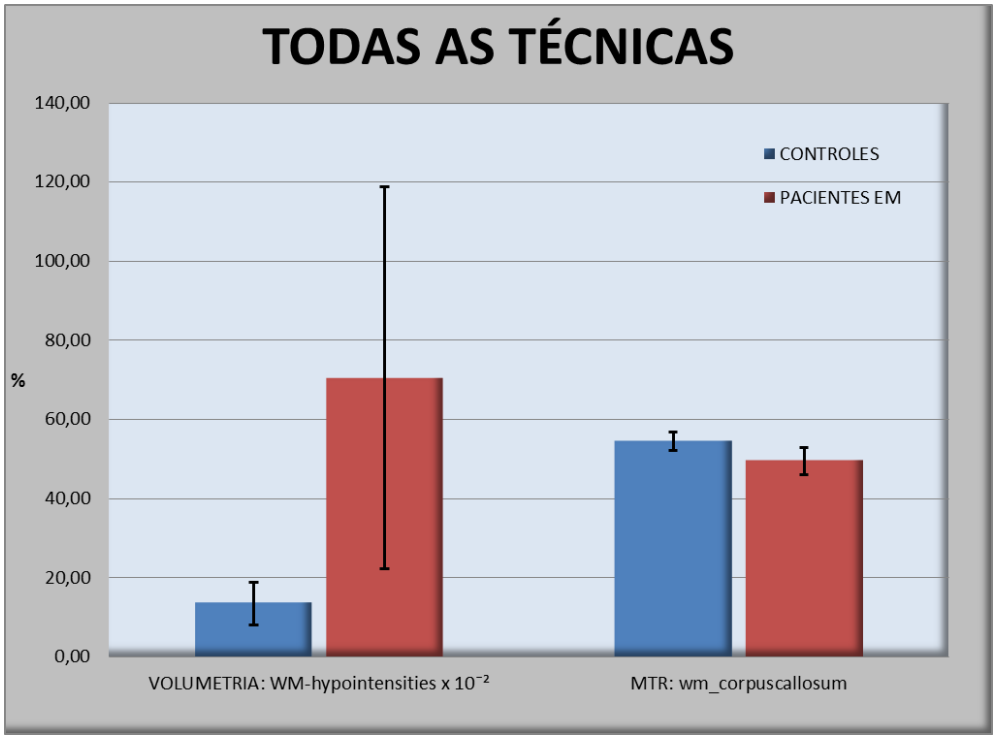

Figura 3.1 – Valores médios dos melhores atributos obtidos na tarefa de classificação usando o vetor global nos grupos de controles e pacientes. Os valores de desvio padrão são também indicados no gráfico. O eixo vertical deve ser visto com cautela, pois cada técnica indica valores com diferentes significados.

Fonte: Elaborada pelo autor.

De acordo com o gráfico a diferença das médias entre saudáveis e pacientes EM do atributo *VOL_WM-hypointensities* foi grande, com quase 0,60 % de diferença, sendo a média para saudáveis de 0,14 % e para pacientes EM de 0,7 %. O desvio padrão deste atributo para pacientes também foi distante da média, indicando muitos valores distantes da média. A variação entre saudáveis e pacientes EM foi menor no outro atributo *MTR_wm_cospuscallosum* sendo a média de saudáveis 54,6% e a média de pacientes EM 49,8%. O desvio padrão de saudáveis e pacientes EM deste último atributo foi perto da média e a variação foi pequena.

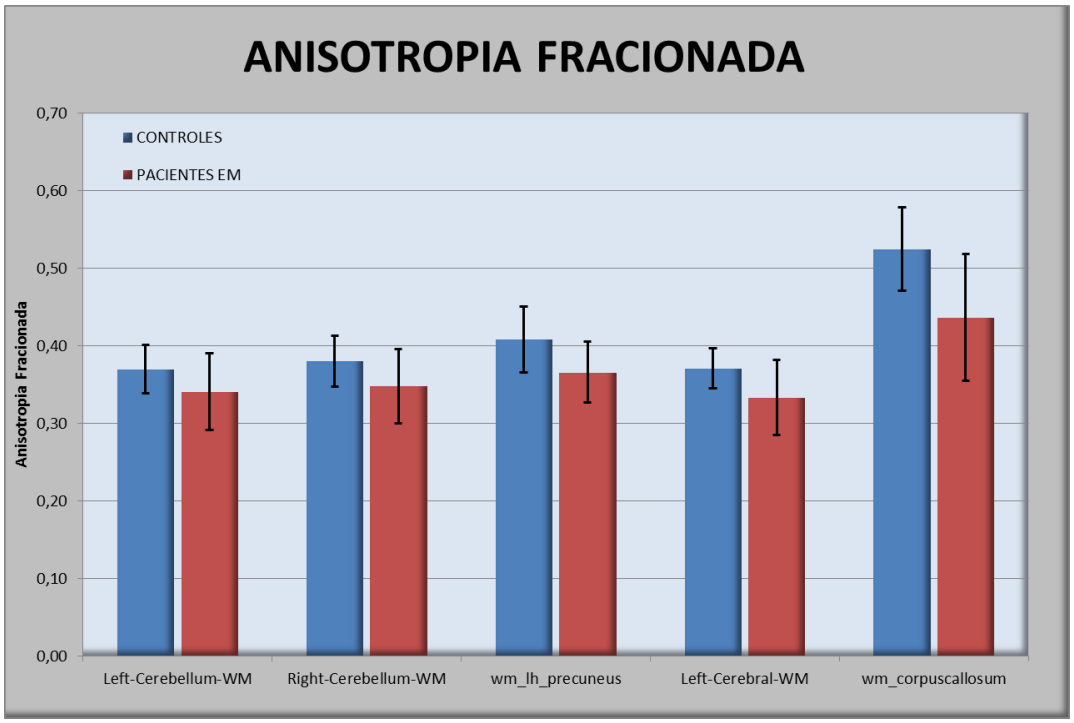

Figura 3.2 – Valores médios dos melhores atributos obtidos na tarefa de classificação usando o vetor técnica Anisotropia Fracionada nos grupos de controles e pacientes. Os valores de desvio padrão são também indicados no gráfico. O eixo vertical é adimensional.
Fonte: Elaborada pelo autor.

Para o vetor técnica anisotropia fracionada os atributos que mais apareceram na seleção de atributos foram *Left-Cerebellum-White-Matter*, *Right-Cerebellum-White-Matter*, *wm_lh_precuneus*, *Left-Cerebral-White-Matter* e *wm_corpuscallosum*. A figura 3.2 mostra o gráfico de média e desvio padrão de anisotropia fracionada. As unidades de valores de anisotropia fracionada não existem, ou seja, os valores são adimensionais. Em todos os casos as médias de saudáveis ficaram maiores que a média de pacientes EM. O atributo *wm_corpuscallosum* atingiu as maiores médias sendo 0,52 para saudáveis e 0,44 para pacientes EM. As médias de todos os outros atributos ficaram abaixo de 0,41. A variação máxima do desvio padrão entre estes melhores atributos foi de 0,06. Todos esses atributos representam regiões da SB, sendo 3 do hemisfério esquerdo, 1 do hemisfério direito e 1 dos dois hemisférios.

Os melhores atributos do vetor difusibilidade média foram *wm_lh_precuneus*, *wm_rh_isthmucingulate*, *Left-Cerebellum-White-Matter* e *wm_corpuscallosum*. A figura 3.3 mostra o gráfico de média e desvio padrão de difusibilidade média. O atributo *wm_lh_corpuscallosum* foi o que teve as maiores valores de média e desvio padrão para saudáveis e pacientes EM atingindo 1,25 x 10^{-3} mm²/s para saudáveis e 1,52 10^{-3} mm²/s

para pacientes EM. Para todos os melhores atributos de difusibilidade média os valores de média e desvio padrão dos pacientes EM foram maiores que os de saudáveis.

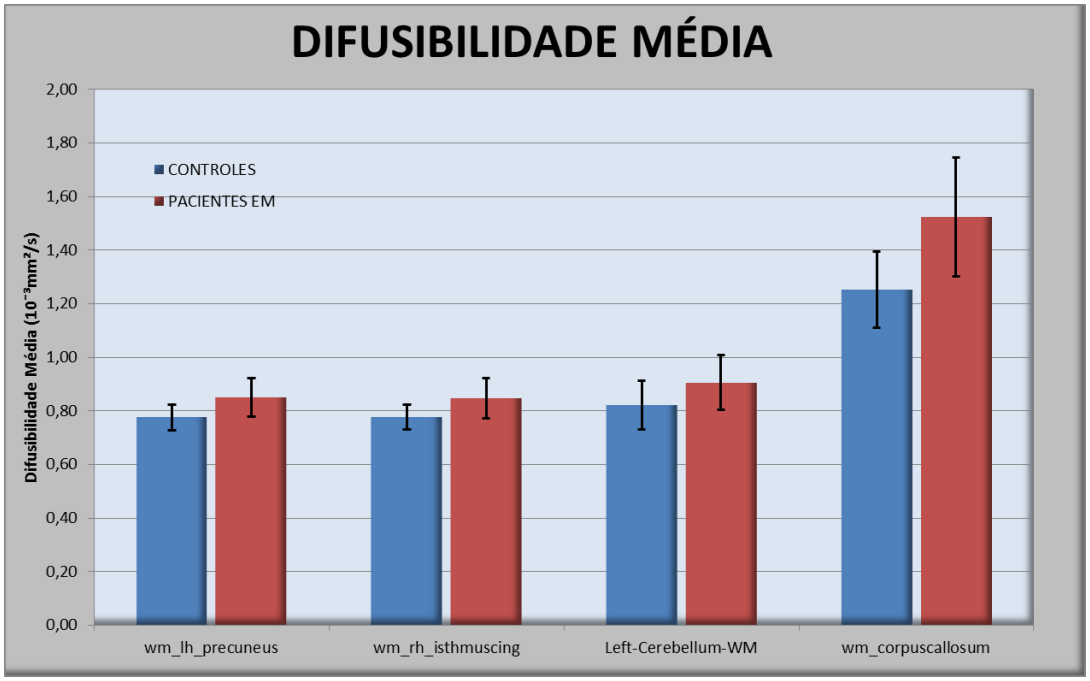

Figura 3.3 – Valores médios dos melhores atributos obtidos na tarefa de classificação usando o vetor técnica Difusibilidade Média nos grupos de controles e pacientes. Os valores de desvio padrão são também indicados no gráfico. A unidade do eixo vertical é 10^{-3}mm^2/s.
Fonte: Elaborada pelo autor.

Os valores de média dos outros atributos ficaram acima de 0,77 x 10^{-3} mm²/s e abaixo de 1 x 10^{-3} mm²/s. Todos os atributos foram de regiões da SB, sendo dois do hemisfério esquerdo, 1 do hemisfério direito e 1 dos dois hemisférios.

O vetor difusibilidade paralela trouxe como melhores atributos *wm_corpuscallosum, wm_lh_medialorbitofrontal, wm_rh_caldalmiddlefrontal* e *wm_rostralanteriorcingulate.* A figura 3.4 mostra o gráfico de média e desvio padrão para difusibilidade paralela. Em todos esses atributos os valores de média e desvio padrão de pacientes EM foram maiores que os de saudáveis. As maiores médias foram para o atributo *wm_corpuscallosum* chegando a 1,96 x 10^{-3} mm²/s para saudáveis e 2,19 x 10^{-3} mm²/s para pacientes EM. O atributo *wm_rostralanteriorcingulate* apresentou valores de média de 1,40 x 10^{-3} mm²/s para saudáveis e 1,66 x 10^{-3} mm²/s para pacientes EM. As médias dos outros dois atributos ficaram acima de 1,05 x 10^{-3} mm²/s e abaixo de 1,24 x 10^{-3} mm²/s. Entre esses atributos, 2 representam regiões dos dois hemisférios do cérebro, 1 representa o hemisfério esquerdo e 1 o hemisfério direito, sendo todos regiões da SB do cérebro.

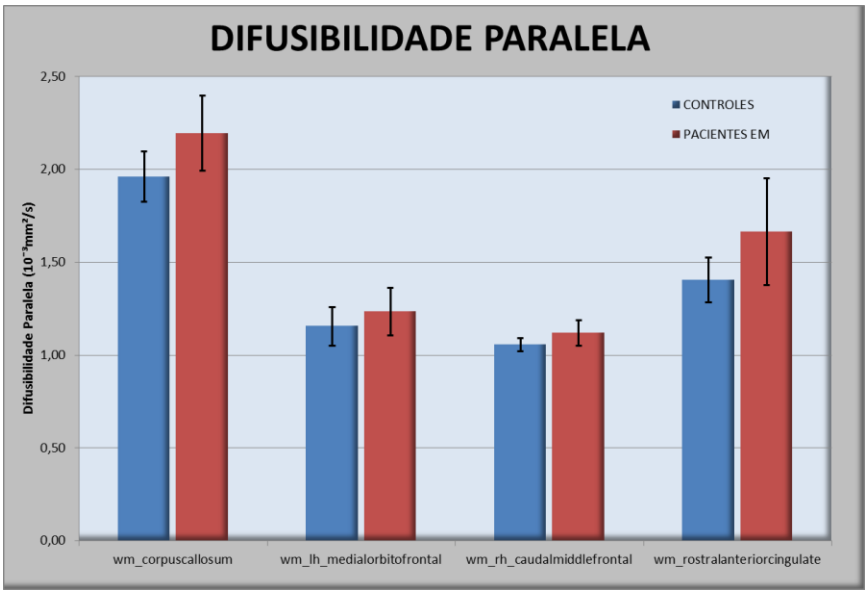

Figura 3.4 – Valores médios dos melhores atributos obtidos na tarefa de classificação usando o vetor técnica Difusibilidade Paralela nos grupos de controles e pacientes. Os valores de desvio padrão são também indicados no gráfico. A unidade do eixo vertical é $10^{-3}mm^2/s$.
Fonte: Elaborada pelo autor.

Os valores da classificação para difusibilidade perpendicular são mostrados no na figura 3.5. Em todos os atributos os valores de pacientes foram maiores que os de saudáveis. Todas as médias ficaram abaixo de 0,7 x 10^{-3} mm²/s, indicando que dentre as difusibilidades, esta foi a que apresentou os menores valores. Os melhores atributos para difusibilidade perpendicular foram *wm_lh_isthmuscingulate*, *wm_lh_inferiortemporal*, *wm_rh_inferiortemporal* e *wm_lh_fusiform*.

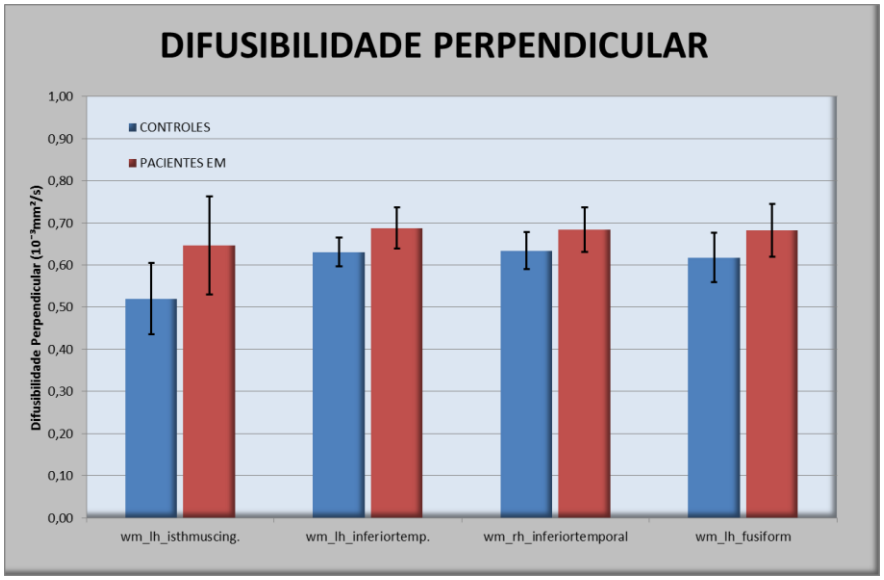

Figura 3.5 – Valores médios dos melhores atributos obtidos na tarefa de classificação usando o vetor técnica Difusibilidade Perpendicular nos grupos de controles e pacientes. Os valores de desvio padrão são também indicados no gráfico. A unidade do eixo vertical é $10^{-3}mm^2/s$. Fonte: Elaborada pelo autor.

Houve pouca variação entre as médias e desvio padrões dos valores de difusibilidade perpendicular. As médias ficaram acima de 0,52 x 10^{-3} mm²/s e abaixo de 0,69 x 10^{-3} mm²/s. Os desvios padrão ficaram acima de 0,03 x 10^{-3} mm²/s e abaixo de 0,12 x 10^{-3} mm²/s. As regiões que mais apareceram foram do hemisfério esquerdo com 3 e 1 do hemisfério direito.

O vetor de espessura cortical apresentou como melhores atributos *lh-fusiform*, *lh-lingual*, *pericalcarine* e *lh-precentral*. A figura 3.6 mostra o gráfico de média e desvio padrão do vetor espessura cortical. Observa-se que dentre estes 4 atributos, 3 são do hemisfério esquerdo e 1 dos dois hemisférios. Em todos os casos os valores de saudáveis foram maiores que os valores de pacientes EM, sendo os maiores valores encontrados no atributo *lh-fusiform* com 2,87 mm e 2,61 mm de média de saudáveis e pacientes EM respectivamente.

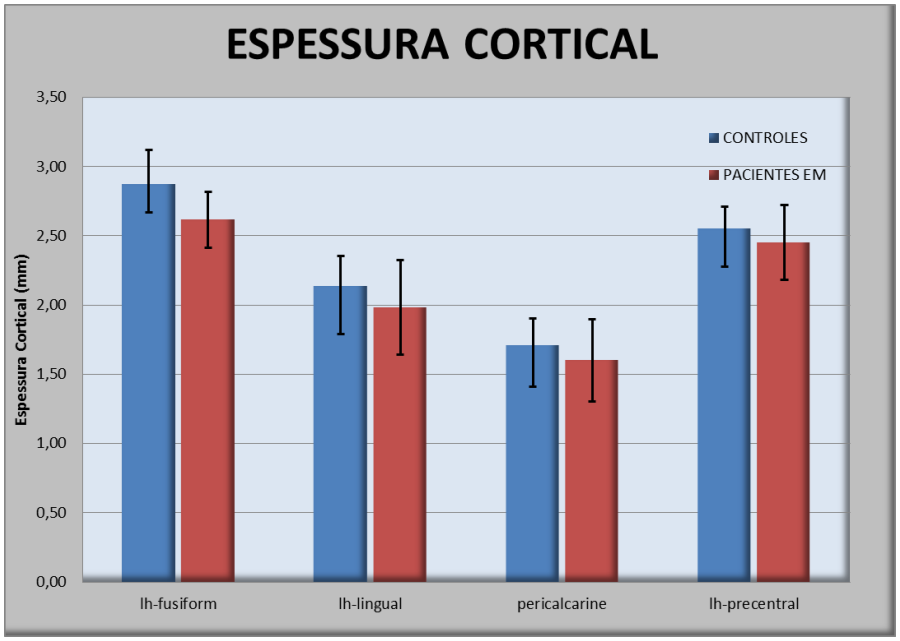

Figura 3.6 – Valores médios dos melhores atributos obtidos na tarefa de classificação usando o vetor técnica Espessura Cortical nos grupos de controles e pacientes. Os valores de desvio padrão são também indicados no gráfico. A unidade do eixo vertical é mm.
Fonte: Elaborada pelo autor.

O atributo *lh-precentral* também apresentou valores altos alcançando médias de 2,55 mm para saudáveis e 2,45 mm para pacientes EM. Os outros dois atributos atingiram médias acima de 1,60 mm e abaixo de 2,14 mm.

Os atributos que mais apareceram para o MTR foram *Left-Hippocampus*, *wm_lh_fusiform*, *wm_lh_caudalmiddlefrontal*, *Brain-Stem* e *wm_parahippocampal*. A figura 3.7 mostra o gráfico de média e desvio padrão de MTR. Os valores tiveram pouca variação

entre atributos e entre classes, ficando entre 47% a 60% os valores de média. A variação do desvio padrão também foi pequena em todos os casos, apresentando no máximo 1,7 % de variação entre o menor e o maior valor. Em todos os atributos os valores de saudáveis foram maiores que os de pacientes EM.

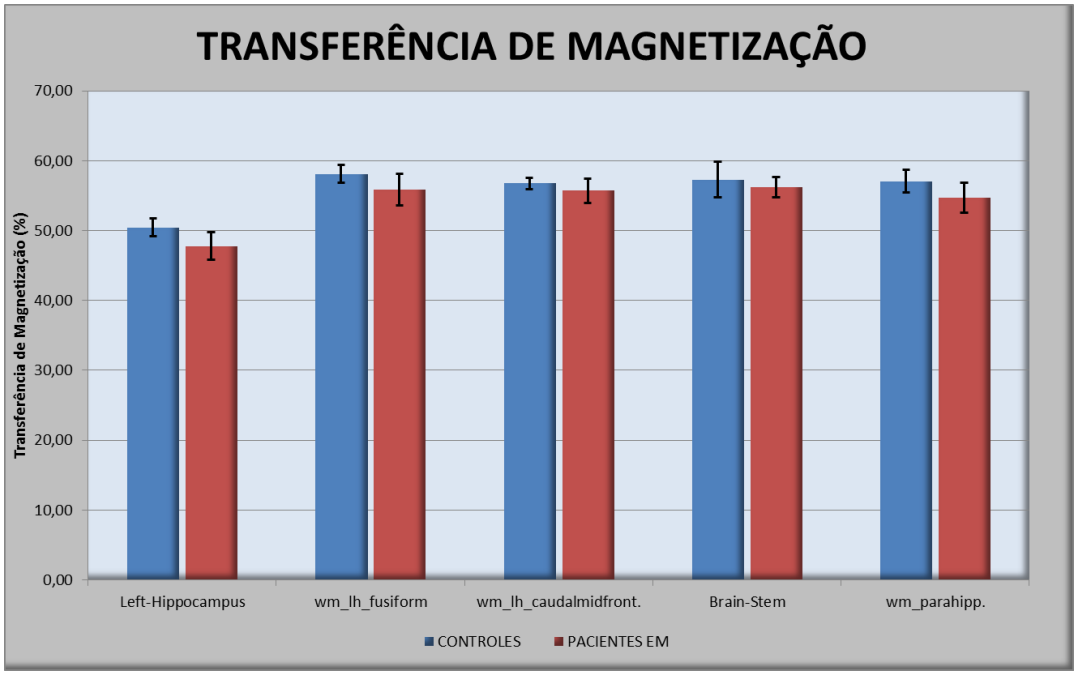

Figura 3.7 – Valores médios dos melhores atributos obtidos na tarefa de classificação usando o vetor técnica Transferência de Magnetização nos grupos de controles e pacientes. Os valores de desvio padrão são também indicados no gráfico. A unidade do eixo vertical é %.
Fonte: Elaborada pelo autor.

O atributo que apresentou os maiores valores de média foi o *wm_lh_fusiform* com 58,09% para saudáveis e 55,85% para pacientes EM. Os menores valores de média ficaram com atributo *Left-Hippocampus* com 50,45% para saudáveis e 47,78% para pacientes EM. Dos 5 atributos encontrados, 3 são do hemisfério esquerdo e 2 são dos dois hemisférios.

O vetor de Relaxometria apresentou os seguintes melhores atributos: *wm_lh_fusiform, Pallidum, wm_corpuscallosum, wm_lh_precuneus* e *wm_lh_posteriorcingulate*. A figura 3.8 mostra o gráfico de média e desvio padrão de Relaxometria. Os valores de pacientes EM foram maiores em relação a saudáveis, com exceção do atributo *Pallidum* onde os valores de saudáveis foram maiores. A maior média ficou com o atributo *wm_corpuscallosum* atingindo 145 ms para pacientes EM e 112 ms para saudáveis.

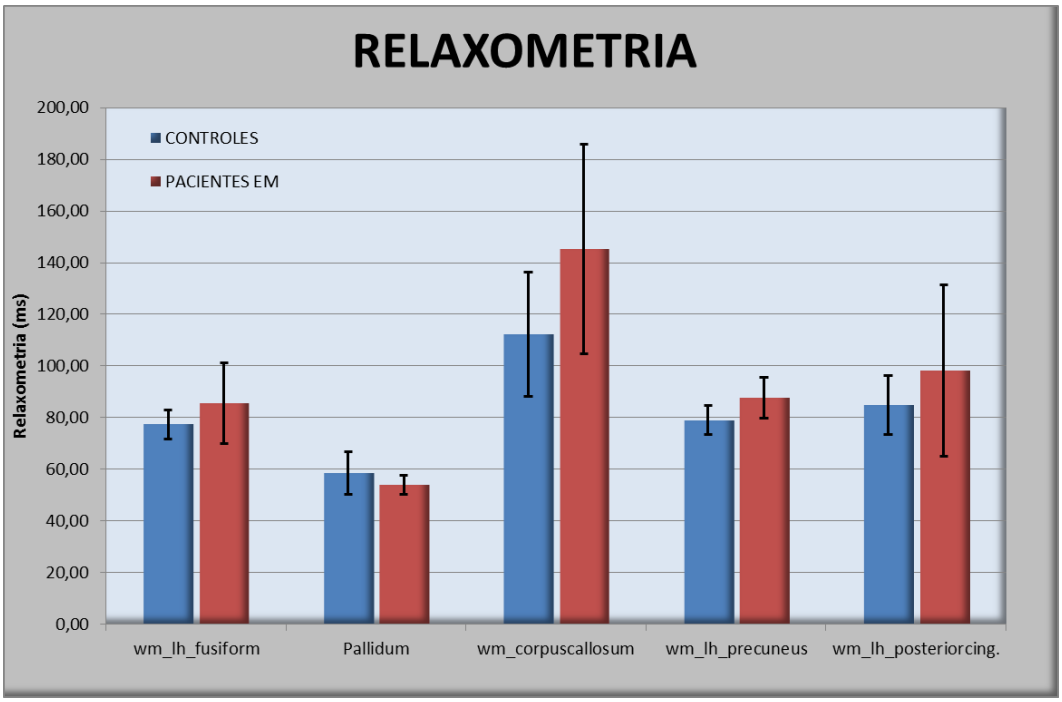

Figura 3.8 – Valores médios dos melhores atributos obtidos na tarefa de classificação usando o vetor técnica Relaxometria nos grupos de controles e pacientes. Os valores de desvio padrão são também indicados no gráfico. A unidade do eixo vertical é ms.

Fonte: Elaborada pelo autor.

Dentre os atributos encontrados, 4 são somente de SB e o outro de SC. Três atributos são do hemisfério esquerdo e 2 de ambos os hemisférios.

Os melhores atributos para volumetria são *WM-hypointensities, Right-VentralDC e lh-lateralocciptal*. A figura 3.9 mostra a média e o desvio padrão de volumetria. O atributo que apresentou a maior média foi o *lh-lateralocciptal* com 0,94 % para saudáveis e 0,73 % para pacientes EM. Dentre todas as técnicas, a volumetria foi a que apresentou a maior variação de valores entre atributos e entre classes. A variação entre a menor e maior média de pacientes chegou a 0,52 %. O atributo *WM-hypointensities* apresentou a maior variação entre média e desvio padrão de saudáveis e pacientes EM, sendo 0,52 % e 0,40 % a variação entre médias e desvios padrão. Para os atributos *Right-VentralDC* e *lh-lateraloccipital*, os valores de saudáveis foram maiores e para o atributo *WM-hypointensities* os valores de pacientes EM foram maiores. O atributo *WM-hypointensities* apresentou a maior variação entre médias e desvios padrão entre todos os melhores atributos desta pesquisa.

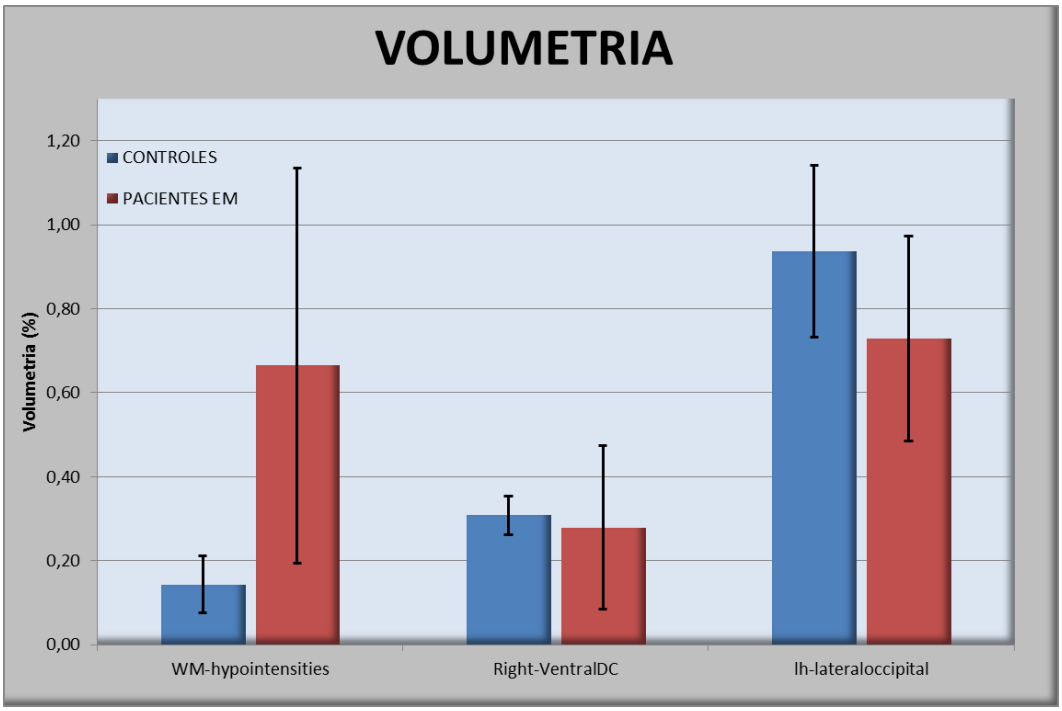

Figura 3.9 – Valores médios dos melhores atributos obtidos na tarefa de classificação usando o vetor técnica Volumetria nos grupos de controles e pacientes. Os valores de desvio padrão são também indicados no gráfico.
Fonte: Elaborada pelo autor.

Após a seleção dos melhores atributos por todas as instâncias e validação cruzada para os vetores, aqueles que mais apareceram entre todos os vetores técnica são mostrados na figura 3.10. O atributo *wm_lh_corpuscallosum* foi o que mais apareceu entre todos, totalizando 19 vezes. Em segundo ficou *wm_lh_precuneus* com 13. Na sequência tem os atributos *WM-hypointensities* e *wm_lh_fusiform* que apareceram 11 vezes e *Left-Cerebellum-White-Matter* que apareceu 10 vezes. Todos esses atributos em destaque são de regiões da SB do cérebro como era de se esperar, pois as lesões da EM acontecem na SB, fazendo com que esses valores sejam diferentes entre pacientes EM e indivíduos saudáveis. Todos esses atributos configuram como os melhores para classificar indivíduos saudáveis e pacientes EM, por apresentarem valores bem heterogêneos entre as classes.

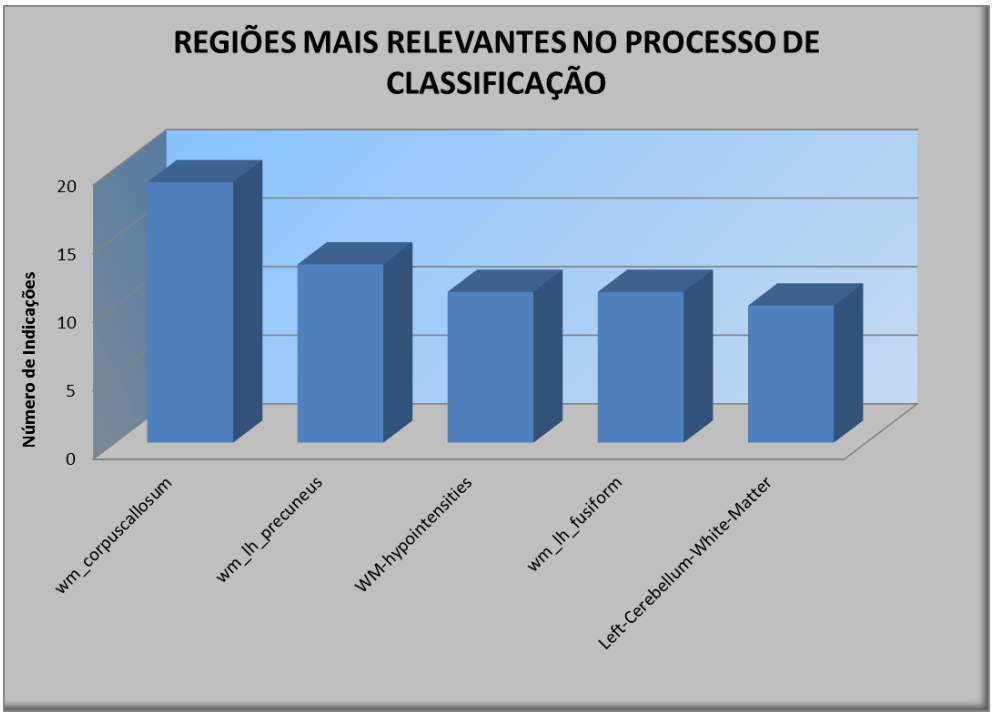

Figura 3.10 – Regiões cerebrais com maior número indicações de melhores atributos na classificação usando os vetores técnicas quantitativas. O máximo número de indicações possíveis nos testes executados seria 48. Fonte: Elaborada pelo autor.

A figura 3.11 mostra os principais atributos encontrados considerando a redução de regiões sugeridas.

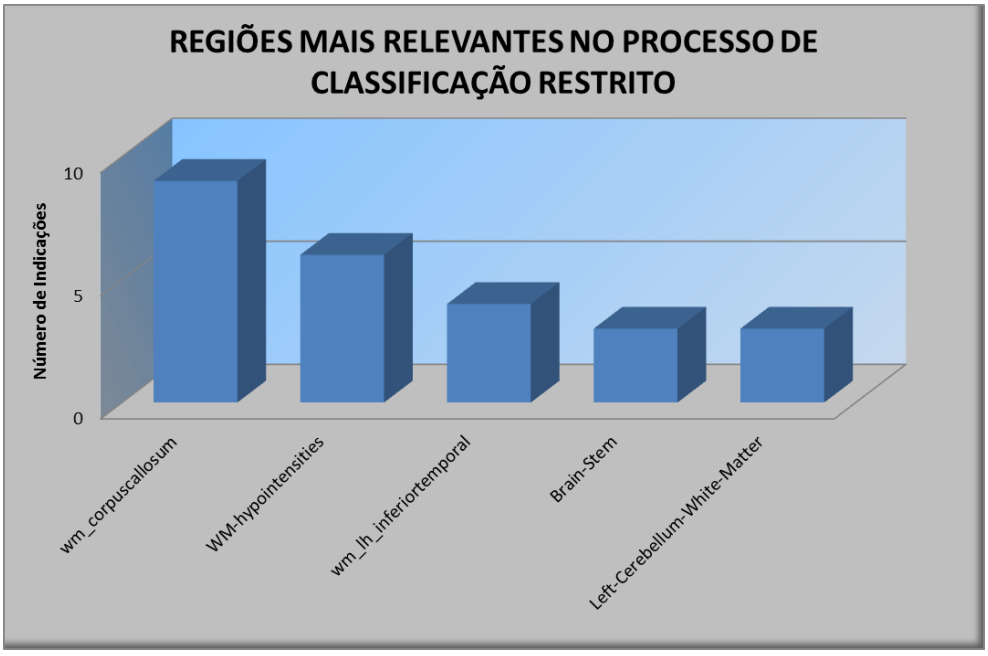

Figura 3.11 – Regiões cerebrais com maior número indicações de melhores atributos na classificação restrita usando os vetores técnicas quantitativas. O máximo número de indicações possíveis nos testes executados seria 12. Fonte: Elaborada pelo autor.

Conforme mostrados nas figuras 3.10 e 3.11 as classificações principal e restrita tiveram algumas regiões em comum numa comparação de posições no *ranking*. O *wm_corpuscallosum* apareceu em 1º lugar em ambas as classificações, sendo considerada a região mais afetada pela EM e o melhor atributo para separação das classes. O atributo *WM_hypointensities* também ganhou destaque por aparecer em ambas as classificações ficando em 2º lugar na classificação restrita e em 3º lugar na classificação principal. A região *Left-Cerebellum-White-Matter* também apareceu nas duas classificações na 5ª posição. As regiões *wm_lh_inferiortemporal* e *Brain-Stem* (Tronco Cerebral), que já tinham sido pré-selecionadas pelo neurologista para a classificação restrita, apareceram em 3º e 4º lugares respectivamente apenas na classificação restrita. Na classificação principal observam-se com destaque as regiões *wm_lh_precuneus* e *wm_lh_fusiform* que apareceram em 2º e 4º lugares respectivamente, configurando como regiões não mencionadas na classificação restrita e que ganharam importância no *ranking* da classificação principal.

3.2 Classificação

Depois de realizada a seleção dos melhores atributos, as tarefas de classificação foram executadas considerando diferentes atributos e parâmetros do algoritmo KNN. A principal ferramenta de análise em cada classificação foi a curva ROC, pois indica os valores de sensibilidade e especificidade trazendo a porcentagem de acertos e erros da classificação através da matriz de confusão.

Nas figuras 3.12, 3.13 e 3.14 são mostrados as porcentagens de área sob a curva ROC usando KNN1, KNN3 e KNN5, respectivamente. Nestes três gráficos observa-se que a classificação considerando todos os atributos, sem nenhuma seleção previa, teve o pior desempenho entre todas as outras classificações. Isto já poderia ser esperado e sugere que a escolha previa de atributos é um passo a ser considerado e ainda ressalta o significado de alguns atributos.

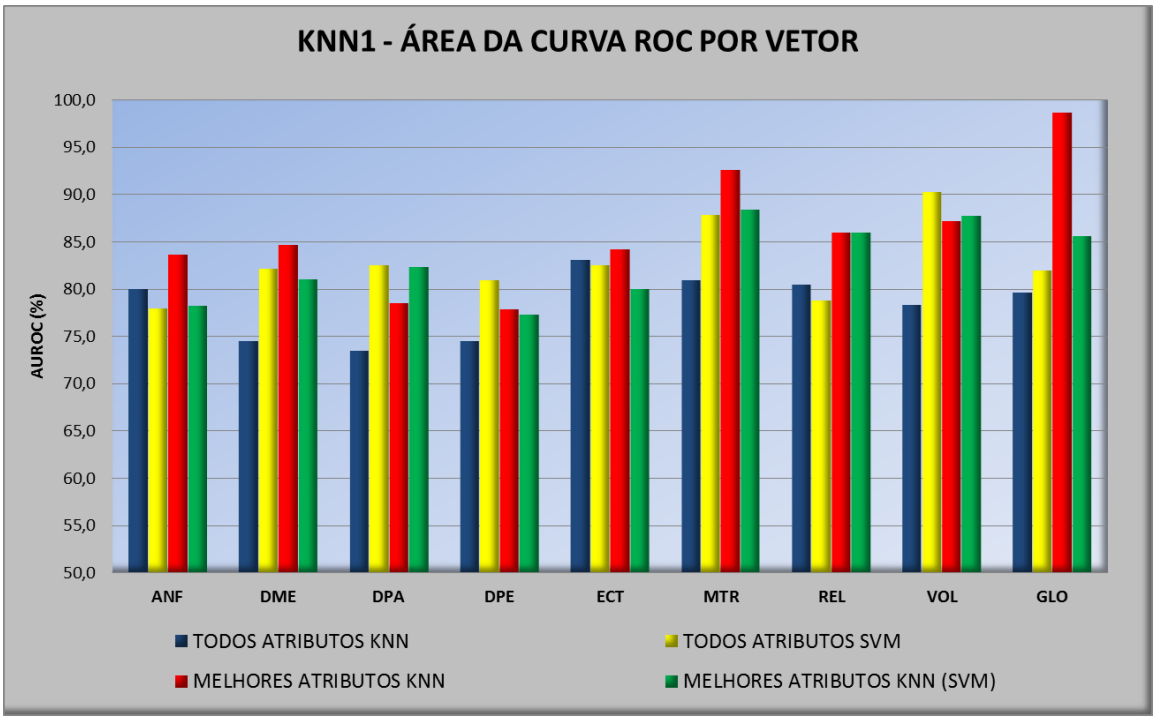

Figura 3.12 – Valores de AUROC por vetor de atributos usando KNN1 e considerando diferentes esquemas de pré-seleção de atributos.
Fonte: Elaborada pelo autor.

De acordo com os gráficos, as classificações com seleções de atributos tiveram melhores desempenhos em comparação com a classificação sem seleção de atributos. Em geral, a classificação por KNN5 teve o melhor desempenho geral em relação a KNN1 e KNN3. A classificação por todas as instâncias teve um desempenho melhor que a validação cruzada provavelmente por treinar todas as instâncias, diferente da validação cruzada que por treinar grupos, estes podem apresentar características particulares, interferindo no resultado final da classificação.

Apesar da classificação sem seleção de melhores atributos apresentar a menor porcentagem de acertos, esta classificação foi feita com os dois algoritmos, ou seja, o KNN e o SVM com isso foi possível a comparação entre eles. Em geral, para KNN1 a seleção de todos os atributos por SVM teve um melhor desempenho em relação à seleção de todos os atributos por KNN. Para as classificações KNN3 e KNN5 a seleção de todos os atributos por KNN teve um melhor desempenho em relação à seleção de todos os atributos por SVM.

A seleção de melhores atributos por KNN em geral teve melhor desempenho do que a seleção de melhores atributos por SVM. Apenas em algumas técnicas na classificação KNN1 foi que a seleção de melhores atributos SVM teve melhor desempenho que a seleção de melhores atributos por KNN.

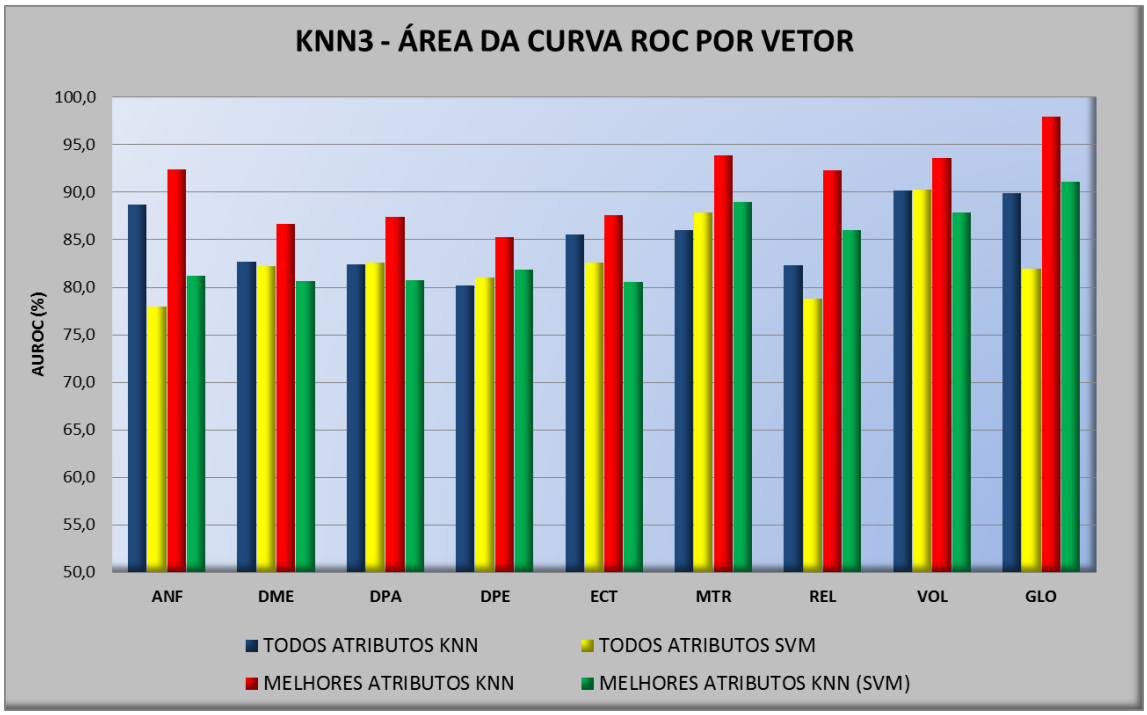

Figura 3.13 – Valores de AUROC por vetor de atributos usando KNN3 e considerando diferentes esquemas de pré-seleção de atributos.
Fonte: Elaborada pelo autor.

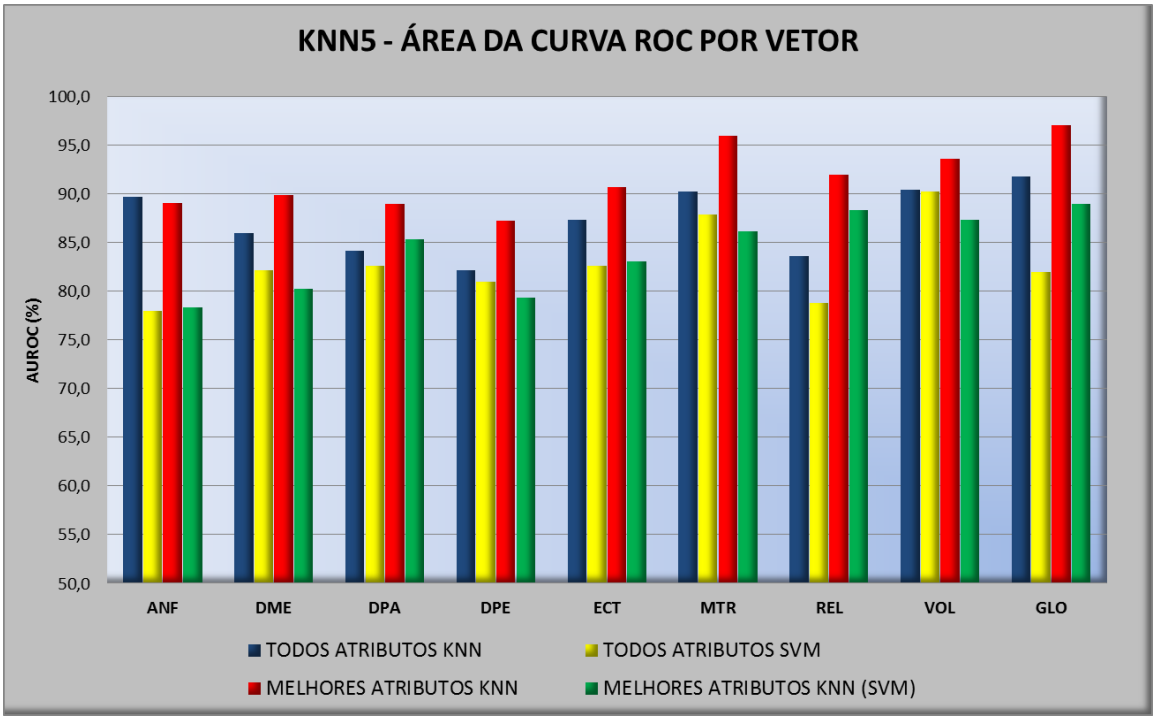

Figura 3.14 – Valores de AUROC por vetor de atributos usando KNN5 e considerando diferentes esquemas de pré-seleção de atributos.
Fonte: Elaborada pelo autor.

Dentre as técnicas quantitativas a MTR foi a que atingiu o maior valor de área sob a curva ROC (AUROC) pela seleção de melhores atributos KNN5, atingindo 96% de acertos.

Porém, o vetor global foi o que atingiu as maiores valores de área sob a curva ROC dentre os outros vetores pela seleção de melhores atributos. Surpreendentemente a seleção de melhores atributos por KNN1 no vetor global foi a que obteve a maior área sob a curva ROC porcentagem de acertos, chegando a 98,7% seguidos pelo KNN3 com 98% e KNN5 com 97,1%.

De acordo com a figura 3.15 que configura uma visão diferente dos classificadores por técnica quantitativa, dentre todos os classificadores o SVM foi o que obteve os piores desempenhos chegando ao máximo de 91,1% de acertos no vetor global. Dentre os classificadores KNN o que obteve o pior número de acertos foi o KNN1 por técnicas ficando sempre abaixo de 93%, com exceção no vetor global que alcançou 98,7 % de acertos. O KNN5 teve os melhores resultados por técnica em geral, chegando a 96% de acertos em MTR. O vetor global foi o que apresentou as maiores porcentagens de acertos chegando a 98,7% com KNN1, 98% com KNN3, 97,1 com KNN5 e 91,1% com SVM. Também, de acordo com o gráfico, a pior técnica quantitativa para classificação foi a difusibilidade perpendicular com valores abaixo de 88% e a melhor técnica quantitativa foi a transferência de magnetização com acertos chegando a 96%.

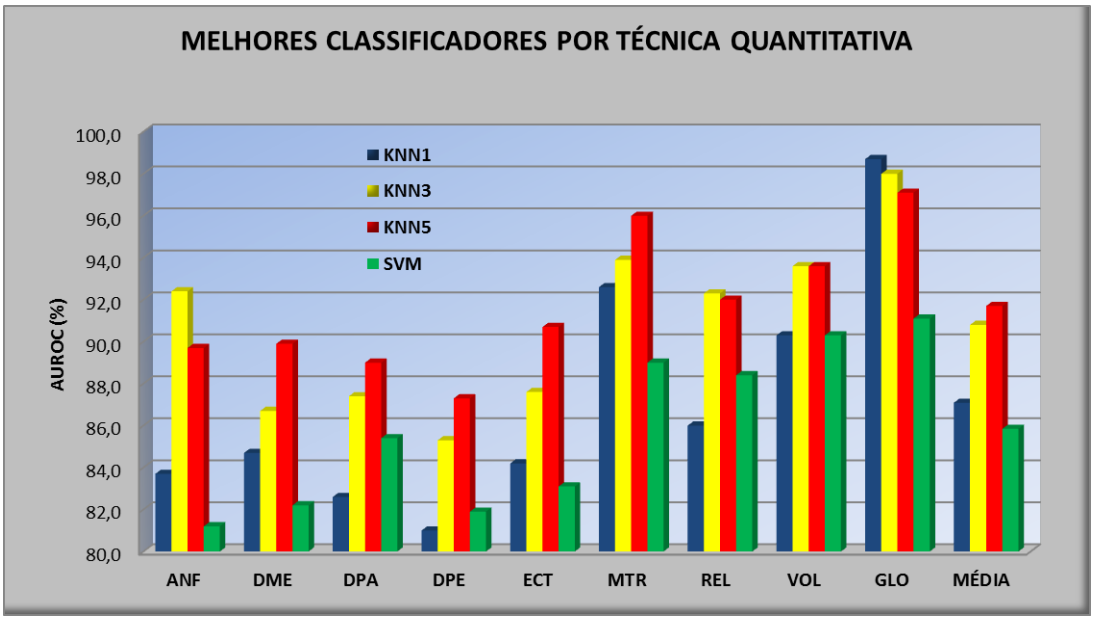

Figura 3.15 – Áreas sob a curva ROC por vetor e por algoritmo de classificação, considerando a melhor estratégia de pré-seleção de atributos em cada caso. Também está incluída a média por classificador.

Fonte: Elaborada pelo autor.

Entre as técnicas a melhor de todas foi a MTR seguida pela Volumetria com maior número de acertos. A pior delas foi a Difusibilidade Perpendicular. O SVM apresentou o pior

desempenho geral entre todos, ficando em quarto lugar. Em terceiro lugar aparece o KNN1, em segundo o KNN3 e em primeiro lugar configurando como sendo o melhor classificador em geral foi o KNN5 por não apresentar muitas diferenças e conter a maior área sob a curva ROC.

Na figura 3.16 são mostrados alguns gráficos que mostram o comportamento da matriz de confusão nas melhores situações de classificação. O gráfico (A) mostra a melhor classificação encontrada que foi para o vetor global com KNN1 aonde a área sob a curva ROC chegou a 98,7% errando apenas 1 paciente de EM. O gráfico (B) mostra a melhor classificação MTR com KNN de 5 que atingiu 96% na área sob a curva ROC.

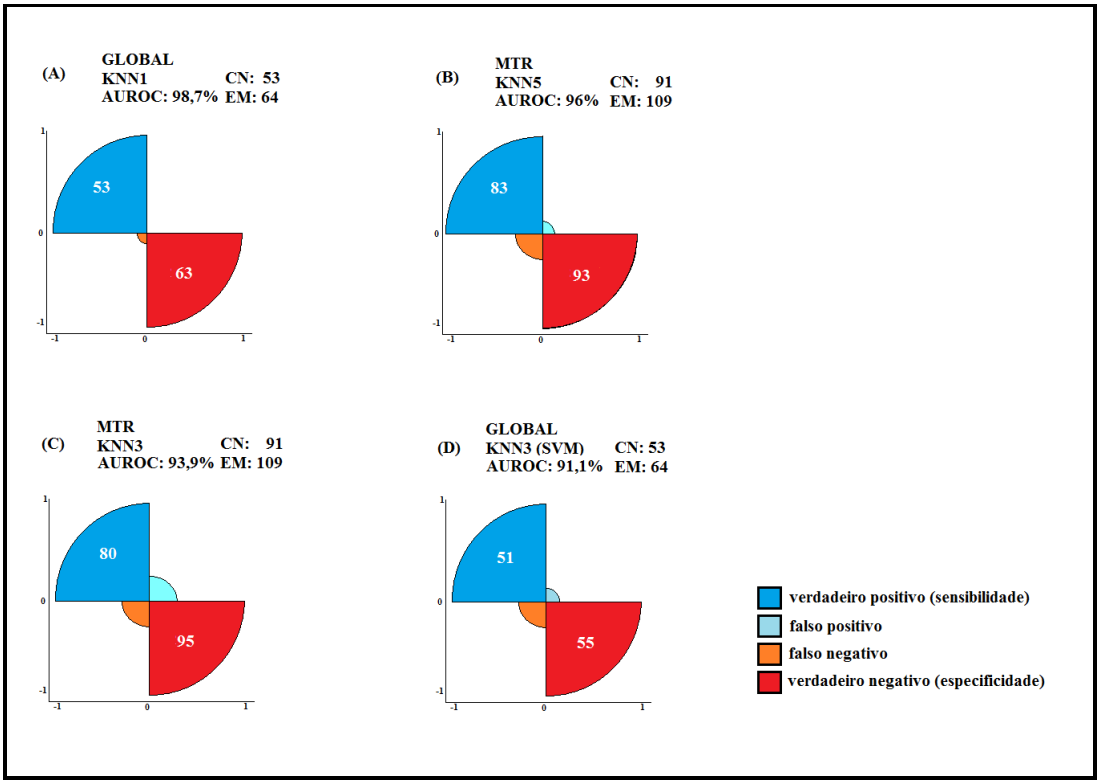

Figura 3.16 – Valores de sensibilidade e especificidade nas melhores classificações usando os vetores Global ou MTR.

Fonte: Elaborada pelo autor.

O gráfico (C) mostra outra melhor classificação MTR com KNN3 que teve 93,9% de área sob a curva ROC. E no gráfico (D) têm-se outra melhor classificação do vetor global com KNN3, porém utilizando o SVM que atingiu 91,1% de área sob a curva ROC.

Em um modelo 3D, na figura 3.17 são indicadas as 4 regiões anatômicas, segundo o atlas aqui utilizado, que apareceram envolvidas nos melhores atributos na classificação dos diferentes vetores das técnicas (Figura 3.10), elas são: *wm_corpuscallosum*, *wm_lh_precuneus*, *wm_lh_fusiform* e *Left-Cerebellum-White-Matter*.

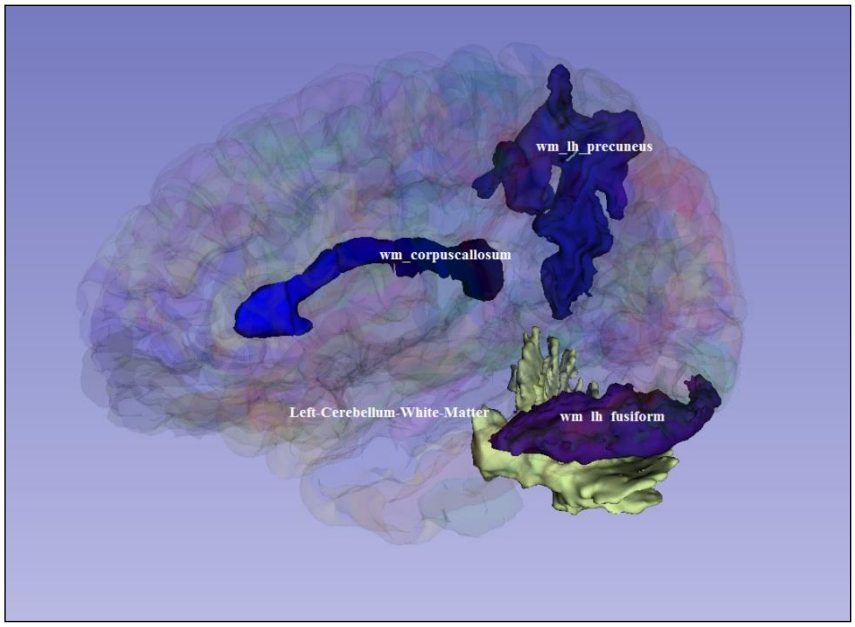

Figura 3.17 – Modelo em 3D do cérebro indicando a localização das quatro regiões com os melhores atributos na classificação usando os diferentes vetores das técnicas. Fonte: Elaborada pelo autor.

A figura 3.18 mostra outro conjunto de regiões do cérebro que também apareceram envolvidas nos melhores atributos na classificação dos diferentes vetores das técnicas, porém com menos destaque que as anteriores da figura 3.17. Dentre elas estão *Right-Cerebellum-WM, Right-VentralDC, Left-Hippocampus, lh-lingual e lh-precentral*.

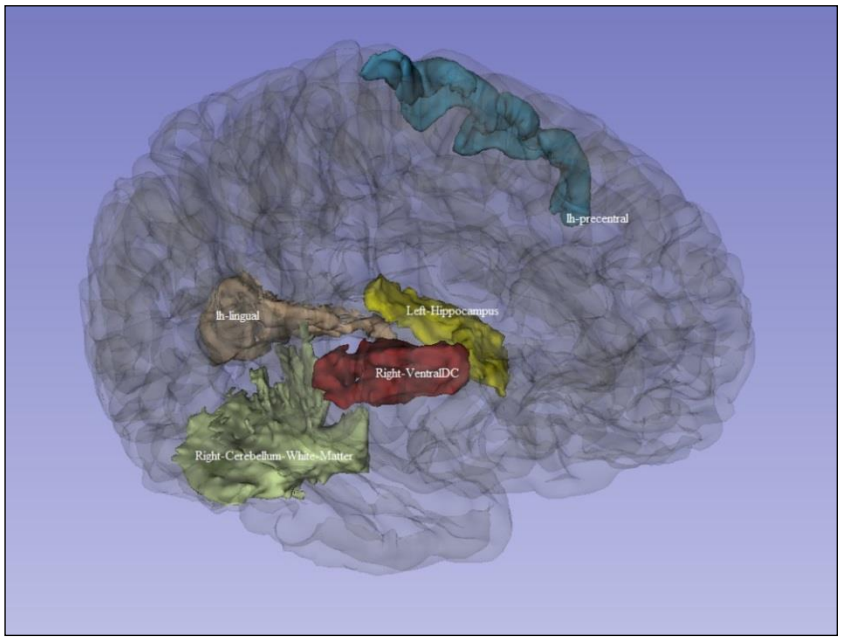

Figura 3.18 – Modelo em 3D do cérebro indicando a localização de outras regiões envolvidas nos melhores atributos na classificação usando os diferentes vetores das técnicas, porém com menor relevância que as apresentadas na figura 3.17. Fonte: Elaborada pelo autor.

A classificação geral também foi comparada com a classificação restrita. O quadro 3.1 representa esta comparação e um resumo das melhores classificações. Em geral, as médias da classificação geral foram maiores que as médias da classificação restrita. No entanto, a diferença entre as médias foi baixa, mostrando certa equivalência entre as duas classificações.

Quadro 3.1 – Comparativo de porcentagem dos valores de AUROC para as classificações geral e restrita.

KNN5	**CLASSIFIC. GERAL (%)**		**CLASSIFIC. RESTRITA (%)**	
	TA	**MA**	**TA**	**MA**
ANF	89,7	92,0	76,5	87,4
DME	86,0	90,1	85,9	90,1
DPA	84,2	91,9	79,3	88,9
DPE	82,2	89,2	84,2	90,0
ECT	87,4	91,0	82,2	85,0
MTR	90,3	96,5	87,1	91,0
REL	83,6	92,0	83,5	90,6
VOL	90,5	95,5	90,6	96,5
GLO	91,8	97,8	91,7	97,3
MÉDIA	**87,3**	**92,9**	**84,6**	**90,8**

Fonte: Elaborada pelo autor.

Para KNN5 a Classificação Geral teve média de acertos de 87,3% para todos os atributos (TA) e 92,9% para os melhores atributos (MA). Na Classificação Restrita a média de acertos foi de 84,6% para TA e 90,8% para MA. Os resultados dessas classificações se mostraram bem próximos em geral.

A partir do item 4 são discutidos os principais resultados encontrados, divididos por melhores atributos e classificações.

4. DISCUSSÃO

A principal característica para que os classificadores encontrem os melhores atributos é que os valores que eles representam devem ser bem diferenciados entre os grupos. No caso deste estudo os classificadores basearam-se na diferença de dados entre as mesmas regiões do cérebro de indivíduos saudáveis e pacientes EM. Consideramos que essas diferenças tenham sido causadas pela doença visto que o efeito envelhecimento foi minimizado no pareamento dos grupos (Quadro 2.1). Existem mais mulheres do que homens entre os pacientes EM com uma proporção de 2,27. Em geral a EM tem uma predominância maior entre as mulheres (CREE, 2007; HAUSER; GOODWIN, 2008;), o que acabou aparecendo nesta casuística, pois dentre os 144 pacientes de EM, 100 foram mulheres e o restante homens.

A partir da figura 2.1 pode ser notado que na a casuística é constituída majoritariamente por pacientes em um estágio avançado da doença. Esta característica da amostra deve-se a sua origem de um hospital do setor terciário da saúde publica. Ter muitos casos da doença em estágio avançado foi útil para a classificação, pois tivemos muitos dados alterados das regiões do cérebro devido às lesões em pacientes EM, o que contribuiu para facilitar a descoberta de atributos relevantes.

Em geral, o resultado da classificação acertou em indicar que as regiões mais afetadas com a EM estavam contidas na substância branca do cérebro, em concordância com a literatura (KALB, 2000; SANTOS, 2007; GOLDENBERG, 2012). Para as técnicas quantitativas os valores das médias dos atributos mais relevantes nas classificações parecem estar de acordo com o esperado para cada técnica. As classificações utilizando KNN e SVM puderam analisar os dados sob diversos aspectos e trouxeram resultados condizentes com a metodologia aplicada.

A partir do item 4.1 são discutidos os principais aspectos dos resultados encontrados neste estudo divididos na mesma forma como eles foram expostos para facilitar a compreensão.

4.1 Melhores Atributos

A partir da utilização dos dados fornecidos por todas as técnicas, os melhores atributos foram: *VOL_WM-hypointensities* e *MTR_wm_cospuscallosum* (Figura 3.1). O FreeSurfer não segmenta diretamente lesões de EM, porém consideramos que os resultados aqui obtidos sugerem que este *label* pode representar áreas com algumas lesões desmielinizantes. Neste

label a diferença das médias no volume normalizado entre pacientes e controles normais chegou 0,60 % sendo um forte indicador da diferença entre os grupos (Figura 3.1). O alto desvio padrão deste *label* nos pacientes (0,51%) é um indicativo de indivíduos em diferentes estágios da doença. Já nos sujeitos saudáveis o valor médio foi um quase insignificante 0,14% com um pequeno desvio padrão (0,05%), isto poderia ser associado a pequenos artefatos sistemáticos na imagem que ocasionem hipointensidades e não a lesões como nos pacientes.

O valor de MTR reflete sobre a integridade mielínica e o corpo caloso é uma estrutura com uma alta densidade de fibras. Desta forma, é totalmente esperado que um dos principais atributos na separação dos grupos seja o *MTR_wm_cospuscallosum.* No caso dos pacientes existe uma diminuição de 4.8% em relação aos controles, denotando perda mielínica. Novamente a maior variabilidade nos pacientes sugere os diferentes estágios da doença.

Ao utilizar os dados fornecidos pelas técnicas individuais (Figuras 3.2-9), os melhores atributos obtidos na separação dos grupos sugerem as diferentes tendências de cada um dos parâmetros quantitativos. Em geral no grupo de pacientes pode ser notado: uma diminuição da anisotropia fracionada (Figura 3.2), um aumento da difusibilidade média (Figura 3.3), paralela (Figura 3.4) e perpendicular (Figura 3.5) e uma redução no valor da transferência de magnetização (Figura 3.7). Os achados anteriores são indicativos de dano mielínico em diversas regiões da substância branca do cérebro.

O aumento no valor de T2 no grupo de pacientes (Figura 3.8), também em regiões com predominância de substancia branca aparentemente normal, indica geralmente um aumento no conteúdo de água, sugerindo inflamação e edema (TOFTS, 2003). Deve ser destacado que não apareceu como melhor atributo variações de T2 no *label WM-hypointensities*. Isto pode ser atribuído aos diferentes estágios das lesões no grupo de pacientes estudado (Figura 2.1), pois em lesões agudas existe um aumento de T2 e já em lesões crônicas existe um maior espectro dos valores de T2.

As figuras 3.6 e 3.9 indicam atrofia em regiões corticais predominantemente a esquerda e uma região subcortical a direita (diencéfalo ventral) dos pacientes. Devemos notar que o diencéfalo não é uma estrutura anatômica única e sim compreende um conjunto de estruturas cuja segmentação fica difícil. Estas atrofias são indicativas de perda neuronal. A atrofia do corpo caloso é uma característica reconhecida na EM (KLAWITER *et al*., 2014) mesmo em pacientes com baixos níveis de deficiência e disfunção cognitiva.

Na classificação de dados de volumetria os atributos destacados foram *WM_hypointensities, Right_VentralDC e lh-lateraloccipital*. Os atributos *Right_VentralDC* (diencéfalo) e *lh-lateraloccipital* (localizado no lobo temporal) tiveram valores de média 0,28

e 0,73 % respectivamente contra 0,31 e 0,94 % de seus controles normais, constatando que também houve uma perda volumétrica dos pacientes. O atributo *WM_hypointensities* teve valores de média 0,66 % e desvio padrão 0,47 % para pacientes EM contra valores de média 0,14 % e desvio padrão 0,07 % para controles normais, ou seja, tiveram grande variação de média e desvio padrão, sendo este atributo da substância branca um dos mais afetados pela EM. Na pesquisa de (VOLLMER, 2015) as lesões foram variáveis entre os estudos o que pode ser devido às diferenças no algoritmo de volume cerebral e volumes das lesões, bem como o curso da doença. Apesar de a perda de volume cerebral ser maior no início da doença, observaram que em alguns casos as taxas de perda de volume cerebral se abrandaram durante o curso da EM.

Na análise regional resultante da seleção dos melhores atributos a partir dos dados individuais de cada técnica, cinco regiões foram consideradas mais relevantes no processo de classificação como mostrado na figura 3.10. A região *wm_corpuscallosum* foi a que mais apareceu na seleção de melhores atributos com uma quantidade de 19 vezes O *label wm_hypointensities* apareceu 11 vezes na seleção de melhores atributos indicando áreas bem afetadas pela EM. Este resultado era quase esperado a partir dos resultados do vetor global. Ele confirma o comentário anterior de serem áreas representativas de lesões de EM. A região *Left-Cerebellum-White-Matter* também apareceu na seleção de melhores atributos com uma quantidade de 10 vezes. De acordo com (TORNES; CONWAY; SHEREMATA, 2014), as manifestações no cerebelo em EM são comuns, especialmente quando a doença se torna progressiva. Os pacientes que possuem manifestações cerebelares no início da doença tendem a ter um curso mais agressivo, com maior deficiência e mais carga da doença.

Ao considerar a redução de atributos, em termos de melhores atributos a seleção trouxe quase todas as regiões encontradas na classificação principal (Figura 3.11). O atributo que mais apareceu foi o *wm_corpuscallosum* seguidos por *WM_hypointensities*, *wm_lh_inferiortemporal*, *Brain-Stem* e *Left-Cerebellum-White-Matter*. Em geral a classificação restrita teve 60% dos seus cinco melhores atributos iguais aos da classificação principal e 40% de atributos que não tinham sido selecionados entre os cinco melhores na classificação principal. Os atributos da classificação restrita que não apareceram entre os cinco primeiros na classificação principal foi o *wm_lh_inferiortemporal* e o *Brain-Stem* posicionados em 3º e 4º respectivamente, no entanto eles aparecem posicionados em 8º e 7º respectivamente no *ranking* da classificação principal (Apêndice R). Os atributos *wm_lh_precuneus* e o *wm_lh_fusiform* não apareceram na classificação restrita porque não foram selecionados pelo neurologista para a classificação restrita, porém mostraram-se

atributos importantes na separação das classes atingindo posições em 2º e 4º lugar na classificação principal. Posteriores revisões na literatura devem confirmar estes achados. Observa-se que as regiões encontradas na classificação por regiões foram bem próximas as da classificação principal. Desta forma, concluímos que o uso das regiões envolvidas na classificação restrita conjuntamente com o *precuneus* e o *fusiform* seria a estratégia mais indicada para um eficiente processo de classificação de controles normais e pacientes EM.

Nas duas classificações (principal e restrita) observa-se uma grande quantidade de melhores atributos as regiões do hemisfério esquerdo do cérebro, o que indica que as regiões mais afetadas são no hemisfério dominante, pensamento lógico e competência comunicativa.

4.2 Classificação

Em geral, a maior quantidade de acertos de classificações ocorreu para o vetor global alcançando valores próximos a 99% de área ROC com a seleção de melhores atributos KNN. Isto provavelmente ocorreu pela grande quantidade de atributos classificados, onde a classificação foi feita para um número maior de valores, tendo maiores chances de comparação e definição de melhores atributos de classificação. Em geral, as classificações feitas com todos os atributos tiveram pior desempenho do que as classificações com seleção de melhores atributos como vistos nas figuras 3.12, 3.13 e 3.14. Isto ocorreu, pois a seleção de atributos KNN pôde analisar o vetor inteiro com seus parâmetros e descobrir quais são os atributos mais relevantes para a classificação. As técnicas MTR e volumetria obtiveram em geral os melhores resultados de classificação com alguma vantagem para MTR como vistos na figura 3.15. As classificações com as técnicas de difusão obtiveram em geral piores índices em relação às outras técnicas como vistos na figura 3.15.

A classificação dos vetores técnicas teve pior desempenho que a do vetor global porque contou com menor quantidade de atributos, tornando a classificação mais complexa e com maior seleção de atributos. Isto ocorreu provavelmente devido aos valores dos atributos serem muito aproximados também.

A figura 3.15 mostra as melhores classificações ocorridas por vetor. A melhor situação de classificação ocorreu para KNN5 alcançando uma média de 91,7% de acertos, seguido por KNN3 com 90,8%, KNN1 com 87,1% e SVM com 85,8%. Isto ocorreu possivelmente pela heterogeneidade da amostra como é na realidade clínica, onde muitos poucos vizinhos não seriam representativos do grupo. A pior situação de classificação ocorreu com SVM que não conseguiu definir a melhor margem entre as classes. Essa margem utilizada que contou com o

kernel linear, sendo basicamente uma reta na tentativa de melhor separar as classes. Acontece que uma margem reta não foi muito eficiente na separação, pois dificilmente as características das classes definem uma fronteira linear, sendo que ocorre na maioria dos casos existe uma grande mistura de classes. Outros tipos de *kernel* poderiam ter sido tentados, mas foram descartados pelo alto custo computacional; em trabalhos posteriores isto poderia ser considerado. Nas técnicas quantitativas que possuem características mais definidas, como é o caso da volumetria onde a perda de tecido cerebral é mais acentuada em pacientes EM a classificação SVM atingiu melhores resultados, chegando a valores acima de 90% de área ROC. O algoritmo KNN teve mais acertos em geral, pois definiu margens não lineares entre as características das classes, realizando uma melhor separação. E à medida que se aumentou o fator de KNN obtiveram-se melhores resultados ainda, pois foi possível delimitar melhor as margens de separação entre as classes.

A classificação de dados obteve em geral uma boa quantidade de acertos de indivíduos saudáveis e pacientes EM com no mínimo 73% de acerto da área da curva ROC. Isto deve entre outros fatores por se tratar de dados provenientes de imagens de RM com ROI determinado por um atlas predefinido pelo FreeSurfer que diminui o agrupamento entre os *voxels* e reduz a dimensionalidade do agrupamento de *voxels* com características anatômicas e funcionais semelhantes (LI *et al*., 2014).

A comparação de classificações geral e restrita trouxe resultados aproximados. Os resultados são mostrados no quadro 3.11 onde se tem que a classificação principal teve em geral uma quantidade de acertos maior do que classificação por regiões. Como esperado, a classificação principal mostrou que o vetor global é melhor que o vetor técnica. Isto aconteceu porque o vetor global disponibilizou todos os atributos para os classificadores e com isso pode-se realizar uma classificação mais apurada, comparando todos os atributos existentes e selecionando os mais importantes na separação entre pacientes e controles. Também foi possível perceber que a classificação por seleção de melhores atributos foi melhor do que a que não tem seleção de melhores atributos. Esta seleção de melhores atributos feita pelos classificadores pôde selecionar quais atributos são melhores para classificação das classes. Com isso a classificação se tornou mais rápida e eficiente, trazendo melhores resultados de acertos na classificação. As classificações mostraram quais são as melhores técnicas quantitativas para classificação que foram MTR e Volumetria. Devido às lesões da EM a MTR foi bem alterada e houve diminuição do volume cerebral e por isso são técnicas que tiveram resultados mais diferentes em relação às outras técnicas e em relação aos controles normais. A classificação restrita é boa, porém inferior à classificação principal. Isto porque a

classificação principal analisou um universo maior de atributos, contribuindo para uma melhor classificação.

A classificação principal teve 87,3% e 92,9% de média nos acertos para todos os atributos (TA) e melhores atributos (MA), respectivamente. A classificação por regiões trouxe 84,6% e 90,8% de média nos acertos para TA e MA, respectivamente. A classificação por regiões teve um melhor desempenho para as técnicas DPE e VOL. Em DPE atingiu respectivamente 84,2% e 90% na classificação restrita contra 82,2% e 89,2% da classificação principal. Em VOL atingiu respectivamente 90,6% e 96,5% na classificação restrita contra 90,5% e 95,5% da classificação principal. Para a classificação principal foram mostradas as melhores situações de classificação, não necessariamente sendo somente por KNN5.

Concluímos então que a classificação por regiões teve em geral um pior desempenho do que a classificação principal. Embora com métodos tradicionais feitos até o momento como comparações em teste t tinham-se como conhecidos estes parâmetros. Porém quando incorporamos mais regiões nós melhoramos a classificação, sugerindo que não somente estes são os mais envolvidos.

A classificação por técnicas teve comportamentos ligeiramente diferentes, de acordo com as especificações de cada fenômeno físico e os erros envolvidos na estimativa de cada parâmetro quantitativo. Os números puderam confirmar que a doença varia de acordo com a técnica quantitativa empregada.

A difusão das moléculas de água pode ou não sofrer mudanças ao longo do caminho. Se o meio estiver isotrópico as propriedades da água geralmente mantém os mesmos valores e intensidade independente da direção e sentido. Agora caso o meio seja anisotrópico as propriedades possuem valores diferentes segundo a direção considerada. Uma das causas do meio estar mais anisotrópico são as lesões que ocorrem no cérebro de pacientes com EM. Na figura 4.1 são mostrados exemplos de imagens de difusão.

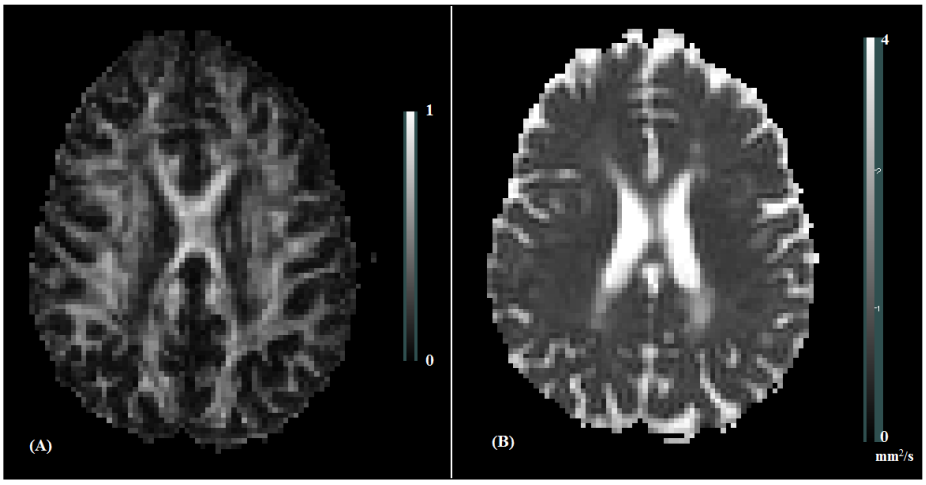

Figura 4.1 – Exemplos de mapas de neuroimagens obtidos a partir de dados de DTI de uma paciente de EM (feminino, 51 anos): (A) Anisotropia Fracionada ou Fraccional. (B) Difusibilidade Média.
Fonte: Elaborada pelo autor.

Valores próximos de zero significam que a difusão é mais isotrópica em todos os sentidos e valores mais distantes de zero indicam um meio mais anisotrópico. De acordo com (FILIPPI *et al.*, 2001) a difusão é afetada pelas propriedades do meio onde o movimento molecular ocorre e a medição da difusão nos tecidos biológicos fornecem informações sobre as estruturas dos tecidos em nível microscópico. Sendo assim o movimento da água pode ser dificultado pela presença de barreiras estruturais a níveis celulares. Os processos patológicos que alteram a organização do tecido dificultam a passagem das moléculas de água causando anormalidades na difusão da água. Na substância branca do cérebro a mobilidade molecular não é a mesma em todas as direções. Os resultados da classificação de difusão da água estão de acordo com os resultados esperados. Para a anisotropia fracionada os melhores atributos selecionados foram de SB e em todos os casos os valores de CN foram maiores que os valores de pacientes EM. Estes valores estão de acordo com os resultados de (FILIPPI, *et al.*, 2001) pois concluíram que a SB de pacientes EM tem maior difusibilidade e menor anisotropia fracionada do que a substância branca de controles normais. Isto provavelmente indica que o meio é mais anisotrópico para pacientes e menos anisotrópico para controles normais. Os valores de difusibilidade mostram o inverso, ou seja, os maiores valores ficaram com pacientes EM e os menores valores com CN. Isto indica que a difusão da água é mais prejudicada em pacientes e mantém a normalidade em controles normais. Segundo (FILIPPI, *et al.*, 2001) os elementos patológicos de EM tem o potencial de alterar as permeabilidade ou geometria de barreiras estruturais à difusão molecular da água no cérebro. Coerente com isso os resultados do estudo constatou que a difusibilidade da água em lesões da EM é maior do que nas regiões de SB, onde é maior na SB de voluntários saudáveis. Os resultados nos levam

a dizer que possivelmente a redução da inflamação e remielinização podem ser os mecanismos que levam a lesões ou inversamente a falha de reparação pode ser o mecanismo que conduz as lesões com prejuízos irreversíveis. O estudo de (FILIPPI *et al*., 2001) demonstrou que a difusão da água correspondem a áreas com aumento da permeabilidade da barreira sangue-cérebro e são associadas a edema, inflamação, desmielinização e perda axonal. A acumulação de células inflamatórias e a decomposição de produtos da mielina poderiam potencialmente restringir a difusão de água e diminuir a anisotropia fracionada por causa da presença de barreiras à difusão. Isto está de acordo com os resultados de classificação de anisotropia fracionada, pois mostram valores menores de pacientes em relação a controles normais. De acordo com (FILIPPI *et al*., 2001) as áreas mais afetadas pela EM são regiões da SB do cérebro, possivelmente ocasionando hiperplasia difusa, edema desigual, infiltração perivascular, glioses, mielina anormalmente fina e perda axonal. Essas mudanças na SB são generalizadas, mas tendem a ser mais graves em locais como áreas periventriculares e SB do lobo frontal.

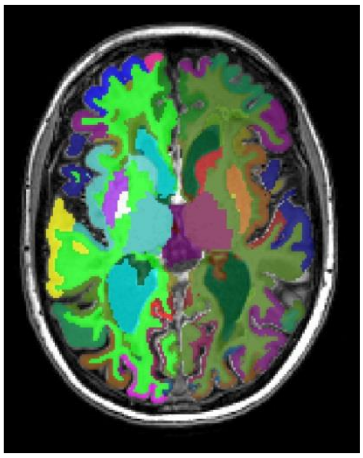

Figura 4.2 – Exemplo do resultado de uma segmentação de estruturas corticais e subcorticais realizada pelo FreeSurfer a partir de uma imagem por RM ponderada em T1 de uma paciente de EM, feminina, 47 anos.
Fonte: Elaborada pelo autor.

A espessura cortical é uma medida morfométrica utilizada para descrever a espessura combinada das camadas do córtex cerebral (THORSTENSEN; HOFER; SAPIRO, 2006). Ela é muitas vezes um indicativo da capacidade cognitiva do indivíduo. Em muitos casos a EM provoca atrofia cerebral difusa que é o resultado da redução do tecido tanto na SC quanto SB (SAILER *et al*., 2003). A técnica de espessura cortical consegue muitas vezes avaliar a perda de volume do córtex cerebral em pacientes EM. Na figura 4.2 é mostrado um exemplo de neuroimagem de Espessura Cortical e Volume.

Os resultados da classificação dos dados de espessura cortical mostram as regiões *lh_fusiform, lh_lingual, pericalcarine* e *lh_precentral* como mais relevantes na classificação entre controles normais e pacientes EM. Em todos os casos os valores de controles normais foram maiores do que os de pacientes EM. Isto indica que nessas regiões mostradas houve provavelmente uma atrofia de volume do córtex devido às lesões provocadas pela EM. De acordo com (SAILER *et al*., 2003) a visualização de diferença estatística entre espessura e toda a superfície cortical revelou um distinto padrão de atrofia cortical em pacientes com EM. A espessura cortical média também foi reduzida em pacientes EM em comparação com CN. No estudo de (SAILER *et al*., 2003) na região frontal a espessura média de pacientes EM foi de 2,37 mm e de CN 2,73 mm, ou seja, houve uma redução em pacientes EM. Na classificação deste estudo, um dos atributos selecionados foi o *lh_precentral* que fica também na região frontal sendo que pacientes EM ficaram com média 2,45 mm e controles normais com média de 2,55 mm, ou seja, também houve redução em pacientes EM. Para as regiões temporais em (SAILER *et al*., 2003) a espessura cortical foi de 2,65 mm em pacientes EM e 2,95 mm em CN. Neste estudo o atributo *lh_fusiform* também fica na região temporal e teve uma média de 2,61 mm para pacientes EM e 2,87 mm de média para controles normais, ou seja, também houve uma redução nesta região temporal. De acordo com (SAILER *et al*., 2003) cinco pacientes EM com incapacidade grave demonstraram semelhantes atrofias corticais nas regiões temporal parietal e frontal em comparação com pacientes de capacidade moderada. Neste estudo não foi realizado uma escala de incapacidade da doença, no entanto o atributo *lh_precentral* apareceu na classificação e fica próximo dessas regiões. Na EM o aumento do volume da lesão causa uma perda irreversível nos tecidos possivelmente levando a um dado neuronal com transecção de axônios e subsequente lesão axonal de neurônios corticais. A maior parte das lesões evolui dos ventrículos para as áreas frontais, temporais e motoras do córtex. Para fins clínicos e de monitoramento, os resultados também demonstraram a utilidade in vivo das medições da atrofia cortical em outros do que aqueles associados com distúrbios neurodegeneração clássica.

A volumetria é utilizada para estudos de atrofia e na EM ela tem sido utilizada para medir a perda de volume cerebral causadas pelas lesões e as próprias lesões de EM. Também de acordo com (VOLLMER, 2015) os resultados mostraram que as taxas de perda de volume cerebral em pacientes EM estão acima do esperado para o envelhecimento normal. Verificaram que a queda mais acentuada no volume do cérebro de pacientes EM acontece na fase precoce da doença, onde existe maior nível inflamatório da doença. Fizemos uma classificação de gravidade de pacientes EM conforme a figura 2.1 que indica 12% de

pacientes em uma fase inicial da doença, porém nos baseamos por vários estágios da doença. Em (VOLLMER, 2015) observaram que a perda volumétrica em pacientes EM foi de 0,7% ao ano sendo que no envelhecimento normal esta taxa é de 0,1 a 0,3% ao ano, constatando que a perda de volume cerebral de pacientes EM é maior em relação a indivíduos saudáveis.

De acordo com (RORIZ, 2010) a taxa de transferência de magnetização é uma importante técnica para avaliar a integridade mielínica na substância cerebral, por isso ela foi utilizada nesse sentido. Também segundo (SANTOS, 2007) a MTR é sensível à destruição das macromoléculas de membranas, o que no SNC ocorre com a mielina. Na figura 4.3 são mostrados exemplos de imagens MTR e de Relaxometria.

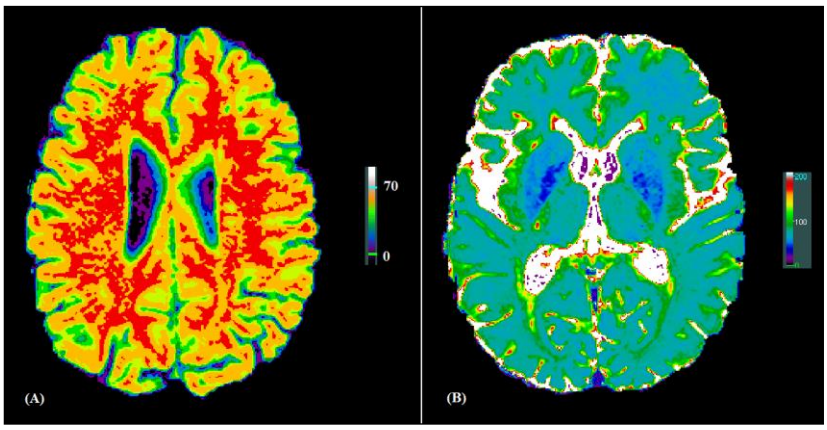

Figura 4.3 – Exemplos de mapas de neuroimagens obtidos a partir de dados de IRMQ de uma paciente de EM (feminino, 51 anos): (A) MTR (B) Relaxometria.
Fonte: Elaborada pelo autor.

A redução da medida da MTR indica desnaturação da mielina. O MTR mostra-se um bom indicador da lesão tecidual seja na SB ou SC aparentemente normal, quer seja nas lesões visíveis, sendo este parâmetro capaz de diferenciar pacientes EM de controles normais Os resultados da classificação do vetor MTR mostrou que grande parte dos atributos encontrados foi de substância branca cerebral. Dentre os valores destes atributos observou-se que os valores de pacientes EM foram menores do que de controles normais. Isto pode ser explicado porque as lesões e desestruturações das macromoléculas nos pacientes fazem com que haja uma queda proporcional na taxa de transferência de magnetização (KUCHARCZYK *et al*., 1994). Essas lesões podem significar desmielinização e lesão axonal na SB e perda de membrana plasmática na SC (KUCHARCZYK *et al*., 1994). Em comparação com o estudo de (CERCIGNANI *et al*., 2001) os resultados de pacientes EM também foram menores que os de controles normais e as regiões mais afetadas foram de SB do cérebro. Algumas regiões de SC também foram afetadas em pacientes EM (CERCIGNANI *et al*., 2001), bem como nesse

estudo que a classificação trouxe as regiões *Left-Hippocampus e Brain-Stem* que também fazem parte da SC. Algumas mudanças sutis que ocorrem na SB segundo (CERCIGNANI *et al*., 2001) pode ser a hiperplasia difusa de astrócitos, edema desigual, infiltração perivascular, gliose, mielina anormalmente fina, perda axonal, entre outros. Todas estas mudanças podem modificar as relações de proporções de água livre e ligada, bem como a quantidade de restringir barreiras presentes na SB, influenciando a sua MT.

A Relaxometria mede o tempo de relaxação dos tecidos, sendo que dependendo do tecido ou região do cérebro, existem diferentes tempos de relaxação. Principalmente, em regiões que foram lesionadas ocorre um aumento do tempo de relaxação devido a problemas ocasionados nessa região (SOUZA, 2010). Nos resultados de classificação de Relaxometria observamos que em todos os casos os valores de pacientes foram maiores do que de controles normais. Isto porque provavelmente as lesões ocasionadas pela EM aumentou a concentração de água nos tecidos podendo indicar edemas, desmielinizações, glioses ou disfunções axionais. E com isto o tempo de relaxação dessas regiões aumentou principalmente devido aos danos ocorridos na região lesionada. Segundo a pesquisa de (NEEMA *et al*., 2009) as áreas pesquisadas e que possivelmente teriam mais lesões foram o lobo frontal, o lobo parietal e o corpo caloso e que uma análise regional mostrou que os danos na SB foi generalizado, envolvendo áreas frontal, parietal e corpo caloso. Neste estudo evidenciou que provavelmente essas regiões realmente foram afetadas, pois a classificação trouxe entre outros os atributos de SB tais como *wm_corpuscallosum* e *wm_lh_fusiform*. A substância branca parece ter sido a maior afetada neste estudo, assim como em (NEEMA *et al*., 2009) onde as medições foram capazes de detectar mudanças estruturais e na SB tornando a anormal em pacientes com EM tanto a nível global e regional. Os achados deste estudo possivelmente também mostram uma perda substancial das fibras axonais do corpo caloso em paciente com EM assim como em (NEEMA *et al*., 2009). Também de acordo com (NEEMA *et al*., 2009) tais achados podem representar desmielinização, inflamação, proliferação glial e perda axonal.

5. CONCLUSÕES

De acordo com os resultados das classificações os melhores atributos encontrados, configurando como melhores regiões do cérebro para separar indivíduos saudáveis e pacientes EM foram: *wm_corpuscallosum*, *wm_lh_precuneus*, *WM-hypointensities*, *wm_lh_fusiform* e *Left-Cerebellum-White-Matter*. Como era esperado, são todas regiões da substância branca do cérebro, conhecida como a região mais afetada pela EM.

O algoritmo KNN5 foi considerado o melhor classificador para separar indivíduos saudáveis de pacientes EM resultando em 91,7% de valor médio de área sob a curva ROC. Através dos seus parâmetros foi possível alcançar em geral os melhores resultados de acertos da área da curva ROC, possibilitando a descoberta de atributos mais relevantes ao processo de classificação. O KNN3 também obteve bons resultados, configurando como segundo melhor classificador com resultados de 90,8% de valor médio de área sob a curva ROC, seguido em terceiro pelo KNN1 com 87,1% de valor médio de área sob a curva ROC. O SVM não alcançou bons índices de acerto, ficando na quarta posição entre os classificadores com resultados de 85,8% de valor médio de área sob a curva ROC, fato atribuído à utilização de um *kernel* linear. Porém o problema maior não foi a classificação e sim a definição dos atributos mais importantes para a classificação.

As classificações KNN e SVM sem seleção de atributos tiveram em geral os piores índices de acerto do que as classificações com seleção de atributos que obtiveram melhores índices de acerto. Isto já era esperado devido à escolha de atributos mais relevantes no processo de classificação os índices certamente aumentariam, o que sugere que esta estratégia é importante na utilização clinica desta metodologia.

O vetor global teve melhor desempenho em geral de acertos, principalmente na seleção de melhores atributos KNN. De acordo com os resultados, a técnica quantitativa com um maior poder de diferenciação foi a MTR e a mais fraca foi a difusibilidade perpendicular. O SVM teve um desempenho inferior comparado com o KNN no problema específico.

Em geral, todas as técnicas mostraram que a SB é a mais afetada pela EM. Entre as regiões mais afetadas o atributo *WM_hypointensities* teve a maior variação de resultados entre pacientes EM e controles normais, sendo determinado como a região mais afetada pelas lesões de EM e assim sugestivo das lesões.

Os resultados da classificação dos dados de difusão mostram que as lesões da EM provocam barreiras na difusão natural da água do cérebro, tornando o meio anisotrópico. Foi

o que mostrou os resultados menores de pacientes EM em anisotropia e maiores nas difusibilidades em comparação com indivíduos saudáveis.

Os resultados de espessura cortical e volumetria mostraram atrofias no córtex e no volume do cérebro devido às lesões de EM. Também mostraram que as taxas de perda de volume cerebral em pacientes EM estão acima do esperado para o envelhecimento normal. A queda mais acentuada no volume do cérebro de pacientes EM acontece mais na fase precoce da doença, onde existe maior nível inflamatório da doença.

A melhor técnica de classificação foi a MTR com 92,9% de valor médio de área sob a curva ROC, porém com uma análise global elevou-se a sensibilidade para 96,2% de valor médio de área sob a curva ROC. Os resultados da classificação do vetor MTR mostrou uma queda proporcional na taxa de transferência de magnetização nos pacientes em diversas regiões de substância branca aparentemente normal. Todas estas mudanças podem modificar as relações de proporções de água livre e ligada, bem como a quantidade de restringir barreiras presentes na substância branca, influenciando a sua MTR.

Os resultados de Relaxometria mostram que as lesões ocasionadas pela EM aumentou o conteúdo de água nos tecidos podendo indicar edemas, desmielinizações, glioses ou disfunções axionais. E com isto o tempo de relaxação dessas regiões aumentou principalmente devido aos danos ocorridos na região lesionada. Os resultados mostraram que os danos na substância branca foram generalizados, envolvendo áreas como o lobo frontal, o lobo parietal e o corpo caloso.

Em geral as técnicas mostraram que a EM causa: desmielinização, lesões na SB, edema, perda axonal, gliose, perda de volume cerebral, entre outros.

A classificação restrita trouxe em geral os mesmos atributos que a classificação principal, porém com um índice menor de acertos, atingindo acertos 2,1% de média menor que a classificação principal e 60% de acertos entre os cinco melhores atributos. Desta forma, uma estratégia sugerida neste problema de classificação seria o uso dos dados de todas as técnicas quantitativas aqui discutidas proveniente das regiões envolvidas na classificação restrita conjuntamente com o *precuneus* e o *fusiform*, aplicando o KNN5 com seleção de melhores atributos.

As técnicas de classificação se mostraram, em geral, uma ferramenta adequada para diferenciação entre indivíduos saudáveis e pacientes EM, indicando ainda os melhores atributos que se mostraram coerentes ao esperado do ponto de vista da fisiopatologia da doença.

De forma geral podemos concluir que as técnicas de reconhecimento de padrões são excelentes ferramentas na avaliação dos achados de imagem por Ressonância Magnética em Esclerose Múltipla. As aplicações das mesmas corroboraram que as regiões cerebrais mais afetadas nesta doença estão contidas na substância branca com um padrão de distribuição difuso, com maior envolvimento no hemisfério esquerdo. A técnica de transferência de magnetização foi confirmada como um importante biomarcador no processo de desmielinização. A classificação principal trouxe melhorias em relação à classificação restrita, sendo um importante recurso na separação dos grupos.

Este trabalho tem algumas limitações que poderiam ter sido abordadas como, por exemplo:

-Ter explorado mais os estágios da doença usando uma casuística bem maior de pacientes distribuída homogeneamente entre os diferentes estágios;

-Ter uma casuística maior de sujeitos controles mais idosos e do sexo feminino para efetuar um melhor pareamento;

-Poderia ter sido feita a seleção de melhores atributos utilizando o SVM para melhorar a classificação de dados e explorar outros *kernels* para o SVM e não somente o linear;

-Utilizar mais algoritmos de classificação para aumentar as chances de descobrir o melhor classificador.

Num trabalho futuro os resultados quantitativos das lesões EM poderiam ser utilizados em uma comparação das classificações. Também poderia ser realizado um trabalho sobre auxilio diagnóstico a partir do treinamento e classificação dos dados.

A classificação dos dados contribuiu para descobrir quais regiões são mais importantes a serem analisadas na esclerose múltipla e com isso pode-se diminuir o protocolo de aquisição, análise e processamento de imagens, otimizando o processo.

Pode-se também utilizar as técnicas de reconhecimento de padrões em dados quantitativos de outras doenças a fim de descobrir os comportamentos padrões específicas das técnicas quantitativas e da própria doença.

As lesões de EM também podem ser analisadas através do processamento das imagens pré e pós-contraste utilizando o reconhecimento de padrões para verificar se as lesões são crônicas ou inflamatórias.

6. REFERÊNCIAS BIBLIOGRÁFICAS

BAGNATO, F. et al. T1 cortical hypointensities and theirassociation with cognitive disability in multiple sclerosis. **Multiple Sclerosis**, v. 16, n. 10, p. 1203-1212, 2010.

BASSER, P. J.; PIERPAOLI, C. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. **Journal of Magnetic Resonance**, v. 111, n. 3, p. 209-219, 1996.

BENDFELDT, K. et al. Multivariate pattern classification of gray matter pathology in multiple sclerosis. **Neuroimage**, v. 60, n. 1, p. 400-408, 2012.

BLONDA, P. et al. Detection of multiple sclerosis lesions in MRI's with neural networks. **World Scientific,** p. 157-164, 2002.

CALABRESI, P. A. Diagnosis and management of multiple sclerosis. **Am Fam Physician**, v. 70, n. 10, p. 1935–1944, 2004.

CARMO, S. S. **Características do envolvimento do Sistema Nervoso Central na Polirradiculoneuropatia Inflamatória Desmielinizante Crônica: um estudo mediante técnicas quantitativas de Imagem por Ressonância Magnética**. 2014. 260 p. Tese (Doutorado em Neurologia) – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2014.

CERCIGNANI, M. et al. Magnetisation transfer ratio and mean diffusivity of normal appearing white and grey matter from patients with multiple sclerosis**. Journal of Neurology**, **Neurosurgery & Psychiatry**, v. 70, n. 3, p. 311-317, 2001.

CITIZENDIUM, F. **Cortical thickness**, London, 2010. Disponível em: <http://en.citizendium.org/wiki/Cortical_thickness>. Acesso em: 10 Out. 2015.

CRAIG, W. **M. D. Diffusion**, 2016. Disponível em: <http://highbloodpressure.about.com/od/glossary/g/diff_glos.htm>. Acesso em: 15 Jul 2015.

CREE, B. A. C. **Multiple sclerosis**. In: Brust JCM, ed. Current Diagnosis and Treatment in Neurology. New York: Lange Medical Books/McGraw-Hill Medical, 2007.

DAMANGIR, S. et al. Multispectral MRI segmentation of age related white matter changes using a cascade of support vector machines. **Journal of the Neurological Sciences**, v. 322, n. 2, p. 211-216, 2012.

DE-PAULA, V. et al. **Alzheimer's Disease**. São Paulo: Sub-cellular Biochemistry, 2012. p. 329-352.

FILIPPI, M. et al. Diffusion tensor magnetic resonance imaging in multiple sclerosis. **Neurology**, v. 56, n. 3, p. 304-311, 2001.

FREDERIKSEN, J. L.; LARSSON, H. B.; OLESEN, J. Correlation of magnetic resonance imaging and CSF findings in patients with acute monosymptomatic optic neuritis. **Acta Neurol Scand**, v. 86, n. 3, p. 317-322, 1992.

GEREMIA, E. et al. Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. **NeuroImage**, v. 57, n. 2, p. 378-390, 2011.

GIRALDI, G. A. et al. Statistical Learning Approaches for Discriminant Features Selection. **Journal of the Brazilian Computer Society**, v. 14, n. 2, p. 7-22, 2008.

GOLDENBERG, M. M. Multiple Sclerosis Review. **Pharmacy and Therapeutics**, v. 37, n. 3, p. 175-184, 2012.

HAACKE, E. M. et al. Imaging iron stores in the brain using magnetic resonance imaging. **Magnetic resonance imaging**, v. 23, n. 1, p. 1-25, 2005.

HALL, M.; FRANK, E.; HOLMES, G. et al. The Weka data mining software: an update. **SIGKDD Explorations**, v.11 p.10-18. Waikato. 2015. Disponível em: <http://www.cs.waikato.ac.nz/ml/weka/index.html>. Acesso em 05 Abr. 2015.

HAUSER, S. L; GOODWIN, D. S. **Multiple sclerosis and other demyelinating diseases**. eds. Harrison's Principles of Internal Medicine, vol. II, 17th ed. New York: McGraw-Hill Medical, 2008. p. 2611–2621.

KALB, R.C. **Esclerose Múltipla Perguntas e Respostas**. São Paulo: ABEM – Associação Brasileira de Esclerose Múltipla, 2000. 477 p.

KAMBER, M. et al. Model-based 3-D segmentation of multiple sclerosis lesions in magnetic resonance brain images. **Medical Imaging**, v. 14, n. 3, p. 442-453, 1995.

KLAWITER, E. C. et al. Corpus Callosum Atrophy Correlates with Gray Matter Atrophy in Patients with Multiple Sclerosis. **American Society of Neuroimaging**, v. 25, n. 1, p. 62-67, 2014.

KOHAVI, R.; JOHN G. H. Wrappers for featuresubsetselection. **Artificial Intelligence**, v. 97, n. 2, p. 273-324, 1997.

KRITAS, S. K. et al. **Impact of mast cells on multiple sclerosis: inhibitory effect of natalizumab**. Int J Immunopathol Pharmacol**,** v. 27, n. 3, p. 331-5, 2014 Jul-Sep 2014. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/25280024>. Acesso em 15 Mai. 2015.

KUCHARCZYK, W. et al. Relaxivity and magnetization transfer of white matter lipids at MR imaging: importance of cerebrosides and pH**. Radiology**, v. 192, n. 2, p. 521-529, 1994.

LAO, Z. et al. Computer-assisted segmentation of white matter lesions in 3D MR images using support vector machine. **Academic radiology**, v. 15, n. 3, p. 300-313, 2008.

LASSMANN, H. The pathology of multiple sclerosis and its evolution. **Philos Trans R Soc Lond B Biol Sci**, v. 354, n. 1390, p. 1635-1640, 1999.

LI, M. et al. Discriminative analysis of multivariate features from structural MRI and diffusion tensor images. **Magnetic Resonance Imaging**, v. 32, n. 8, p. 1043-1051, 2014.

LIU, J.; BUISMAN-PIJLMAN, F.; HUTCHINSON, M. R. **Toll-like receptor 4: innate immune regulator of neuroimmune and neuroendocrine interactions in stress and major depressive disorder**. Front Neurosci**,** v. 8, p. 309, 2014. ISSN 1662-4548. Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/25324715>. Acesso em: 20 Abr. 2015.

LUCCHINETTI, C. F. Multiple sclerosis and the spectrum of CNS inflammatory demyelinating diseases. **Semin Neurol**, v. 28, n. 1, p. 3-6, 2008.

NEEMA, M. et al. 3T MRI relaxometry detects T2 prolongation in the cerebral normal-appearing white matter in multiple sclerosis. **NeuroImage**, v. 46, n. 3, p. 633-641, 2009.

OLEK, M. J. **Epidemiology, risk factors and clinical features of multiple sclerosis in adults**. Wolters Kluwer. Disponível em: http://www.uptodate.com/contents/epidemiology-and-clinical-features-of-multiple-sclerosis-inadults. Acesso em: 20 Ago. 2015.

OLIVEIRA JUNIOR, P. P. M. **Medida da espessura cortical com imagens de ressonância magnética: técnica e aplicações na doença de Alzheimer e na polimicrogiria**. 2008. 101 p. Tese (Doutorado em Ciências) – Faculdade de Medicina de São Paulo, Universidade de São Paulo, São Paulo, 2008.

PESSINI, R. A. **Arquitetura de Software para a Implementação de Arquivo Digital de Pós-Processamento de Imagens Médicas.** 2012. 95 p. [Monografia] (Especialista em Banco de Dados) – Centro de Ciências Exatas, Naturais e Tecnológicas, Universidade de Ribeirão Preto, Ribeirão Preto, 2012.

PESSOTI, H. C. **Desenvolvimento de um framework para classificação de doenças pulmonares difusas apoiado por técnicas de recuperação de imagens baseada em conteúdo e recuperação de informação textual** [Monografia]. Ribeirão Preto: Universidade de São Paulo, 2008.

QUDDUS, A. Adaboost and Support Vector Machines for white matter lesion segmentation in MR images. **Annual international conference of the ieee engineering in medicine and biology society**, n. 1, p. 463-466, 2005.

RORIZ, S. T. S. **Alterações neuropsicológicas em pacientes na fase inicial da esclerose múltipla**. 2010. 118 p. Dissertação (Mestrado em Neurologia/Neurociências) – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2010.

SAILER, M. et al. Focal thinning of the cerebral cortex in multiple sclerosis. **Brain**, v. 126, n. 8, p. 1734-1744, 2003.

SANTOS, A. C. **Quantificação das mudanças no encéfalo causadas pela Esclerose Múltipla usando Ressonância Magnética Quantitativa**. 2007. 133 p. Tese (Livre Docência em Medicina) - Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2007.

SOUZA, D. G. B. **Perfil HLA-DRB, aspectos clínicos e ressonância magnética nuclear em pacientes brasileiros com esclerose múltipla e neuromielite óptica**. 2010. 110 p. Tese (Doutorado em Neurologia) – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2010.

SCHMANSKY, N. FreeSurferWiki. **Martinos Center for Biomedical Imaging**. 2015. Disponível em: <http://freesurfer.net/fswiki>. Acesso em: 12 Jun. 2015.

SHIMO, H. K. **Aplicação de Filtros de Gabor para auxílio ao reconhecimento de padrões em doenças pulmonares difusas**. 2009. [Monografia]. Ribeirão Preto: Universidade de São Paulo, 2009.

THORSTENSEN, N. et al. Measuring Cortical Thickness From Volumetric Mri Data. **VA Medical Center**, p.1-4, 2006.

TOFTS, P, **Quantitative MRI of the Brain**. England: John Wiley & Sons Ltd, 2003. 657p.

TORNES, L.; CONWAY, B.; SHEREMATA, W. Multiple Sclerosis and the Cerebellum. **Clinical Neurology**, v. 32, n. 4, p. 957-977, 2014.

TRAUGOTT, U.; REINHERZ, E. L.; RAINE, C. S. Multiple sclerosis: distribution of T cell subsets within active chronic lesions. **Science**, v. 219, n. 4582, p. 308-310, 1983.

VOLLMER, T. et al. The natural history of brain volume loss among patients with multiple sclerosis: A systematic literature review and meta-analysis. **Journal of the Neurological Sciences**, v. 357, n. 2, p. 8-18, 2015.

WHEELER, A.; VOINESKOS, A. A review of structural neuroimaging in schizophrenia: from connectivity to connectomics. **Frontiers in human neuroscience**, n. 8**,** p. 653, 2014.

WAXMAN, S. G. Demyelinating diseases-new pathological insights, new therapeutic targets. **N Eng J Med**, v. 338, n. 5, p. 323-325, 1998.

WEINSHENKER, B. C. Epidemiology of multiple sclerosis. **Neurol Clin**, v. 14, n. 2, p. 291–308, 1996.

WEYGANDT, M. et al. MRI Pattern Recognition in Multiple Sclerosis Normal-Appearing Brain Areas. **Plos One**, v. 6, n. 6, 2011.

YAFFE, R. et al. Physiology of functional and effective networks in epilepsy. **Clinical Neurophysiology**, v. 126, n. 2, p. 227-236, 2014.

APÊNDICES

Parâmetros de Classificações no WEKA

APÊNDICE A – Parâmetros de Classificação de Atributos KNN

Classificador	KNN1, KNN3 e KNN5
crossValidate	*False*
debug	*False*
distanceWeighting	*No distance weighting*
meanSquared	*False*
nearestNeighbourSearchAlgorithm	*EuclideanDistance*
measurePerformance	*False*
skipIdentical	*False*
windowsize	0
Cross-validation Folds	10

APÊNDICE B – Parâmetros de Classificação de Atributos SVM

Classificador	SVM
SVMType	*C-SVC (classification)*
cacheSize	40
coef0	0
cost	1
debug	*False*
degree	*False*
doNotReplaceMissingValues	*False*
eps	0,001
gama	0
kernelType	*linear: u'*v*
loss	0,1
normalize	*False*
nu	0,5
seed	1
shrinking	*True*
weights	Em branco

APÊNDICE C – Parâmetros de Seleção de Atributos KNN

Classificador	KNN1, KNN3 e KNN5
Attribute Evaluator	*WrapperSubsetEval*
Folds	5 (TI) e 10 (VC)
Seed	1
Threshold	0,01
Search Method	*Best First D1 - N5*
Attribute Selection Mode	*Use full training set (TI) e Cross-validation Folds 10 (VC)*

Relatórios de Melhores Classificações pelo WEKA

APÊNDICE D - Global KNN3

```
=== Run information ===

Scheme:weka.classifiers.lazy.IBk -K 1 -W 0 -A
"weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-
last\""
Relation:     NRM-EMP-weka.filters.unsupervised.attribute.Remove-R1-249,251-254,256-
284,286-433,435-557,560-663
Instances:    117
Attributes:   7
              ECT_parsorbitalis
              ECT_lh-precentral
              ECT_rh-supramarginal
              RELAX_Corpo_Caloso
              VOL_Right-VentralDC
              VOL_WM-hypointensities
              class
Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         116               99.1453 %
Incorrectly Classified Instances         1                0.8547 %
Kappa statistic                          0.9828
Mean absolute error                      0.0177
Root mean squared error                  0.0921
Relative absolute error                  3.571  %
Root relative squared error             18.4837 %
Total Number of Instances              117

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 1         0.016      0.981     1         0.991      0.987    nrm
                 0.984     0          1         0.984     0.992      0.987    emp
Weighted Avg.    0.991     0.007      0.992     0.991     0.991      0.987

=== Confusion Matrix ===

  a  b   <-- classified as
 53  0 |  a = nrm
  1 63 |  b = emp
```

APÊNDICE E - MTR KNN5

```
=== Run information ===

Scheme:weka.classifiers.lazy.IBk -K 5 -W 0 -A
"weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-
last\""
Relation:     MTR_NRM-EMP-weka.filters.unsupervised.attribute.Remove-R1-3,5-9,12-
21,23-34,36-53,56-69,71-77,79-80,82,86,89,91,93,95-97,99,101-105,107-108,111-123,125
-127
Instances:    200
Attributes:   25
              Left-Cerebellum-White-Matter
              Brain-Stem
              Left-Hippocampus
              Right-VentralDC
              lh-lingual
              temporalpole
              rh-caudalmiddlefrontal
              rh-precentral
              wm_lh_caudalmiddlefrontal
              wm_lh_fusiform
              wm_lh_inferiortemporal
              wm_lh_isthmuscingulate
              wm_lh_lateraloccipital
              wm_lh_lingual
              wm_lh_medialorbitofrontal
              wm_parahippocampal
              wm_lh_parsopercularis
              wm_lh_pericalcarine
              wm_lh_precuneus
              wm_lh_rostralmiddlefrontal
              wm_rh_fusiform
              wm_rh_isthmuscingulate
              wm_rh_lateraloccipital
              wm_rh_superiorfrontal
              class
Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 5 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         176               88      %
Incorrectly Classified Instances        24               12      %
Kappa statistic                          0.7598
Mean absolute error                      0.1528
Root mean squared error                  0.2771
Relative absolute error                 30.798  %
Root relative squared error             55.6446 %
Total Number of Instances              200

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.912     0.147      0.838     0.912     0.874      0.96     nrm
                 0.853     0.088      0.921     0.853     0.886      0.96     emp
Weighted Avg.    0.88      0.115      0.883     0.88      0.88       0.96

=== Confusion Matrix ===

  a  b   <-- classified as
 83  8 |  a = nrm
 16 93 |  b = emp
```

APÊNDICE F - MTR KNN3

```
=== Run information ===

Scheme:weka.classifiers.lazy.IBk -K 3 -W 0 -A
"weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-
last\""
Relation:     MTR_NRM-EMP-weka.filters.unsupervised.attribute.Remove-R1-3,6-9,12-
15,17-21,23-53,56-76,80,82,85-86,89-92,94-95,97-99,101-105,107,109-113,115-125,127
Instances:    200
Attributes:   24
              Left-Cerebellum-White-Matter
              Left-Cerebellum-Cortex
              Brain-Stem
              Left-Hippocampus
              Right-Cerebellum-White-Matter
              Right-VentralDC
              temporalpole
              rh-caudalmiddlefrontal
              wm_caudalanteriorcingulate
              wm_lh_caudalmiddlefrontal
              wm_corpuscallosum
              wm_lh_fusiform
              wm_lh_inferiortemporal
              wm_lh_isthmuscingulate
              wm_lh_lingual
              wm_lh_medialorbitofrontal
              wm_lh_parstriangularis
              wm_lh_posteriorcingulate
              wm_lh_rostralmiddlefrontal
              wm_rh_fusiform
              wm_rh_inferiortemporal
              wm_rh_middletemporal
              wm_rh_superiortemporal
              class
Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 3 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         175               87.5    %
Incorrectly Classified Instances        25               12.5    %
Kappa statistic                          0.7486
Mean absolute error                      0.1546
Root mean squared error                  0.2998
Relative absolute error                 31.1691 %
Root relative squared error             60.206  %
Total Number of Instances              200

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.879     0.128       0.851    0.879      0.865      0.939    nrm
                 0.872     0.121       0.896    0.872      0.884      0.939    emp
Weighted Avg.    0.875     0.124       0.876    0.875      0.875      0.939

=== Confusion Matrix ===

  a  b   <-- classified as
 80 11 |  a = nrm
 14 95 |  b = emp
```

APÊNDICE G - Global KNN3

```
=== Run information ===

Scheme:weka.classifiers.functions.LibSVM -S 0 -K 0 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0
-C 1.0 -E 0.001 -P 0.1 -seed 1
Relation:     NRM-EMP-weka.filters.unsupervised.attribute.Remove-R1-85,87-249,251-
362,364-558,560-576,578-663
Instances:    117
Attributes:   6
              DME_wm_lh_posteriorcingulate
              ECT_parsorbitalis
              MTR_wm_corpuscallosum
              VOL_WM-hypointensities
              VOL_parsorbitalis
              class
Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

LibSVM wrapper, original code by Yasser EL-Manzalawy (= WLSVM)

Time taken to build model: 0.05 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         106               90.5983 %
Incorrectly Classified Instances        11                9.4017 %
Kappa statistic                          0.8124
Mean absolute error                      0.094
Root mean squared error                  0.3066
Relative absolute error                 18.9575 %
Root relative squared error             61.5612 %
Total Number of Instances              117

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.962     0.141      0.85      0.962     0.903      0.911    nrm
                 0.859     0.038      0.965     0.859     0.909      0.911    emp
Weighted Avg.    0.906     0.084      0.913     0.906     0.906      0.911

=== Confusion Matrix ===

  a  b   <-- classified as
 51  2 |  a = nrm
  9 55 |  b = emp
```

APÊNDICE H - Regiões do Cérebro Analisadas pela Classificação

1	*Left-Cerebral-White-Matter*
2	*Left-Cerebral-Cortex*
3	*Left-Cerebellum-White-Matter*
4	*Left-Cerebellum-Cortex*
5	*Left-Thalamus-Proper*
6	*Left-Caudate*
7	*Left-Putamen*
8	*Pallidum*
9	*Brain-Stem*
10	*Left-Hippocampus*
11	*Amygdala*
12	*Left-VentralDC*
13	*Right-Cerebral-White-Matter*
14	*Right-Cerebral-Cortex*
15	*Right-Cerebellum-White-Matter*
16	*Right-Cerebellum-Cortex*
17	*Right-Thalamus-Proper*
18	*Right-Caudate*
19	*Right-Putamen*
20	*Right-Hippocampus*
21	*Right-VentralDC*
22	*WM-hypointensities*
23	Corpo_Caloso
24	*caudalanteriorcingulate*
25	*lh-caudalmiddlefrontal*
26	*lh-cuneus*
27	*entorhinal*
28	*lh-fusiform*
29	*lh-inferiorparietal*
30	*lh-inferiortemporal*
31	*isthmuscingulate*
32	*lh-lateraloccipital*
33	*lh-lateralorbitofrontal*
34	*lh-lingual*
35	*lh-medialorbitofrontal*
36	*lh-middletemporal*
37	*parahippocampal*
38	*lh-paracentral*

39	*lh-parsopercularis*
40	*parsorbitalis*
41	*lh-parstriangularis*
42	*pericalcarine*
43	*lh-postcentral*
44	*lh-posteriorcingulate*
45	*lh-precentral*
46	*lh-precuneus*
47	*rostralanteriorcingulate*
48	*lh-rostralmiddlefrontal*
49	*lh-superiorfrontal*
50	*lh-superiorparietal*
51	*lh-superiortemporal*
52	*lh-supramarginal*
53	*temporalpole*
54	*rh-caudalmiddlefrontal*
55	*rh-cuneus*
56	*rh-fusiform*
57	*rh-inferiorparietal*
58	*rh-inferiortemporal*
59	*rh-lateraloccipital*
60	*rh-lateralorbitofrontal*
61	*rh-lingual*
62	*rh-medialorbitofrontal*
63	*rh-middletemporal*
64	*rh-paracentral*
65	*rh-parsopercularis*
66	*rh-parstriangularis*
67	*rh-postcentral*
68	*rh-posteriorcingulate*
69	*rh-precentral*
70	*rh-precuneus*
71	*rh-rostralmiddlefrontal*
72	*rh-superiorfrontal*
73	*rh-superiorparietal*
74	*rh-superiortemporal*
75	*rh-supramarginal*
76	*wm_caudalanteriorcingulate*
77	*wm_lh_caudalmiddlefrontal*
78	*wm_corpuscallosum*
79	*wm_cuneus*

80	*wm_lh_fusiform*
81	*wm_lh_inferiorparietal*
82	*wm_lh_inferiortemporal*
83	*wm_lh_isthmuscingulate*
84	*wm_lh_lateraloccipital*
85	*wm_lh_lateralorbitofrontal*
86	*wm_lh_lingual*
87	*wm_lh_medialorbitofrontal*
88	*wm_lh_middletemporal*
89	*wm_parahippocampal*
90	*wm_lh_paracentral*
91	*wm_lh_parsopercularis*
92	*wm_lh_parstriangularis*
93	*wm_lh_pericalcarine*
94	*wm_lh_postcentral*
95	*wm_lh_posteriorcingulate*
96	*wm_lh_precentral*
97	*wm_lh_precuneus*
98	*wm_rostralanteriorcingulate*
99	*wm_lh_rostralmiddlefrontal*
100	*wm_lh_superiorfrontal*
101	*wm_lh_superiorparietal*
102	*wm_lh_superiortemporal*
103	*wm_lh_supramarginal*
104	*wm_rh_caudalmiddlefrontal*
105	*wm_rh_fusiform*
106	*wm_rh_inferiorparietal*
107	*wm_rh_inferiortemporal*
108	*wm_rh_isthmuscingulate*
109	*wm_rh_lateraloccipital*
110	*wm_rh_lateralorbitofrontal*
111	*wm_rh_lingual*
112	*wm_rh_medialorbitofrontal*
113	*wm_rh_middletemporal*
114	*wm_rh_paracentral*
115	*wm_rh_parsopercularis*
116	*wm_rh_parstriangularis*
117	*wm_rh_pericalcarine*
118	*wm_rh_postcentral*
119	*wm_rh_posteriorcingulate*
120	*wm_rh_precentral*

121	*wm_rh_precuneus*
122	*wm_rh_rostralmiddlefrontal*
123	*wm_rh_superiorfrontal*
124	*wm_rh_superiorparietal*
125	*wm_rh_superiortemporal*
126	*wm_rh_supramarginal*

Ranking **Melhores Atributos**

APÊNDICE I - Anisotropia Fracionada

POS.	ATRIBUTO	QTD. TA	QTD. VC	TOTAL
1º	*Left-Cerebellum-White-Matter*	3	3	6
1º	*Right-Cerebellum-White-Matter*	3	3	6
2º	*wm_lh_precuneus*	3	2	5
2º	*Left-Cerebral-White-Matter*	2	3	5
2º	*wm_corpuscallosum*	2	3	5
2º	*WM-hypointensities*	2	3	5
2º	*wm_lh_pericalcarine*	2	3	5
3º	*Right-Cerebral-White-Matter*	1	3	4
3º	*wm_caudalanteriorcingulate*	2	2	4
3º	*wm_lh_fusiform*	1	3	4
3º	*wm_rh_posteriorcingulate*	2	2	4
3º	*wm_rostralanteriorcingulate*	2	2	4
4º	*wm_lh_lingual*	0	3	3
4º	*wm_lh_medialorbitofrontal*	2	1	3
4º	*wm_lh_isthmuscingulate*	1	2	3
4º	*wm_lh_supramarginal*	1	2	3
5º	*wm_lh_superiortemporal*	2	0	2
5º	Corpo_Caloso	0	2	2
5º	*wm_rh_middletemporal*	0	2	2
5º	*wm_lh_superiorfrontal*	1	1	2
5º	*wm_lh_superiorparietal*	1	1	2
5º	*wm_rh_pericalcarine*	1	1	2
6º	*wm_rh_superiortemporal*	1	1	2
6º	*wm_lh_inferiortemporal*	1	0	1
6º	*wm_rh_inferiortemporal*	1	0	1
6º	*wm_rh_precuneus*	1	0	1
6º	*wm_rh_superiorfrontal*	1	0	1
6º	*wm_lh_posteriorcingulate*	0	1	1
6º	*wm_rh_isthmuscingulate*	0	1	1
6º	*wm_rh_precentral*	0	1	1

APÊNDICE J - Difusibilidade Média

POS.	ATRIBUTO	QTD. TA	QTD. VC	TOTAL
1º	*wm_lh_precuneus*	2	3	5
1º	*wm_rh_isthmuscingulate*	2	3	5
2º	*Left-Cerebellum-White-Matter*	1	3	4
2º	*wm_corpuscallosum*	1	3	4
3º	*wm_lh_fusiform*	2	1	3
3º	*WM-hypointensities*	2	1	3
3º	*Left-Cerebral-White-Matter*	1	2	3
3º	*wm_rostralanteriorcingulate*	1	2	3
4º	*wm_rh_rostralmiddlefrontal*	2	0	2
4º	*wm_lh_caudalmiddlefrontal*	0	2	2
4º	*wm_lh_parstriangularis*	1	1	2
4º	*wm_rh_superiorfrontal*	1	1	2
4º	*wm_lh_inferiortemporal*	0	2	2
4º	*wm_lh_paracentral*	0	2	2
4º	*wm_rh_inferiortemporal*	0	2	2
5º	*Brain-Stem*	0	1	1
5º	*wm_lh_superiorfrontal*	0	1	1
5º	*wm_rh_paracentral*	0	1	1
5º	*wm_rh_posteriorcingulate*	0	1	1

APÊNDICE K – Difusibilidade Paralela

POS.	ATRIBUTO	QTD. TA	QTD. VC	TOTAL
1º	*wm_corpuscallosum*	3	3	6
2º	*wm_lh_medialorbitofrontal*	2	3	5
2º	*wm_rh_caudalmiddlefrontal*	2	3	5
3º	*wm_rostralanteriorcingulate*	1	3	4
4º	*wm_lh_inferiortemporal*	0	3	3
4º	*Left-Cerebellum-White-Matter*	2	1	3
4º	*wm_lh_paracentral*	2	1	3
4º	*wm_lh_lateraloccipital*	1	2	3
4º	*wm_lh_lateralorbitofrontal*	1	2	3
5º	*wm_rh_isthmuscingulate*	2	0	2
5º	*wm_lh_pericalcarine*	0	2	2
5º	*wm_parahippocampal*	0	2	2
5º	*wm_rh_rostralmiddlefrontal*	0	2	2
5º	*Brain-Stem*	1	1	2
6º	Corpo_Caloso	1	0	1
6º	*wm_caudalanteriorcingulate*	1	0	1
6º	*wm_lh_inferiorparietal*	1	0	1

6º	*wm_lh_middletemporal*	1	0	1
6º	*wm_lh_posteriorcingulate*	1	0	1
6º	*wm_lh_precentral*	1	0	1
6º	*wm_lh_superiorfrontal*	1	0	1
6º	*wm_lh_superiortemporal*	1	0	1
6º	*wm_rh_inferiortemporal*	1	0	1
6º	*wm_rh_medialorbitofrontal*	1	0	1
6º	*wm_rh_superiorfrontal*	1	0	1
6º	*wm_rh_superiortemporal*	1	0	1
6º	*Left-Cerebral-White-Matter*	0	1	1
6º	*wm_lh_isthmuscingulate*	0	1	1
6º	*wm_rh_lateraloccipital*	0	1	1
6º	*wm_rh_middletemporal*	0	1	1
6º	*wm_rh_parstriangularis*	0	1	1

APÊNDICE L – Difusibilidade Perpendicular

POS.	ATRIBUTO	QTD. TA	QTD. VC	TOTAL
1º	*wm_lh_isthmuscingulate*	3	3	6
2º	*wm_lh_inferiortemporal*	2	2	4
3º	*wm_rh_inferiortemporal*	1	2	3
3º	*wm_lh_fusiform*	0	2	2
3º	*wm_lh_lateralorbitofrontal*	0	2	2
3º	*wm_lh_posteriorcingulate*	0	2	2
3º	*wm_corpuscallosum*	1	1	2
3º	*wm_lh_medialorbitofrontal*	1	1	2
3º	*wm_lh_precuneus*	1	1	2
3º	*wm_rh_parsopercularis*	1	1	2
3º	*wm_rostralanteriorcingulate*	1	1	2
4º	*Brain-Stem*	1	0	1
4º	*Right-Cerebral-White-Matter*	1	0	1
4º	*wm_lh_inferiorparietal*	1	0	1
4º	*wm_lh_middletemporal*	1	0	1
4º	*wm_lh_superiorfrontal*	1	0	1
4º	*wm_lh_superiortemporal*	1	0	1
4º	*wm_rh_caudalmiddlefrontal*	1	0	1
4º	*wm_rh_inferiorparietal*	1	0	1
4º	*wm_rh_isthmuscingulate*	1	0	1
4º	*wm_rh_middletemporal*	1	0	1
4º	*wm_rh_parstriangularis*	0	1	1
4º	*wm_rh_posteriorcingulate*	0	1	1
4º	*wm_rh_posteriorcingulate*	0	1	1
4º	*wm_rh_superiortemporal*	0	1	1
4º	*wm_rh_superiortemporal*	0	1	1
4º	Corpo_Caloso	0	1	1

4º	Idade	0	1	1
4º	*Right-Cerebellum-White-Matter*	0	1	1
4º	*wm_lh_lateraloccipital*	0	1	1
4º	*wm_rh_lateraloccipital*	0	1	1
4º	*wm_rh_superiorfrontal*	0	1	1

APÊNDICE M - Espessura Cortical

POS.	ATRIBUTO	QTD. TA	QTD. VC	TOTAL
1º	*lh-fusiform*	3	3	6
1º	*lh-lingual*	3	3	6
1º	*pericalcarine*	3	3	6
2º	Idade	2	3	5
2º	*lh-precentral*	2	3	5
3º	*lh-parstriangularis*	1	3	4
3º	*rh-lingual*	1	3	4
3º	*entorhinal*	2	2	4
3º	*lh-lateralorbitofrontal*	2	2	4
3º	*lh-medialorbitofrontal*	2	2	4
3º	*lh-precuneus*	2	2	4
3º	*lh-superiorfrontal*	2	2	4
3º	*rh-superiortemporal*	2	2	4
4º	*rostralanteriorcingulate*	3	0	3
4º	*parsorbitalis*	2	1	3
4º	*rh-inferiortemporal*	2	1	3
4º	*rh-precentral*	1	2	3
4º	*rh-parstriangularis*	3	0	3
4º	*rostralanteriorcingulate*	3	0	3
4º	*temporalpole*	2	1	3
4º	*lh-middletemporal*	1	2	3
4º	*lh-postcentral*	1	2	3
5º	*lh-cuneus*	0	2	2
5º	*lh-rostralmiddlefrontal*	2	0	2
5º	*isthmuscingulate*	1	1	2
5º	*lh-inferiortemporal*	1	1	2
5º	*parahippocampal*	1	1	2
5º	*rh-fusiform*	1	1	2
5º	*rh-lateraloccipital*	1	1	2
5º	*rh-middletemporal*	0	2	2
5º	*rh-postcentral*	0	2	2
6º	*rh-caudalmiddlefrontal*	1	0	1
6º	*rh-medialorbitofrontal*	1	0	1
6º	*rh-parsopercularis*	1	0	1
6º	*rh-rostralmiddlefrontal*	1	0	1

6º	*rh-superiorfrontal*	1	0	1
6º	*rh-supramarginal*	1	0	1
6º	*lh-inferiorparietal*	0	1	1
6º	*lh-superiortemporal*	0	1	1
6º	*rh-paracentral*	0	1	1
6º	*rh-precuneus*	0	1	1

APÊNDICE N - Transferência de Magnetização

POS.	ATRIBUTO	QTD. TA	QTD. VC	TOTAL
1º	*Left-Hippocampus*	3	3	6
1º	*wm_lh_fusiform*	3	3	6
2º	*wm_lh_caudalmiddlefrontal*	3	2	5
3º	*Brain-Stem*	2	2	4
3º	*wm_parahippocampal*	2	2	4
4º	*wm_rh_fusiform*	3	0	3
4º	*Right-VentralDC*	0	3	3
4º	*temporalpole*	0	3	3
5º	*rh-caudalmiddlefrontal*	2	0	2
5º	*Left-Cerebellum-Cortex*	1	1	2
5º	*wm_rh_isthmuscingulate*	1	1	2
5º	*wm_rh_lateraloccipital*	1	1	2
5º	*wm_rh_superiortemporal*	1	1	2
6º	*rh-precentral*	1	0	1
6º	*Right-Cerebral-White-Matter*	1	0	1
6º	*Right-VentralDC*	1	0	1
6º	*wm_caudalanteriorcingulate*	1	0	1
6º	*wm_lh_inferiortemporal*	1	0	1
6º	*wm_lh_isthmuscingulate*	1	0	1
6º	*wm_lh_lateralorbitofrontal*	1	0	1
6º	*wm_lh_lingual*	1	0	1
6º	*wm_lh_parsopercularis*	1	0	1
6º	*wm_lh_pericalcarine*	1	0	1
6º	*wm_rh_superiorfrontal*	1	0	1
6º	*lh-fusiform*	0	1	1
6º	*lh-lingual*	0	1	1
6º	*rh-inferiortemporal*	0	1	1
6º	*rh-posteriorcingulate*	0	1	1
6º	*Right-Cerebral-Cortex*	0	1	1
6º	*Right-Hippocampus*	0	1	1
6º	*wm_corpuscallosum*	0	1	1
6º	*wm_lh_lateraloccipital*	0	1	1
6º	*wm_lh_lateralorbitofrontal*	0	1	1
6º	*wm_lh_medialorbitofrontal*	0	1	1
6º	*wm_lh_parstriangularis*	0	1	1

6º	*wm_lh_pericalcarine*	0	1	1
6º	*wm_lh_posteriorcingulate*	0	1	1
6º	*wm_lh_rostralmiddlefrontal*	0	1	1
6º	*wm_rh_fusiform*	0	1	1
6º	*wm_rh_inferiortemporal*	0	1	1
6º	*wm_rh_middletemporal*	0	1	1
6º	*wm_rh_posteriorcingulate*	0	1	1

APÊNDICE O - Relaxometria

POS.	ATRIBUTO	QTD. TA	QTD. VC	TOTAL
1º	*wm_lh_fusiform*	3	2	5
1º	*Pallidum*	2	3	5
2º	*wm_corpuscallosum*	2	2	4
3º	*wm_lh_precuneus*	0	3	3
3º	*wm_lh_posteriorcingulate*	2	1	3
3º	Corpo_Caloso	1	2	3
4º	*Left-Hippocampus*	1	1	2
4º	*wm_caudalanteriorcingulate*	1	1	2
4º	*lh-paracentral*	1	1	2
5º	*Left-Cerebellum-Cortex*	1	0	1
5º	*Left-VentralDC*	0	1	1
5º	*wm_cuneus*	0	1	1
5º	*wm_lh_isthmuscingulate*	0	1	1
5º	*wm_lh_posteriorcingulate*	0	1	1
5º	*wm_lh_superiortemporal*	0	1	1
5º	*lh-lateralorbitofrontal*	1	0	1
5º	*lh-superiorparietal*	1	0	1
5º	*pericalcarine*	1	0	1
5º	*rh-paracentral*	1	0	1
5º	*wm_lh_inferiortemporal*	1	0	1
5º	*wm_lh_medialorbitofrontal*	1	0	1
5º	*wm_lh_precuneus*	1	0	1
5º	*wm_rh_parsopercularis*	1	0	1
5º	*wm_rh_posteriorcingulate*	1	0	1
5º	*WM-hypointensities*	1	0	1

APÊNDICE P - Volumetria

POS.	ATRIBUTO	QTD. TA	QTD. VC	TOTAL
1º	*Right-VentralDC*	3	3	6
1º	*WM-hypointensities*	3	3	6
2º	*lh-lateraloccipital*	2	2	4
2º	*parsorbitalis*	1	3	4
3º	*temporalpole*	1	2	3
4º	*lh-inferiortemporal*	0	2	2
4º	*wm_rh_inferiortemporal*	2	0	2
4º	*lh-medialorbitofrontal*	1	1	2
4º	*rh-inferiortemporal*	1	1	2
4º	*wm_lh_lateraloccipital*	1	1	2
5º	*Left-Cerebellum-Cortex*	1	0	1
5º	*parahippocampal*	1	0	1
5º	*parsorbitalis*	1	0	1
5º	*rh-superiortemporal*	1	0	1
5º	*temporalpole*	1	0	1
5º	*wm_lh_lingual*	1	0	1
5º	*wm_lh_medialorbitofrontal*	1	0	1
5º	*entorhinal*	0	1	1
5º	*Left-Cerebellum-White-Matter*	0	1	1
5º	*Right-Cerebellum-White-Matter*	0	1	1
5º	*wm_lh_fusiform*	0	1	1
5º	*wm_rh_inferiortemporal*	0	1	1

APÊNDICE Q - Global

POS.	ATRIBUTO	QTD. TA	QTD. VC	TOTAL
1º	*ECT_parsorbitalis*	3	3	6
1º	*VOL_WM-hypointensities*	3	3	6
2º	*MTR_wm_corpuscallosum*	2	1	3
3º	*VOL_parsorbitalis*	1	1	2
4º	*ECT_lh-precentral*	1	0	1
4º	*ECT_rh-middletemporal*	1	0	1
4º	*ECT_rh-supramarginal*	1	0	1
4º	*RELAX_Corpo_Caloso*	1	0	1
4º	*RELAX_isthmuscingulate*	1	0	1
4º	*RELAX_wm_rh_superiortemporal*	1	0	1
4º	*VOL_Right-VentralDC*	1	0	1
4º	*DME_wm_lh_posteriorcingulate*	0	1	1
4º	*DME_wm_lh_precuneus*	0	1	1

APÊNDICE R - Regiões mais relevantes na classificação principal

Posição	Atributo	Quantidade
1º	*wm_corpuscallosum*	19
2º	*wm_lh_precuneus*	13
3º	*WM-hypointensities*	11
3º	*wm_lh_fusiform*	11
4º	*Left-Cerebellum-White-Matter*	10
5º	*Right-Cerebellum-White-Matter*	6
5º	*wm_lh_isthmuscingulate*	6
5º	*lh-fusiform*	6
5º	*lh-lingual*	6
5º	*pericalcarine*	6
5º	*Left-Hippocampus*	6
5º	*Right-VentralDC*	6
6º	*Left-Cerebral-White-Matter*	5
6º	*wm_lh_pericalcarine*	5
6º	*wm_rh_isthmuscingulate*	5
6º	*wm_lh_medialorbitofrontal*	5
6º	*wm_rh_caudalmiddlefrontal*	5
6º	Idade	5
6º	*lh-precentral*	5
6º	*wm_lh_caudalmiddlefrontal*	5
7º	*wm_rostralanteriorcingulate*	4
7º	*wm_lh_inferiortemporal*	4
7º	*Brain-Stem*	4
7º	*wm_parahippocampal*	4
7º	*lh-lateraloccipital*	4
8º	*wm_rh_inferiortemporal*	3
8º	*Pallidum*	3
9º	*wm_lh_posteriorcingulate*	2

APÊNDICE S - Regiões mais relevantes na classificação restrita

Posição	Atributo	Quantidade
1º	*wm_corpuscallosum*	9
1º	*WM-hypointensities*	6
2º	*wm_lh_inferiortemporal*	4
3º	*Brain-Stem*	3
3º	*Left-Cerebellum-White-Matter*	3
3º	*lh-inferiortemporal*	3
3º	*Right-Cerebellum-White-Matter*	3
3º	*wm_rh_caudalmiddlefrontal*	3
3º	*wm_rh_inferiortemporal*	3
3º	*wm_rh_middletemporal*	3

APÊNDICE T - Área sob a curva ROC (%) usando KNN1

VETOR	TA KNN (%)	TA SVM (%)	MA KNN (%)	MA SVM (KNN) (%)
ANF	80,0	78,0	83,7	78,3
DME	74,5	82,2	84,7	81,1
DPA	73,5	82,6	78,5	82,4
DPE	74,5	81,0	77,9	77,3
ECT	83,1	82,6	84,2	80,0
MTR	81,0	87,9	92,6	88,4
REL	80,5	78,8	86,0	86,0
VOL	78,4	90,3	87,2	87,8
GLO	79,7	82,0	98,7	85,6

APÊNDICE U - Área sob a curva ROC (%) usando KNN3

VETOR	TA KNN (%)	TA SVM (%)	MA KNN(%)	MA SVM (KNN) (%)
ANF	88,7	78,0	92,4	81,2
DME	82,7	82,2	86,7	80,7
DPA	82,4	82,6	87,4	80,8
DPE	80,2	81,0	85,3	81,9
ECT	85,6	82,6	87,6	80,6
MTR	86,0	87,9	93,9	89,0
REL	82,3	78,8	92,3	86,0
VOL	90,2	90,3	93,6	87,9
GLO	89,9	82,0	98,0	91,1

APÊNDICE V - Área sob a curva ROC (%) usando KNN5

VETOR	TA KNN (%)	TA SVM (%)	MA KNN (%)	MA SVM (KNN) (%)
ANF	89,7	78,0	89,1	78,4
DME	86,0	82,2	89,9	80,3
DPA	84,2	82,6	89,0	85,4
DPE	82,2	81,0	87,3	79,4
ECT	87,4	82,6	90,7	83,1
MTR	90,3	87,9	96,0	86,2
REL	83,6	78,8	92,0	88,4
VOL	90,5	90,3	93,6	87,4
GLO	91,8	82,0	97,1	89,0

APÊNDICE W - Porcentagem das Melhores Classificações por Vetor com Média

VETOR	KNN1 (%)	KNN3 (%)	KNN5 (%)	SVM (%)
ANF	83,7	92,4	89,7	81,2
DME	84,7	86,7	89,9	82,2
DPA	82,6	87,4	89,0	85,4
DPE	81,0	85,3	87,3	81,9
ECT	84,2	87,6	90,7	83,1
MTR	92,6	93,9	96,0	89,0
REL	86,0	92,3	92,0	88,4
VOL	90,3	93,6	93,6	90,3
GLO	98,7	98,0	97,1	91,1
MÉDIA	**87,1**	**90,8**	**91,7**	**85,8**

ANEXOS

ANEXO 1 - Ofício do Comitê de Ética

HOSPITAL DAS CLÍNICAS DA FACULDADE DE MEDICINA
DE RIBEIRÃO PRETO DA UNIVERSIDADE DE SÃO PAULO

FMRP-USP
RIBEIRÃO PRETO

Ribeirão Preto, 21 de outubro de 2015.

Ofício nº 3778/2015
CEP/LVG

PROCESSO HCRP nº 14124/2015

Prezados Pesquisadores,

O trabalho intitulado **"APLICAÇÃO DE TÉCNICAS DE RECONHECIMENTO DE PADRÕES EM DADOS QUANTITATIVOS DE NEUROIMAGENS POR RESSONÂNCIA MAGNÉTICA",** foi analisado pelo Comitê de Ética em Pesquisa, em sua 416ª Reunião Ordinária realizada em 19/10/2015, e enquadrado na categoria: **APROVADO,** bem como a solicitação de dispensa de aplicação do Termo de Consentimento Livre e Esclarecido.

De acordo com Carta Circular nº 003/2011/CONEP/CNS, datada de 21/03/2011, o sujeito de pesquisa ou seu representante, quando for o caso, deverá rubricar todas as folhas do Termo de Consentimento Livre e Esclarecido – TCLE – apondo sua assinatura na última do referido Termo; o pesquisador responsável deverá da mesma forma, rubricar todas as folhas do Termo de Consentimento Livre e Esclarecido – TCLE – apondo sua assinatura na última página do referido Termo.

Este Comitê segue integralmente a Conferência Internacional de Harmonização de Boas Práticas Clínicas (IGH-GCP), bem como a Resolução nº 466/2012 CNS/MS.

Lembramos que devem ser apresentados a este CEP, o Relatório Parcial e o Relatório Final da pesquisa.

Atenciosamente.

PROF. DR. LUÍS VICENTE GARCIA
Vice-Coordenador do Comitê de Ética em
Pesquisa do HCRP e da FMRP-USP

Ilustríssimos Senhores
RODRIGO ANTONIO PESSINI
PROF. DR. CARLOS ERNESTO GARRIDO SALMON (Orientador)
Depto. de Clínica Médica – Centro de Ciências das Imagens e Física Médica

HOSPITAL DAS CLÍNICAS DA FACULDADE DE MEDICINA DE RIBEIRÃO PRETO DA UNIVERSIDADE DE SÃO PAULO
Campus Universitário – Monte Alegre
14048-900 Ribeirão Preto SP

Comitê de Ética em Pesquisa do HCRP e FMRP-USP
FWA-00002733; IRB-00002186 e
Registro Plataforma Brasil /CONEP nº 5440
(016) 3602-2228
cep@hcrp.usp.br

www.hcrp.usp.br

ANEXO 2 - Ofício do Comitê de Ética (versão 2)

HOSPITAL DAS CLÍNICAS DA FACULDADE DE MEDICINA
DE RIBEIRÃO PRETO DA UNIVERSIDADE DE SÃO PAULO

FMRP-USP
RIBEIRÃO PRETO

Ribeirão Preto, 03 de fevereiro de 2016.

Ofício nº 429/2016
CEP/MGV

PROCESSO HCRP Nº 14124/2015

Prezados Pesquisadores,

O Comitê de Ética em Pesquisa, em sua 420ª Reunião Ordinária, realizada em 01/02/2015, analisou e aprovou o Projeto de Pesquisa Versão 2 – 11/01/2016 e a carta de dispensa do TCLE Versão 2 – 11/01/2016, bem como a alteração do título da pesquisa para **"APLICAÇÃO DE TÉCNICAS DE RECONHECIMENTO DE PADRÕES EM DADOS QUANTITATIVOS DE NEUROIMAGENS POR RESSONÂNCIA MAGNÉTICA EM PACIENTES DE ESCLEROSE MÚLTIPLA".** O CEP também tomou ciência do Relatório de andamento de 11/01/2016.

Atenciosamente,

DRª MARCIA GUIMARÃES VILLANOVA
Coordenadora do Comitê de Ética em
Pesquisa do HCRP e da FMRP-USP

Ilustríssimos Senhores
RODRIGO ANTONIO PESSINI
PROF. DR. CARLOS ERNESTO GARRIDO SALMON (Orientador)
Depto. de Clínica Médica – Centro de Ciências das Imagens e Física Médica

Campus Universitário – Monte Alegre
14048-900 Ribeirão Preto SP

Comitê de Ética em Pesquisa do HCRP e FMRP-USP
FWA-00002733; IRB-00002186 e
Registro Plataforma Brasil /CONEP nº 5440
(016) 3602-2228
cep@hcrp.usp.br

www.hcrp.usp.br