



Universidade Estadual de Campinas

FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO

Mariana Pinheiro Bento

ANÁLISE DE LESÕES NA SUBSTÂNCIA BRANCA DO CÉREBRO  
A PARTIR DE IMAGENS DE RESSONÂNCIA MAGNÉTICA

Campinas  
2013



Universidade Estadual de Campinas

FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO

ANÁLISE DE LESÕES NA SUBSTÂNCIA BRANCA DO CÉREBRO  
A PARTIR DE IMAGENS DE RESSONÂNCIA MAGNÉTICA

*Mariana Pinheiro Bento*

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas para obtenção do título de Mestra em Engenharia Elétrica.  
Área de concentração: Engenharia de Computação.

Orientador: Prof. Dr. Roberto de Alencar Lotufo

Co-orientadora: Dra. Letícia Rittner

Este exemplar corresponde à versão final da tese defendida pela aluna Mariana Pinheiro Bento, e orientada pelo Prof. Dr. Roberto de Alencar Lotufo

---

Campinas  
2013

Ficha catalográfica  
Universidade Estadual de Campinas  
Biblioteca da Área de Engenharia e Arquitetura  
Rose Meire da Silva - CRB 8/5974

B446a Bento, Mariana Pinheiro, 1988-  
Análise de lesões na substância branca do cérebro a partir de imagens de ressonância magnética / Mariana Pinheiro Bento. – Campinas, SP : [s.n.], 2013.

Orientador: Roberto de Alencar Lotufo.  
Coorientador: Letícia Rittner.  
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Imagens de ressonância magnética. 2. Processamento de imagens. 3. Diagnostico por imagem. I. Alencar Lotufo, Roberto de, 1955-. II. Rittner, Letícia. III. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. IV. Título.

Informações para Biblioteca Digital

**Título em inglês:** Analysis of white matter lesions using magnetic resonance images

**Palavras-chave em inglês:**

Magnetic resonance image

Image processing

Imaging diagnostic

**Área de concentração:** Engenharia de Computação

**Titulação:** Mestra em Engenharia Elétrica

**Banca examinadora:**

Roberto de Alencar Lotufo [Orientador]

Jane Maryam Rondina

Romis Ribeiro de Faissol Attux

**Data de defesa:** 20-02-2013

**Programa de Pós-Graduação:** Engenharia Elétrica

## COMISSÃO JULGADORA - TESE DE MESTRADO

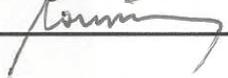
**Candidata:** Mariana Pinheiro Bento

**Data da Defesa:** 20 de fevereiro de 2013

**Título da Tese:** "Análise de Lesões na Substância Branca do Cérebro a partir de Imagens de Ressonância Magnética"

Prof. Dr. Roberto de Alencar Lotufo (Presidente):  \_\_\_\_\_

Profa. Dra. Jane Maryam Rondina:  \_\_\_\_\_

Prof. Dr. Romis Ribeiro de Faissol Attux:  \_\_\_\_\_

# Agradecimentos

Agradeço,

em primeiro lugar, à Deus, pela força e coragem durante toda esta longa caminhada.

a todos os professores que me acompanharam durante o mestrado, em especial ao Prof. Lotufo, por todas as orientações e o acompanhamentos determinantes para a execução deste trabalho.

à minha querida co-orientadora e amiga Dra. Letícia Rittner pela participação tão ativa e importante na minha orientação, além de todo o companheirismo que facilitaram a minha estadia em Campinas.

ao meu noivo Smaylle pela paciência, tranquilidade, força. Sem você ao meu lado nada disso seria possível.

à minha família, pela o incentivo e carinho, em especial meus pais Luís e Graça, irmã Raquel e minhas amadas priminhas Júlia e Letícia.

aos meus tios e primos que me auxiliaram com a difícil mudança para a cidade de Campinas, em especial a tia Célia e Canta.

aos colegas de trabalho mais próximos Roberto, Suelen e Wallace a convivência diária, almoços no RA e cafés na FEA.

aos demais colegas do Departamento de Computação e Automação a ótima convivência.

aos membros da banca examinadora os comentários, sugestões e contribuições, que ajudaram a melhorar a qualidade e a redação final do manuscrito.

à agência CAPES o apoio financeiro concedido durante todo o período de mestrado.

à FEEC/UNICAMP a ótima estrutura que oferece aos estudantes e pesquisadores.

a todos que de alguma forma contribuíram com o meu progresso como aluna e como Ser.

# Resumo

Essa dissertação de mestrado propõe um método de análise de lesões na substância branca do cérebro afim de distinguir as regiões de interesse entre substância branca normal ou não-normal, tarefa denominada identificação de lesões, assim como distinguir diferentes tipos de lesões de acordo com sua etiologia: desmielinizante ou isquêmica, tarefa denominada classificação de lesões. O método combina a análise de textura com o uso de classificadores, como o Máquinas de vetores suporte (SVM), K-vizinhos mais próximos (kNN), Floresta de Caminhos Ótimos (OPF) e Análise Discriminante Linear (LDA). Experimentos realizados em dados reais de ressonância magnética do cérebro mostraram que o método proposto é adequado para identificação e classificação de lesões no cérebro.

Palavras-chave: Imagens de Ressonância Magnética; Hiperintensidades na substância branca; Textura; Classificadores

# Abstract

This dissertation proposes a method for brain white matter lesions analysis in order to distinguish regions of interest between normal and non-normal brain white matter, called lesion identification task, and also to distinguish different types of lesions based on their etiology: demyelinating or ischemic, called lesion classification task. The method combines texture analysis with the use of classifiers such as Support Vector Machine (SVM), k-Nearest Neighbor (kNN), Optimum Path Forest (OPF) and Linear Discriminant Analysis (LDA). Experiments using real brain MRI data have shown that the proposed method is suitable to identify and classify the brain lesions.

Key-words: MRI; White Matter Hyperintensity; Texture; Classifiers

# Lista de Figuras

1.1	Fluxograma de um sistema de reconhecimento de padrões . . . . .	8
1.2	Distâncias e orientações utilizadas no cálculo da matriz de co-ocorrência. . . . .	9
1.3	Exemplo de matriz de co-ocorrência . . . . .	10
1.4	Exemplo de matriz de comprimento de corrida . . . . .	11
1.5	Representação das diferentes orientações nas quais a matriz de comprimento de corrida pode ser calculada (Schwartz, 2005). . . . .	12
1.6	Definição da vizinhança do pixel $x(i,j)$ . . . . .	13
1.7	Definição da vizinhança do pixel $s$ da imagem $f$ . . . . .	14
1.8	Decomposição resultante da transformada Wavelet . . . . .	15
1.9	Representação do método de seleção de atributos árvore de decisão com o conjunto de atributos inicial e reduzido (Adaptada de (Han and Kamber, 2006))	16
2.1	Fluxograma da metodologia proposta . . . . .	28
2.2	Exemplo do conjunto de dados e da extração das regiões de interesse . . . . .	29
3.1	Representação decrescente do grau de importância dos atributos de textura na execução da tarefa Id gerado através do método da árvore de decisão . . . . .	37
3.2	Representação decrescente do grau de importância dos atributos de textura na execução da tarefa Cl gerado através do método da árvore de decisão . . . . .	38
3.3	Representação decrescente do grau de importância dos atributos de textura na execução da tarefa IdCl gerado através do método da árvore de decisão. . . . .	38
3.4	Lista com os rótulos e os nomes correspondentes dos atributos de textura utilizados e extraídos nos seguintes grupos: momentos do histograma (M), matriz de co-ocorrência (MC), matriz de comprimento de corrida (CC), gradiente (G), autoregressivos (AR) e wavelet(W) . . . . .	41

# Lista de Tabelas

1.1	Matriz de relação entre dois classificadores: $c_i$ e $c_j$ . . . . .	21
1.2	Representação tabular da matriz de confusão . . . . .	22
3.1	Melhores acurácias obtidas para as tarefas Id, Cl e IdCl . . . . .	31
3.2	Matriz de confusão normalizada da tarefa de identificação de lesões realizada pelo classificador SVM . . . . .	32
3.3	Matriz de confusão normalizada da tarefa de identificação de lesões realizada pelo classificador OPF . . . . .	32
3.4	Matriz de confusão da tarefa de classificação de lesões realizada pelo classificador LDA . . . . .	32
3.5	Matriz de confusão da tarefa de classificação de lesões realizada OPF utilizando distância manhattan . . . . .	33
3.6	Matriz de confusão das tarefas de identificação e classificação de lesões em uma única etapa realizadas pelo classificador SVM . . . . .	33
3.7	Acurácias obtidas para as tarefas de identificação de lesões (tarefa Id), classificação de lesões (tarefa Cl) e identificação e classificação de lesões em uma única etapa (tarefa IdCl) com ou sem a etapa de pré-processamento (normalização das ROIs) . . . . .	35
3.8	Acurácias obtidas para as tarefas Id, Cl e IdCl variando a quantização utilizada para o cálculo da matriz de co-ocorrência e da matriz de comprimento de corrida . . . . .	36
3.9	Acurácias obtidas nas tarefas Id, Cl e IdCl quando variando o método de seleção de atributos e utilizando o melhor método de classificação para cada tarefa . . . . .	36
3.10	Acurácias obtidas na tarefas IdCl quando variando o método de seleção de atributos . . . . .	36
3.11	Acurácias obtidas na tarefa Cl quando realizando a fusão dos classificadores utilizando uma das técnicas variando o método de seleção de atributos e utilizando o melhor método de classificação para cada tarefa . . . . .	39

# Lista de Acrônimos e Notação

RM	Ressonância Magnética
OPF	Optimum Path Forest (floresta de caminhos ótimos)
IFT	Image Foresting Transform (transformada imagem-floresta)
AVC	Acidente Vascular Cerebral
MST	Minimum Spanning Tree (árvore geradora mínima)
SVM	Support Vector Machine (máquinas de vetores suporte)
kNN	k-Nearest Neighbor (k vizinhos mais próximos)
ROI	Region of Interest (região de interesse)
H	Histograma
MC	Gray Level Co-occurrence Matrices (matriz de co-ocorrência de nível de cinza)
CC	Run-Length Matrix (matriz de comprimento de corrida)
W	Wavelet
G	Gradiente
AR	Auto-regressivos

# Sumário

<b>Introdução Geral</b>	<b>1</b>
0.1 Motivação . . . . .	2
0.2 Objetivos . . . . .	2
0.3 Revisão Bibliográfica . . . . .	3
0.4 Apresentação da Dissertação . . . . .	5
<b>1 Definições</b>	<b>6</b>
1.1 Pré-Processamento . . . . .	6
1.2 Reconhecimento de Padrões . . . . .	7
1.2.1 Extração de Atributos . . . . .	7
1.2.2 Seleção de Atributos . . . . .	15
1.2.3 Classificação de Padrões . . . . .	17
<b>2 Metodologia</b>	<b>23</b>
2.1 Aquisição das imagens, pré-processamento e segmentação manual das regiões de interesse . . . . .	23
2.2 Extração de atributos . . . . .	24
2.3 Seleção de atributos . . . . .	25
2.4 Identificação e Classificação de Lesões . . . . .	25
2.4.1 Fusão de Classificadores . . . . .	26
2.5 Ambiente de Desenvolvimento . . . . .	27
<b>3 Experimentos, Resultados e Discussões</b>	<b>30</b>
3.1 A análise de lesões em substância branca . . . . .	30
3.2 Influência das etapas de pré-processamento, extração de atributos e seleção de atributos nos resultados . . . . .	34
3.2.1 Etapa de pré-processamento . . . . .	34
3.2.2 Quantização das matrizes de co-ocorrência e de comprimento de corrida	35
3.2.3 Seleção de atributos . . . . .	36
3.3 Fusão de Classificadores . . . . .	39

3.4	Considerações Finais . . . . .	40
<b>4</b>	<b>Conclusões e Perspectivas</b>	<b>42</b>
	<b>Bibliografia</b>	<b>45</b>

# Introdução Geral

As lesões em substância branca são comuns no cérebro humano e podem ser detectadas em imagens de ressonância magnética (RM) em pacientes assintomáticos e sintomáticos neurológicos. O frequente aparecimento dessas lesões tem possível associação com a atividade de diversas doenças, ou com distúrbios cognitivos, porém este assunto ainda é controverso, dificultando a correlação entre as manifestações clínicas e os achados de neuroimagem (Walecki et al., 2002).

Neste contexto, a ressonância magnética (RM) é o exame mais sensível para a identificação de alterações na substância branca, pois tem um contraste excelente de tecidos moles e capacidade multiplanar de aquisição das imagens, que proporcionam maior qualidade anatômica nas imagens, sendo portanto o melhor exame para se detectar as lesões cerebrais (Appenzeller et al., 2008). As imagens de RM têm sido utilizadas não somente para o diagnóstico, como também para o acompanhamento dos pacientes, possibilitando que diagnósticos neurológicos possam, em alguns casos, ser feitos mesmo na ausência de sintomas clínicos específicos (Fazekas et al., 1999).

A análise das lesões em substância branca pode ser utilizada tanto para diferenciar tecido saudável de tecido lesionado, tarefa denominada identificação de lesão, como para classificar, dentro do conjunto de pacientes, as diferentes etiologias das lesões, tarefa denominada classificação de lesões. As etiologias de natureza isquêmica e desmielinizantes são as mais frequentemente observadas. As lesões de etiologia desmielinizante são causadas pela esclerose múltipla, enquanto as lesões de etiologia isquêmica são causadas pelo acidente vascular cerebral isquêmico (AVC). Os pacientes com esclerose múltipla têm lesão na bainha de mielina (desmielinização), prejudicando a neurotransmissão e resultando em uma resposta anormal do sistema imunológico do indivíduo (Compston and Coles, 2008). Já as lesões de etiologia isquêmica ocorrem quando há obstrução de um vaso que irriga o cérebro (AVC), geralmente de pequeno calibre, denominado microangiopatias. Este processo causa a diminuição ou parada da circulação sanguínea, provocando a rápida degeneração do tecido cerebral (Ginsberg, 2008).

## 0.1 Motivação

A análise das lesões em substância branca do cérebro em imagens de RM é uma tarefa não-trivial, devido a fatores como a diversidade dos dispositivos de aquisição de imagem e a variedade do tamanho, forma e localização das lesões. Os especialistas geralmente consideram informações clínicas adicionais como idade, exames físicos e histórico do paciente para analisar corretamente essas lesões. Dessa forma, ferramentas e metodologias desenvolvidas com o intuito de auxiliar os especialistas nesta tarefa são relevantes no diagnóstico e acompanhamento dos pacientes.

Essa semi-automatização da análise das lesões em substância branca do cérebro diminui a subjetividade do processo, tornando-o mais robusto e ágil, possibilitando quantificar a severidade das lesões através de medidas quantitativas e acompanhar a evolução dessas lesões ao longo do tempo (estudos longitudinais)

Por fim, os trabalhos na literatura que realizam automatização da análise das lesões em substância branca focam na identificação e segmentação de lesões causadas por uma determinada doença. Nestes casos, a maioria das metodologias propostas não são generalistas para tratar de lesões causadas por diferentes doenças. Além disso, estes trabalhos não realizam estudos sobre as diferentes etiologias das lesões, que possibilitariam em trabalhos futuros propor etiologias a lesões ainda não caracterizadas, como as causadas pela doença lúpus eritematoso sistêmico, baseadas nas etiologias conhecidas (Lapa et al., 2011).

## 0.2 Objetivos

O objetivo principal deste trabalho é analisar as lesões em substância branca do cérebro através da combinação de técnicas de processamento de imagens e reconhecimento de padrões. A análise dessas lesões em imagens de RM compreende a tarefa de identificação, que representa a diferenciação entre tecidos lesionados e tecidos saudáveis, e a tarefa de classificação dos diferentes tipos de lesão de acordo com sua etiologia, isquêmica ou desmielinizante.

Além de desenvolver uma metodologia para a solução do problema proposto, também temos como objetivo estudar e compreender a importância de cada uma das etapas dessa abordagem, com o intuito de determinar a sua influência nos resultados finais.

Um outro objetivo do trabalho é desenvolver uma metodologia que solucione o problema buscando atingir requisitos, como pouca ou nenhuma interação com o usuário, tempo de execução não proibitivo e taxas de acertos aceitáveis para o ambiente clínico, afim de possibilitar em trabalhos futuros o desenvolvimento de uma ferramenta para o auxílio no diagnóstico médico. Além disso, o método proposto deve ser robusto e generalista para analisar corretamente lesões de diferentes etiologias, com tamanhos e formatos variáveis.

### 0.3 Revisão Bibliográfica

Os artigos encontrados na literatura, em sua maioria, apresentam discussões clínicas sobre o assunto, relacionando as lesões causadas por esclerose múltipla ou por infarto com informações como idade ou função cognitiva, como em (Hunt et al., 1989) e (Tiehuis et al., 2008), ou ainda detalhando os tratamentos farmacológicos para prevenir o progresso da lesão (Ovbiagele and Saver, 2006). O estudo populacional apresentado em (Walecki et al., 2002), por sua vez, indica uma prevalência maior de distúrbios cognitivos em indivíduos idosos com lesões de substância branca quando comparados aos pacientes sem esse tipo de lesão. Um outro trabalho teórico correlacionado apresentado em (Kim et al., 2008) propõe uma nova subdivisão teórica para as lesões em substância branca com melhor relevância funcional e etiológica, baseado na sua localização.

A automatização da análise das lesões de substância branca em imagens de ressonância magnética, por se tratar de uma tarefa complexa, torna necessária a combinação de técnicas de diferentes áreas, como processamento de imagens e reconhecimento de padrões. Os trabalhos encontrados na literatura que se propõem a realizar análise automática desse tipo de lesão apresentam como principal foco a identificação e segmentação das lesões causadas por uma doença específica, sem realizar estudos sobre a caracterização das lesões de acordo com a etiologia. O trabalho proposto em (Anbeek et al., 2004), por exemplo, apresenta um método probabilístico e automático de segmentação de lesões em substância branca em imagens de RM. O método gera um mapa de probabilidade, contendo as probabilidades dos voxels serem lesões, através do uso do classificador k-vizinhos mais próximos (kNN). As imagens analisadas foram extraídas de 20 pacientes com AVC.

Um outro trabalho apresenta uma técnica para quantificar e localizar as lesões em substância branca em pacientes com depressão tardia em imagens de ressonância magnética, utilizando o algoritmo *fuzzy-connected* (Wu et al., 2006). O método foi avaliado em imagens adquiridas de 11 pacientes com depressão tardia e 8 controles. Em um outro estudo, Zimring introduz um algoritmo de detecção automática de lesões desmielinizantes em imagens de ressonância magnética (Zimring et al., 1998). O método foi aplicado em 45 imagens adquiridas de 14 pacientes com esclerose múltipla.

Existe ainda um trabalho na literatura que se propõe a comparar diferentes métodos automáticos de detecção de lesões na substância branca em imagens de ressonância magnética (Klöppel et al., 2011). O estudo visa comparar métodos de detecção supervisionados que identificam as lesões a partir de exemplos manuais com abordagens não-supervisionadas. O conjunto de dados apresenta imagens de ressonância magnética FLAIR e ponderadas em T1 de um conjunto de 20 pacientes com demências e déficit cognitivo moderado. O trabalho mostra como resultado um estudo crítico das vantagens e desvantagens de cada método, baseando-se nos respectivos tempos de execução e taxas de acerto quando comparados a detecção manual. Os autores salientam que todos os algoritmos estudados apresentaram um conjunto de parâmetros internos que precisam ser selecionados pelo usuário, fato que dificulta o seu uso em ambiente clínico. A conclusão final do trabalho apresenta o SVM como o melhor

considerando a sua performance global, mas que ainda podem realizadas diversas melhorias nessa linha de pesquisa.

Como dito anteriormente, de maneira geral, os métodos propostos para automatização da análise das lesões em substância branca do cérebro trabalham com a localização e segmentação de lesões causadas por uma doença específica, sem caracterizar as lesões de acordo com sua etiologia. Se essa informação fosse disponível, poderia ser utilizada em muitos casos, como no diagnóstico e tratamento de pacientes neurológicos que apresentam lesões na substância branca de etiologia desconhecida.

Tendo em vista a relevância da caracterização das lesões, o método proposto visa tanto identificar quanto classificar regiões de interesse contendo lesões de etiologia isquêmica ou desmielinizante. A abordagem adotada utiliza análise de textura, seguida por uma etapa de classificação, com a finalidade de extrair e identificar informações relevantes para diferenciar tecidos saudáveis de tecidos lesionados, assim como diferenciar as lesões de acordo com sua etiologia.

A análise de textura é uma área de processamento de imagens que tenta extrair características relevantes das imagens através do estudo dos padrões de textura existentes (Haralick et al., 1973). A análise de textura tem sido utilizada em várias aplicações médicas (Castellano et al., 2004), como, por exemplo, no diagnóstico e acompanhamento da doença de Alzheimer (Freeborough and Fox, 1998) e na caracterização de tecidos no cérebro (Lerski et al., 1993). O trabalho apresentado em (Kovalev et al., 2001) realiza uma análise de textura em imagem de RM do cérebro e demonstra que os atributos de textura podem ser utilizados como uma ferramenta eficiente em diversas aplicações, como classificação de tecidos cerebrais e segmentação de lesões em imagens de difusão. Baseado nesse artigo, um método de segmentação de lesões em imagens de difusão da substância branca do cérebro através da análise de textura é proposto (Kruggel et al., 2008).

A análise de textura deve estar associada a uma etapa de classificação de padrões para a identificação e classificação das lesões em substância branca do cérebro. A classificação é a tarefa mais importante na área de reconhecimento de padrões e tem sido utilizada em muitas aplicações médicas a fim de auxiliar os profissionais da medicina a obter maior eficiência e eficácia, como em (Cocosco et al., 2003), (Glass et al., 1997) e (Maes et al., 1999). Existe uma extensa variedade de classificadores na literatura, mas, neste trabalho, faremos uso da floresta de caminhos ótimos (OPF), máquinas de vetores suporte (SVM), análise discriminante linear (LDA) e do método de k-vizinhos mais próximos (kNN). O classificador OPF foi selecionado por apresentar vantagens como: não assumir separabilidade das classes no espaço de atributos, ser rápido, livre de parâmetros e intrinsecamente multiclasse (Papa et al., 2009). O classificador SVM, por sua vez, foi escolhido por apresentar vários resultados satisfatórios em diversas aplicações na literatura (Cortes and Vapnik, 1995), assim como o LDA (Duda et al., 2000). Por fim, também realizamos estudos comparativos com o kNN (Dudani, 1976), por ser amplamente utilizado na comunidade acadêmica e apresentar similaridades conceituais com o classificador OPF.

## 0.4 Apresentação da Dissertação

Essa dissertação é organizada em quatro capítulos. O capítulo um provê um resumo sobre os materiais e métodos utilizados nos algoritmos implementados e propostos. Este capítulo descreve basicamente as técnicas de extração de atributos de textura, a etapa de seleção de atributos, e os classificadores utilizados ( máquina de vetor suporte, análise discriminante linear, floresta de caminhos ótimos e k-vizinhos mais próximo), e por fim o método de fusão de classificadores. No final do capítulo são introduzidas as métricas de avaliação da metodologia desenvolvida extraídas a partir da matriz de confusão.

O capítulo dois indica as etapas da metodologia desenvolvida. A etapa inicial do processo foi a aquisição do banco de imagens e a segmentação manual, seguida pela etapa de pré-processamento. Posteriormente, foi realizada a análise de textura, que representa a extração dos atributos de textura, seguida pela seleção dos atributos mais relevantes e menos redundantes. O próximo passo adotado foi a etapa de identificação e classificação das lesões, utilizando diferentes classificadores e também a combinação deles.

No capítulo três são indicados e discutidos os resultados experimentais obtidos através da análise da matriz de confusão. O último capítulo é a conclusão deste trabalho de mestrado, mostrando as contribuições do projeto, assim como as possíveis abordagens para serem trabalhadas em trabalhos futuros a partir dos resultados aqui obtidos.

## Definições

Este capítulo tem por objetivo apresentar os principais conceitos e técnicas que serão utilizados na metodologia proposta. Inicialmente, será feita uma introdução teórica na área de pré-processamento de imagens, seguida pela área de descrição de imagens, como as técnicas de extração e seleção de atributos, suas definições e representações matemáticas. A seguir, serão apresentados os métodos de classificação e de fusão de classificadores utilizados para o processamento dos atributos extraídos. Por fim, será indicada a técnica de avaliação dos resultados que foi adotada.

### 1.1 Pré-Processamento

A etapa de pré-processamento é responsável por modificar a imagem adquirida na etapa de aquisição com o objetivo de facilitar a sua visualização e o seu processamento posterior. Existem vários tipos de transformações lineares ou não lineares que podem ser aplicadas com diferentes metas, como: melhoramento de contraste, remoção de ruído, realce de bordas, entre outras. Essas transformações podem ser realizadas localmente ou globalmente, no domínio do espaço ou no domínio da frequência (Woods and Gonzalez, 2000).

Uma técnica de pré-processamento largamente aplicada em imagens de ressonância magnética (RM) é a normalização dos níveis de cinza. Este método tem por objetivo minimizar a influência de características globais da imagem, como contraste e brilho, nos descritores extraídos das regiões de interesse a serem analisadas posteriormente. As aplicações médicas em geral são sensíveis à escolha dos equipamentos de aquisição das imagens e por isso enfrentam dificuldades para serem amplamente utilizadas no ambiente clínico. As intensidades nas imagens de RM, por exemplo, podem variar mesmo se forem utilizados o mesmo equipamento e protocolo de aquisição. Dessa forma, a normalização de imagens é empregada para minimizar a influência dos parâmetros de aquisição das imagens de RM (Collewet et al., 2004).

Existem diferentes maneiras de realizar a normalização das imagens. A forma mais simples é aplicar uma transformação multiplicativa para fixar a mesma faixa de nível de cinza para todas as imagens de um mesmo protocolo de aquisição. Outro método de normalização

largamente utilizado na literatura é a normalização na faixa de valores  $[\mu \pm 3\sigma]$ , onde  $\mu$  representa a média dos níveis de cinza, enquanto  $\sigma$  corresponde ao seu desvio-padrão. Este tipo de normalização tem por objetivo remover a dependência de possíveis deslocamentos do valor médio dos níveis de cinza, assim como remover a dependência de mudanças multiplicativas na intensidade das imagens. É importante salientar que esta normalização, diferentemente da transformação multiplicativa, não preserva a variação relativa entre dois níveis de cinza.

De forma geral, a etapa de pré-processamento é utilizada em sistemas de reconhecimento de padrões com o objetivo de facilitar as etapas posteriores. A técnica ideal para cada aplicação varia de acordo com o conjunto de imagens a ser processado, condições de aquisição e do objetivo final da análise. No ambiente clínico, esta etapa é geralmente utilizada para uniformizar os níveis de cinza das imagens, realçar determinadas características ou extrair ruídos ou artefatos (Shih, 2010).

## 1.2 Reconhecimento de Padrões

O processo de reconhecimento de padrões é uma das tarefas da área de aprendizado de máquina cujo objetivo é classificar e rotular informações baseado em um conhecimento prévio. O processo é desenvolvido de forma que inicialmente seja encontrado um modelo que melhor descreva os dados e distinga as diferentes classes (ou conceitos), para que em um segundo momento, este modelo possa ser utilizado para prever as classes de objetos ainda não caracterizados. A etapa inicial no desenvolvimento e implementação de um sistema de reconhecimento de padrões se inicia com o recolhimento das informações que serão classificadas ou descritas através do uso de sensores. No caso de aplicações de imagens médicas, essa etapa é representada pela aquisição das imagens através do uso de equipamentos variados, como os de ressonância magnética ou tomógrafos. A seguir será aplicada a etapa de extração de atributos na qual serão computadas informações numéricas ou simbólicas dos dados adquiridos. A etapa final é a classificação dos dados, que é intrinsecamente dependente das características extraídas na etapa anterior. A Figura 1.1 representa um sistema completo de reconhecimento de padrões.

Nesta seção será realizada a descrição teórica das etapas de reconhecimento de padrões que serão utilizadas neste trabalho de mestrado, abordando as principais técnicas e os algoritmos mais utilizados no ambiente de pesquisa científica.

### 1.2.1 Extração de Atributos

Uma das tarefas mais complexas presentes na análise de imagens é a definição de um conjunto de características capazes de descrever de maneira efetiva cada região contida em uma imagem, de modo a ser utilizado em processos de mais alto nível, tal como classificação de padrões (Schwartz et al., 2012). Existem descritos na literatura atributos de cor, forma e textura, mas por se tratar de uma aplicação com imagens em nível de cinza, os atributos de cor não serão utilizados. Os atributos de forma, por sua vez, também não serão utilizados,

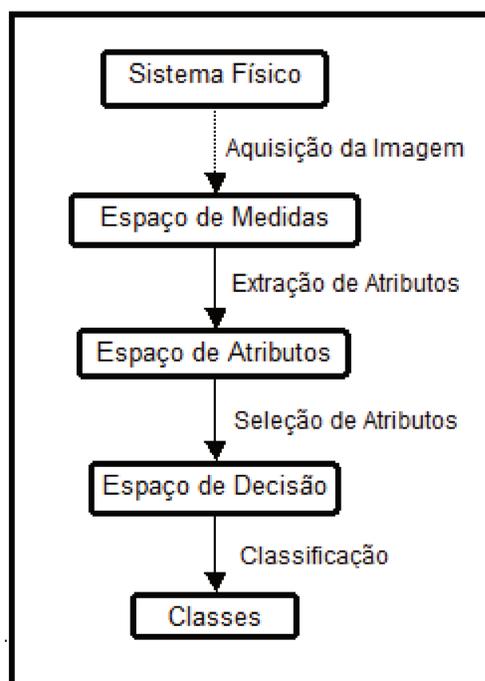


Figura 1.1: Fluxograma de um sistema de reconhecimento de padrões

visto que as regiões de interesse delineadas manualmente não possuem tamanhos ou formatos específicos. Dessa forma, apenas os atributos de textura são utilizados na tarefa de classificação de tecidos lesionados em imagens de ressonância magnética (Klöppel et al., 2011). Existem na literatura várias abordagens de análise de textura. Apresentaremos a seguir as utilizadas neste trabalho: momentos estatísticos extraídos do histograma (H), matriz de co-ocorrência (MC), matriz de comprimento de corrida (CC), gradiente (G), autoregressivos (AR) e *wavelet*(W).

### Abordagem estatística de primeira ordem baseada no histograma

Momentos estatísticos extraídos do histograma (H) oferecem uma descrição das características de um determinado conjunto de atributos e proporcionam uma análise quantitativa da distribuição dos atributos. Dessa forma, através da utilização dos momentos estatísticos, informações relevantes sobre a distribuição da intensidade dos pixels na imagem podem ser extraídas.

O primeiro momento, dado na Equação 1.1, indica a média do histograma (MédiaH) das imagens analisadas. O segundo momento central, representado na Equação 1.2, indica a variância do histograma (VariânciaH) e incorpora as informações sobre a escala ou a amplitude da distribuição dos níveis de cinza. O terceiro momento central, apresentado na Equação 1.3, mede a assimetria dos dados em torno da média amostral (AssimetriaH), descrevendo como os valores dos dados são distribuídos em torno de seu centro. Por fim, o quarto momento, indicado na Equação 1.4, avalia se a distribuição analisada é achatada ou alongada, também

conhecida como curtose do histograma (CurtoseH).

$$\mu_H = \frac{1}{N} \sum_{n=1}^N x_n \quad (1.1)$$

$$\sigma_H^2 = \frac{1}{N} \sum_{n=1}^N (x_n - \mu_H)^2 \quad (1.2)$$

$$\gamma_H = \frac{1}{N} \sum_{n=1}^N (x_n - \mu_H)^3 \quad (1.3)$$

$$\beta_H = \frac{1}{N} \sum_{n=1}^N (x_n - \mu_H)^4 \quad (1.4)$$

Nas Equações 1.1 a 1.4, o valor  $x_n$ , em que  $n = 1 \dots N$ , representa a frequência do histograma para cada valor de *pixel* e  $N$  indica a quantidade de níveis de cinza presentes.

### Matriz de co-ocorrência

As informações de textura estão contidas na relação espacial entre a intensidade de nível de cinza de um *pixel* e de seu vizinho (Haralick et al., 1973). A distribuição de intensidade de nível de cinza pode ser especificada por uma matriz de frequências relativas, em que dois elementos vizinhos de textura ocorrem na imagem. Estes mesmos elementos são separados por uma distância  $d$  em uma orientação  $q$ : o primeiro com valor de nível de cinza  $i$  e o segundo com valor  $j$  conforme ilustrado na Figura 1.2. A orientação  $q$  pode assumir valores iguais a  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  e  $135^\circ$ .



Figura 1.2: Distâncias e orientações utilizadas no cálculo da matriz de co-ocorrência.

Esta matriz de frequências relativas é chamada de matriz de co-ocorrência de níveis de cinza (MC) e quantifica a ocorrência de determinada relação espacial entre *pixels* vizinhos em uma imagem. O tamanho da MC depende da quantização adotada na aquisição da imagem. Uma quantização de 8 bits por exemplo, gera imagens com valores de nível de cinza que variam de  $[0,255]$  e portanto matrizes de co-ocorrência com tamanho de  $256 \times 256$ . As MCs computadas em imagens adquiridas com essa quantização são geralmente esparsas e apresentam maior custo de processamento nas etapas posteriores. Dessa forma, para que sejam geradas MCs mais compactas, basta que seja modificada a quantização utilizada na imagem.

A Figura 1.3 apresenta um exemplo de MC de tamanho 4 x 4, utilizando  $d = 1$  e  $q = 0$ , calculada portanto de uma imagem cujos valores de nível de cinza variam entre  $[0,3]$ . O primeiro elemento dessa matriz, por exemplo indica quantos pixels com valor zero ocorrem com pixels vizinhos a direita com valor também igual zero.

3	2	0	1	0
1	2	1	3	0
3	1	0	2	3
1	2	3	0	3
0	0	0	0	1

(a)

	0	1	2	3
0	3	2	1	1
1	2	0	2	1
2	1	1	0	2
3	2	1	1	0

(b)

Figura 1.3: Exemplo de matriz de co-ocorrência: (a) imagem original; (b) matriz de co-ocorrência utilizando  $d = 1$  e  $q = 0$ .

A MC engloba um conjunto de atributos de textura, alguns dos quais descritos a seguir:

$$Energia = \sum_{i,j} S^2(i,j) \quad (1.5)$$

$$Contraste = \sum_k k^2 \sum_i \sum_j S(i,j) \quad (1.6)$$

$$Correlação = \sum_{i,j} \frac{ijS(i,j) - \mu_i\mu_j}{\sigma_i\sigma_j} \quad (1.7)$$

$$Soma \text{ dos Quadrados} = \sum_{i,j} (i - \mu)^2 S(i,j) \quad (1.8)$$

$$Homogeneidade = \sum_{i,j} \frac{S(i,j)}{1 + (i - j)^2} \quad (1.9)$$

$$Entropia = - \sum_{i,j} S(i,j) \log(S(i,j)) \quad (1.10)$$

A matriz  $S$  representa a MC, o índice de soma  $k$  na Equação 1.6 é definido pela ordem da MC menos um e os índices  $i$  e  $j$  são equivalentes à quantidade de níveis de cinza da imagem analisada e variam de acordo com a quantização adotada. Os parâmetros,  $\mu_i, \mu_j, \sigma_i$  e  $\sigma_j$  representam, respectivamente, a média e o desvio-padrão da linha  $i$  e coluna  $j$  da MC.

A energia definida na Equação 1.5, também conhecida como a uniformidade da energia ou segundo momento angular, expressa a uniformidade dos padrões de textura e é definida como a soma dos quadrados dos elementos na MC. O contraste mede a intensidade de dissimilaridade entre um *pixel* e seu vizinho. A correlação, definida na Equação 1.7, é uma medida estatística que representa a dependência linear entre os tons de nível de cinza presentes na imagem, enquanto a soma dos quadrados mede o grau de dispersão dos elementos da MC da média.

A homogeneidade, por sua vez, informa a similaridade entre os valores de nível de cinza dos *pixels* da imagem. Por fim, a entropia mede o grau de desordem na MC.

As características de homogeneidade e contraste identificam estruturas organizadas na imagem, enquanto a energia, correlação, entropia e a soma dos quadrados caracterizam a complexidade e a natureza das transições de nível de cinza na imagem. Estas características de textura são armazenadas para análise na etapa posterior de classificação dos padrões.

### Matriz de comprimento de corrida

Uma sequência de pixels consecutivos em uma imagem com mesmo nível de cinza define uma corrida. De forma a armazenar e estudar o comportamento dessas corridas, foi proposta a matriz de comprimento de corrida (Schwartz, 2005). Esta matriz  $P(i,j)$  é definida pelo número de corridas com pixels de nível de cinza  $i$  e de comprimento de corrida  $j$ .

A Figura 1.4 apresenta um exemplo de matriz de comprimento de corrida (CC), calculada da mesma imagem original mostrada na Figura 1.3. O primeiro elemento dessa matriz, diferentemente da matriz de co-ocorrência, indica quantos pixels com valor zero ocorrem sem nenhum vizinho na horizontal com o mesmo valor, enquanto o elemento de índice (0,4), que tem valor igual a um, indica que há uma única ocorrência de quatro zeros consecutivos.

3	2	0	1	0
1	2	1	3	0
3	1	0	2	3
1	2	3	0	3
0	0	0	0	1

(a)

	1	2	3	4
0	5	0	0	1
1	6	0	0	0
2	4	0	0	0
3	6	0	0	0

(b)

Figura 1.4: Exemplo de matriz de comprimento de corrida: (a) imagem original; (b) matriz de comprimento de corrida

Em uma textura áspera, espera-se maior frequência de corridas de comprimentos mais longos, enquanto em texturas mais finas ocorrem corridas mais curtas, devido a presença de bordas, responsáveis por variações rápidas e acentuadas nos níveis de cinza dos pixels pertencentes a uma determinada linha da imagem, causando desta maneira, interrupções frequentes nas corridas. Assim como as matrizes de co-ocorrência, a matriz de comprimento de corrida (CC) também pode ser calculada para diferentes orientações, como pode ser visto na Figura 1.5.

Vários atributos de textura podem ser extraídos a partir dessa matriz como a ênfase em corrida curta (SRE), ênfase em corrida longa (LRE), não-uniformidade de nível de cinza (GLN), não-uniformidade de comprimento de corrida (RLN) e percentagem de corrida (RP). As Equações 1.11 a 1.15 apresentam essas medidas de textura, em que  $n_c$  denota o número total de corridas, enquanto  $H_g$  e  $N_r$  representam o valor do nível de cinza máximo e o tamanho da corrida mais longa, respectivamente.

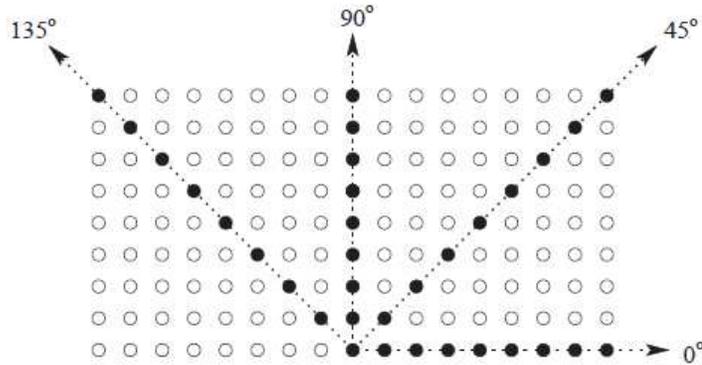


Figura 1.5: Representação das diferentes orientações nas quais a matriz de comprimento de corrida pode ser calculada (Schwartz, 2005).

$$SRE = \frac{1}{n_c} \sum_{i=0}^{H_g} \sum_{j=1}^{N_r} \frac{P(i,j)}{j^2} \quad (1.11)$$

$$LRE = \frac{1}{n_c} \sum_{i=0}^{H_g} \sum_{j=1}^{N_r} j^2 P(i,j) \quad (1.12)$$

$$GLN = \frac{1}{n_c} \sum_{i=0}^{H_g} \left( \sum_{j=1}^{N_r} P(i,j) \right)^2 \quad (1.13)$$

$$RLN = \frac{1}{n_c} \sum_{i=1}^{N_r} \left( \sum_{j=0}^{H_g} P(i,j) \right)^2 \quad (1.14)$$

$$RP = \frac{n_c}{n_p} \quad (1.15)$$

As métricas com SRE e LRE representam, respectivamente, a frequência com que as corridas curtas e longas são encontradas na textura. A medida GLN, por sua vez, descreve a distribuição dos tons de cinza e valores baixos indicam corridas homogêneas nos tons de cinza. Por fim, a RLN descreve a distribuição dos tamanhos de corrida e mede o quão uniforme é a distribuição dos tamanhos de corrida.

## Gradiente

O gradiente (G) de uma imagem descreve mudanças direcionais na intensidade nos níveis de cinza, de forma que o gradiente pode ser utilizado para extrair informações relevantes da imagem (Woods and Gonzalez, 2000). O gradiente é largamente utilizado na literatura como um método de detecção de bordas, no qual os pixels com maior valor de gradiente são denominados pixels de borda. Além dessa aplicação, o gradiente de uma imagem também pode ser utilizado para análise de textura. Os momentos extraídos a partir do gradiente da

imagem são menos suscetíveis a mudanças na iluminação do ambiente, ou nas propriedades do equipamento de aquisição, sendo, portanto, mais invariantes.

G	H	I
L	$x(i,j)$	N
Q	R	S

Figura 1.6: Definição da vizinhança do pixel  $x(i,j)$

O gradiente de uma imagem é definido em uma determinada vizinhança. Supondo uma imagem  $f$  com  $M$  pixels e considerando um pixel  $x(i,j)$  pertencente a esta imagem, cuja vizinhança é apresentada na Figura 1.6, pode-se definir a função de vizinhança deste pixel através da Equação 1.16. A partir dessa função de distância, os atributos de textura média do gradiente ( $\mu_G$ ), variância do gradiente ( $\sigma_G^2$ ), assimetria do gradiente ( $\gamma_G$ ) e curtose do gradiente ( $\beta_G$ ) são definidos nas Equações 1.17 a 1.20.

$$Viz(i,j) = \sqrt{(R - H)^2 + (N - L)^2} \quad (1.16)$$

$$\mu_G = \frac{1}{M} \sum_{i,j \in f} Viz(i,j) \quad (1.17)$$

$$\sigma_G^2 = \frac{1}{M} \sum_{i,j \in f} (Viz(i,j) - \mu_G)^2 \quad (1.18)$$

$$\gamma_G = \frac{1}{(\sqrt{\sigma_G^2})^3} \frac{1}{M} \sum_{i,j \in f} (Viz(i,j) - \mu_G)^3 \quad (1.19)$$

$$\beta_G = \frac{1}{(\sqrt{\sigma_G^2})^4} \frac{1}{M} \sum_{i,j \in f} (Viz(i,j) - \mu_G)^4 - 3 \quad (1.20)$$

### Abordagem baseada em modelos paramétricos auto-regressivo

Nesta abordagem, a textura é considerada como uma amostra extraída de um processo não-determinístico definido por um conjunto de parâmetros que, servindo como modelo para textura, resumem suas características. Com a utilização de parâmetros, pode-se efetuar análise de textura. Após a digitalização, os pixels de uma imagem monocromática assumem valores fixos, e a relação entre pares de pixels é dita determinística. Entretanto, se consideramos que cada pixel pode assumir valores pertencentes a um intervalo com uma dada probabilidade, a imagem torna-se uma cadeia aleatória e a interação de seus componentes passa a ser probabilística (Schwartz et al., 2012).

O modelo autoregressivo (AR) assume uma interação local entre os pixels de uma imagem, de forma que a intensidade de um pixel pode ser representada pela soma ponderada da

intensidade dos pixels vizinhos (Fukunaga, 1991). Assumindo que uma imagem  $f$  é uma cadeia aleatória de média nula, o modelo causal autoregressivo pode ser definido pela Equação 1.21, na qual  $f_s$  é a intensidade da imagem no pixel  $s$ ,  $e_s$  denota ruído independentemente e identicamente distribuído,  $N_s$  representa a vizinhança de  $s$  e  $\theta$  o vetor de parâmetros do modelo. A vizinhança local para o modelo AR, representada por quatro parâmetros, é mostrada na Figura 1.7. A região hachurada na Figura 1.7 indica a região onde a vizinhança válida do modelo AR é localizada.

$$f_s = \sum_{r \in N_s} \theta_r f_r + e_s \quad (1.21)$$

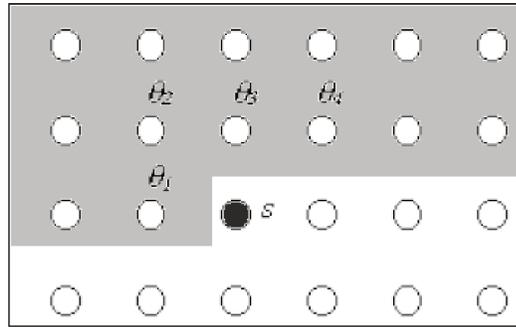


Figura 1.7: Definição da vizinhança do pixel  $s$  da imagem  $f$

Os parâmetros do modelo AR podem ser utilizados para análise textural. No caso da vizinhança definida na Figura 1.7, que apresenta quatro pixels vizinhos, podem ser definidos cinco parâmetros do modelo AR: desvio-padrão do ruído  $e_s$  e o vetor de parâmetros  $\theta = [\theta_1, \theta_2, \theta_3, \theta_4]$ . Através da minimização da soma do erro quadrático, definida na Equação 1.22, os parâmetros responsáveis pela geração de uma determinada textura podem ser estimados pelas equações apresentadas em 1.23 e 1.24, onde  $w_s = col[f_i, i \in N_s]$ .

$$\sum_s e_s^2 = \sum_s (f_s - \theta w_s)^2 \quad (1.22)$$

$$\theta = \left( \sum_s w_s w_s^T \right)^{-1} \left( \sum_s w_s f_s \right) \quad (1.23)$$

$$\sigma^2 = N^{-2} \sum_s (f_s - \theta w_s)^2 \quad (1.24)$$

### Abordagem baseada em processamento de sinais - wavelet haar

A transformada *wavelet haar* (W) aplicada em imagens é discreta e 2D, e convencionalmente implementada de forma separável, resultando da aplicação em cascata de duas

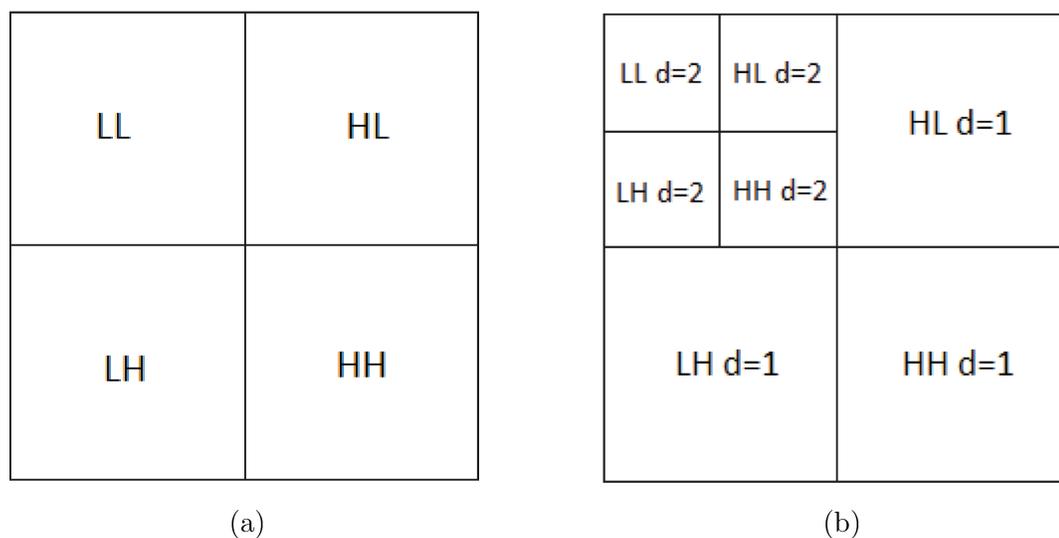


Figura 1.8: Decomposição resultante da transformada Wavelet (a) decomposição em um nível; (b) decomposição em dois níveis Schwartz et al. (2008)

transformadas *wavelet* unidimensionais (Doering, 2005). O resultado da transformada *wavelet* é a decomposição da imagem original em subimagens de acordo com seus componentes de alta ou baixa frequência, como representado na Figura 1.8.

A partir dessas subimagens, pode-se extrair características texturais das regiões de interesse, como a energia de uma subimagem. A Equação 1.25 indica como é realizado o cálculo dessa energia, na qual  $w(l,m)$  representa os coeficientes resultantes da transformada,  $E_{S_i}$  denota a energia da  $i$ -ésima subimagem e  $n_s \times n_s$  apresenta as suas dimensões (Schwartz et al., 2008).

$$E_{S_i} = \frac{1}{n_s^2} \sum_{l=0}^{n_s-1} \sum_{m=0}^{n_s-1} w(l,m)^2 \quad (1.25)$$

## 1.2.2 Seleção de Atributos

Os dados computados na etapa anterior podem conter uma quantidade tão grande de atributos que a sua análise e processamento posterior se tornam impraticáveis. Além disso, alguns destes atributos podem ser redundantes ou irrelevantes. Dessa forma, em um sistema de reconhecimento de padrões, o processo de seleção de atributos reduz o tamanho do conjunto de dados através da remoção de atributos irrelevantes ou redundantes com o objetivo de encontrar um conjunto mínimo de atributos que represente o conjunto original e torne a tarefa de classificação mais fácil e direta. Apresentaremos a seguir duas técnicas de seleção de atributos que possuem metodologias distintas.

## Árvore de Decisão

O método da Árvore de Decisão (Han and Kamber, 2006) foi inicialmente desenvolvido como um algoritmo para classificação, mas também pode ser utilizado como seletor de atributos. O objetivo desta técnica é gerar uma estrutura em forma de árvore que sumarie informações relevantes dos dados de entrada, possibilitando sua correta visualização e interpretação. A estrutura da árvore é gerada de forma que cada nó interno represente um teste realizado em um atributo, enquanto cada ramo indica os possíveis resultados deste teste. Os nós externos, por sua vez, apresentam a predição final das classes.

A Figura 1.9 mostra um conjunto de dados representados na estrutura de uma árvore de decisão. O conjunto inicial de dados apresenta 6 atributos, indicados pelas variáveis:  $A_1$ ,  $A_2$ ,  $A_3$ ,  $A_4$ ,  $A_5$  e  $A_6$ . O algoritmo da árvore de decisão escolhe o melhor atributo para particionar os dados dentre as diversas classes existentes. Dessa forma, no nosso exemplo, o atributo com o maior grau de separabilidade e, portanto, o primeiro a ser selecionado é o  $A_4$ . Os dados então são particionados de acordo com a regra de decisão estabelecida por este atributo, gerando duas ramificações indicadas na Figura como  $S$  (classe positiva) e  $N$  (classe negativa). A seguir, é novamente realizada a escolha do atributo que melhor particiona os dados remanescentes. E o processo continua, até que todas as partições geradas contenham apenas dados de uma determinada classe.

No nosso caso de estudo apenas os atributos  $A_1$ ,  $A_4$  e  $A_6$  foram necessários para a construção da árvore de decisão, de forma que os outros atributos não apresentaram informações relevantes para a análise do conjunto de dados em questão e portanto são ditos irrelevantes ou redundantes. Dessa forma, o conjunto de atributos que constitui a árvore de decisão forma o subconjunto otimizado de atributos.

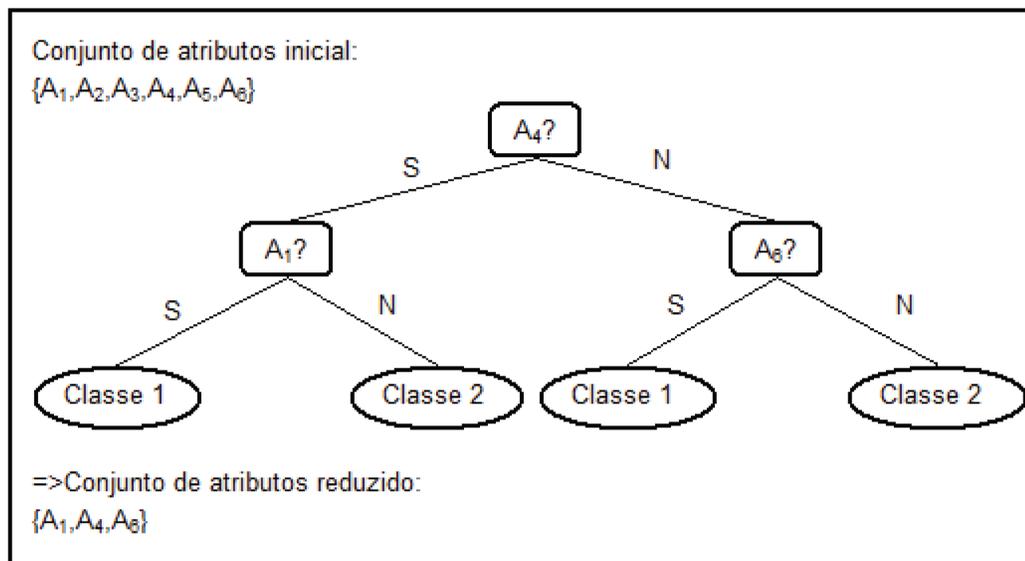


Figura 1.9: Representação do método de seleção de atributos árvore de decisão com o conjunto de atributos inicial e reduzido (Adaptada de (Han and Kamber, 2006))

## Análise por Componentes Principais

A análise por componentes principais (*Principal Component Analysis* - PCA), proposta em 1901 por Karl Pearson (Pearson, 1901), é um procedimento matemático que reduz o conjunto de atributos através da eliminação de informações redundantes. O resultado é um número menor de atributos não correlacionados, chamados componentes principais. Esta representação mais compacta dos atributos faz com que o processo da classificação ou reconhecimento de atributos seja acelerado (Smith, 2002). A PCA também pode identificar padrões no conjunto de dados, buscando similaridades e diferenças.

O cálculo das componentes principais é realizado utilizando-se a matriz de covariância do conjunto de amostras e os autovalores e autovetores correspondentes. O objetivo da técnica é encontrar as direções que indicam a maior variância do conjunto de amostras. Se um autovetor possui um autovalor associado com grande magnitude, significa que se localiza em uma direção em que há grande variância nos padrões.

Os componentes principais são autovetores da matriz de covariância, sendo que o mais relevante componente principal é um autovetor referente ao maior autovalor, o segundo componente principal é um autovetor associado ao segundo maior autovalor, e assim sucessivamente. A redução de atributos é um problema de otimização que busca criar novas características por meio da combinação do conjunto original, mantendo a eficiência no processo de classificação (Pechenizkiy et al., 2004).

### 1.2.3 Classificação de Padrões

A classificação consiste na etapa final de um sistema de reconhecimento de padrões e tem como objetivo analisar os diferentes padrões que os dados possam apresentar. O processo de classificação pode ser dividido em duas etapas principais: treinamento e teste. Na etapa de treinamento, o classificador é construído de forma a descrever um conjunto pré-determinado de classes e conceitos, chamado de conjunto de treinamento. Na etapa de teste, o classificador prediz a classe de uma nova amostra, baseado no modelo construído na etapa anterior.

Existem duas formas conhecidas de aprendizado: o supervisionado e o não supervisionado. No primeiro caso, o processo de aprendizado (treino) do classificador se dá em um conjunto de treinamento cujas classes ou rótulos são conhecidas, enquanto no segundo, o classificador não tem informações sobre as classes nos dados de treinamento (Han and Kamber, 2006).

A literatura apresenta uma grande diversidade de algoritmos de classificação. Neste trabalho, a escolha dos métodos de aprendizado foi realizada de acordo com as seguintes características: lidar de maneira eficiente com problemas multi-classe, apresentar boa performance em trabalhos relacionados, baixo tempo de processamento, fácil entendimento e implementação.

## Máquina de Vetores Suporte

O classificador Máquina de Vetores Suporte (*Support Vector Machine* - SVM) é um método de aprendizado supervisionado que pode ser aplicado em problemas de classificação e

regressão (Boser et al., 1992). O SVM realiza a tarefa de classificação através da construção de um conjunto de hiperplanos em um espaço de alta dimensão que otimiza a separação dos dados entre duas categorias.

No caso dos dados serem linearmente separáveis, o SVM busca o hiperplano com a maior margem entre as classes, que gera a maior distância entre classes. Dessa forma, a distância entre os pontos mais próximos de cada um dos lados do hiperplano é maximizada. Uma versão modificada do SVM linear é o *soft-margin* que permite que alguns amostras de treinamento sejam classificadas erroneamente, afim de evitar o *over-fitting*, mantendo uma margem alta entre classes (Cortes and Vapnik, 1995).

O SVM também pode ser utilizado como um classificador não linear através da utilização dos *kernels*. Os *kernels* são funções de transformação (mapeamento) no espaço de atributos, com o objetivo de torná-lo linearmente separável. A busca por hiperplanos de alta margem entre classes é realizada no espaço de atributos mapeado, e o processo de classificação passa a ser similar ao adotado na classificação dos dados linearmente separáveis (Shawe-Taylor and Cristianini, 2000).

Dessa forma, o SVM apresenta variações na sua forma de funcionamento de acordo com a escolha do *kernel* e dos parâmetros. Mesmo que novos *kernels* estejam sendo propostos em novos estudos, os quatro kernels mais vastamente utilizados são:

- linear :  $K(x_i, x_j) = x_i^T x_j$ ,
- RBF:  $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \gamma > 0$ ,
- polinomial:  $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$ ,
- sigmoide:  $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$ ,

onde  $\gamma, r$  e  $d$  são parâmetros dos *kernels*. Os parâmetros especificados podem assumir valores arbitrários, tornando complexa a seleção da melhor configuração do SVM para determinada aplicação. A fim de minimizar esta dificuldade, foi proposta a técnica de *grid-search*, que busca valores ótimos para cada um dos parâmetros dentro de um intervalo específico (Bergstra and Bengio, 2012). Este método é computacionalmente custoso porque o modelo deve ser avaliado de acordo com uma métrica de performance para diversos valores dentro do *grid* de cada parâmetro.

## Floresta de Caminhos Ótimos

O classificador Floresta de Caminhos Ótimos (*Optimum Path Forest* - OPF) modela a tarefa de classificação de dados como um problema de partição de grafos. A OPF explora a conectividade entre amostras e classes no espaço de atributos. O método pode ser utilizado para classificação supervisionada ou não supervisionada. O algoritmo de aprendizado supervisionado da OPF considera que os dados de treinamento são nós de um grafo completo (com todos os nós conectados) no espaço de atributos. O peso dos arcos é igual a distância

entre os vetores de atributos dos nós. É possível utilizar diferentes funções de distância. Para selecionar os protótipos, é computada a árvore de caminhos mínimos (MST) do grafo completo, e os protótipos são definidos como os nós conectados na MST que possuem rótulos diferentes (Papa et al., 2009).

Na etapa seguinte, é aplicada a transformada imagem-floresta (IFT) para particionar o grafo em uma floresta de caminhos mínimos, em que cada árvore de caminho mínimo é definida por um protótipo (Falcão et al., 2004). A classificação usando OPF, como a maioria dos classificadores, é realizada em duas etapas: treinamento e teste. O treinamento essencialmente consiste em contruir a floresta de caminhos mínimos, em que as amostras em uma árvore de caminho mínimo possuem o mesmo rótulo do seu protótipo. A classificação de uma nova amostra (etapa de predição) é realizada considerando-a como parte do grafo original e o caminho ótimo ao novo nó computado. A amostra de teste recebe o mesmo rótulo do protótipo mais fortemente conectado a ela, i.e. o protótipo que oferece o caminho de menor custo.

### **k Vizinhos mais Próximos**

O classificador  $k$  vizinhos mais próximos (kNN) é uma generalização do classificador vizinho mais próximo (1NN). Este é eficiente, de simples entendimento e implementação e apresenta resultados expressivos na literatura (Cover and Hart, 1967). A regra de decisão do classificador Vizinho Mais Próximo atribui à cada amostra de teste o rótulo da amostra mais próxima do conjunto de treinamento. Existem várias métricas que podem ser utilizadas para calcular a distância entre as amostras, como a distância Minkowski ou a distância de Manhattan, mas a mais comumente utilizada é a distância Euclidiana.

A principal desvantagem deste método é a sua sensibilidade à presença de amostras ruidosas ou irrelevantes. Além disso é importante ressaltar que o tempo de processamento do 1NN quando é utilizado um conjunto de treinamento volumoso pode inviabilizar seu emprego, visto que é necessário calcular a distância de uma nova amostra relativamente a todas as amostras do conjunto de treinamento (Bhatia, 2010). Este método apresenta muitas similaridades conceituais com o classificador OPF, de forma que é possível obter o classificador 1NN a partir do classificador OPF se considerarmos todas as amostras de treinamento da OPF como protótipos (Souza et al., 2012).

O classificador kNN é uma extensão natural do 1NN, e atribui à cada amostra de teste o rótulo mais frequente dentre as  $k$  amostras mais próximas. Como leva em consideração os  $k$  vizinhos, este classificador é menos sensível do que o 1NN à ruído e à *outliers* (Duda et al., 2000).

### **Análise Discriminante Linear**

A Análise Discriminante Linear (*Linear Discriminant Analysis* - LDA) utiliza informações associadas a cada padrão para extrair suas características discriminantes, baseando-se na densidade de probabilidade de cada classe (Webb, 2002).

A análise discriminante linear pode ser utilizada como algoritmo de classificação de padrões, ou como um método de redução de dimensionalidade do problema em questão. A análise é baseada na interpretação geométrica dos dados e visa gerar classes que possam ser diferenciadas entre si, mas que possuam atributos semelhantes. Para tal fim, é realizada a maximização da variância da matriz entre classes, aumentando a distância entre suas médias, e a minimização da variância da matriz dentro das classes.

O algoritmo consiste na computação de uma combinação linear de  $m$  variáveis quantitativas que mais eficientemente separam grupos de amostras em um espaço  $m$ -dimensional através da maximização da razão entre a variância dentro da classe e a variância entre classes. O espalhamento intra-classes é definido na Equação 1.26, em que  $x_i^j$  é o  $i$ -ésimo exemplo da classe  $j$ ,  $u_j$  é a média da classe  $j$ ,  $T_j$  é o número de amostras da classe  $j$ ,  $n_c$  representa a quantidade de classes e o  $t$  indica a operação de transposição. A Equação 1.27 representa o espalhamento inter-classes, em que  $u$  é a média de todas as classes,  $u_j$  e  $u$  são definidos nas Equações 1.28 e 1.29, respectivamente. O número de amostras da classe  $c$  é representado por  $T_c$ .

$$S_w = \sum_{i=1}^{n_c} \sum_i^{T_j} (x_i^j - u_j) \cdot (x_i^j - u_j)^t \quad (1.26)$$

$$S_b = \sum_{j=1}^{n_c} (u_j - u) \cdot (u_j - u)^t \quad (1.27)$$

$$u_j = \frac{1}{T_c} \sum_{j \in c} x_j \quad (1.28)$$

$$u = \frac{1}{T} \sum_j^c T_c u_j \quad (1.29)$$

Uma forma de seguir o critério de Fisher (Fisher, 2011) é maximizar a razão  $S_f = \det(S_b)/\det(S_w)$ , em que  $\det()$  indica o determinante, maximizando a medida inter-classe e minimizando a medida intra-classe. A idéia principal deste método é transformar observações multivariadas em univariadas, fazendo uma combinação linear entre elas, com o intuito de reduzir a dimensão dos dados e permitir a separabilidade das classes.

## Fusão de Classificadores

A fusão de classificadores (Tulyakov et al., 2008) é uma técnica largamente utilizada na literatura e tem como objetivo aumentar a acurácia do processo de classificação. As maneiras mais comuns de fazer essa composição dos métodos de aprendizagem são os algoritmos de *bagging* e de *boosting*. O algoritmo de *bagging* baseia-se no fato de que realizar uma votação entre os classificadores melhora a acurácia obtida. Dessa forma, o resultado final será gerado através de uma votação igualitária entre os classificadores. O *boosting*, por sua vez, também é gerado através da votação dos classificadores, mas diferentemente do método anterior, a

Tabela 1.1: Matriz de relação entre dois classificadores:  $c_i$  e  $c_j$ 

	Positivo $c_i$	Negativo $c_i$
positivo $c_j$	a	b
negativo $c_j$	c	d

votação é ponderada se baseando na acurácia independente de cada um dos métodos de classificação a serem combinados.

Além disso, existe também na literatura a fusão por SVM, na qual os resultados dos classificadores formam um conjunto de dados, utilizados para treinar e testar um classificador SVM. O resultado deste SVM, por sua vez, representa o resultado final do processo de fusão (Faria et al., 2012). Para essa fusão, inicialmente o conjunto de dados é subdividido em conjunto de treino, teste e validação. O conjunto de treino é utilizado para treinar os classificadores individuais, enquanto o conjunto de teste é utilizado para testá-los. A seguir, um classificador SVM é treinado com o resultado dos classificadores individuais no conjunto de teste. Por fim, este classificador SVM é testado com o resultado dos classificadores individuais no conjunto de validação, gerando o resultado da fusão.

Neste contexto de fusão de classificadores, podemos destacar as medidas de diversidade como uma maneira de relacionar classificadores de acordo com suas similaridades quando lidando com o mesmo conjunto de dados. Cada medida de diversidade computa os acordos/desacordos entre um par de classificadores. Logo, para se otimizar o processo de fusão dos classificadores, inicialmente computa-se essas medidas de diversidade, para que em uma etapa posterior apenas os métodos mais diversos sejam combinados (Faria et al., 2012).

As medidas de diversidade são calculadas a partir de uma matriz contendo a relação entre pares de classificadores, denotada  $M$ , com a porcentagem de concordância. Tabela 1.1 representa essa a matriz  $M$ , com a porcentagem de acertos e erros dos classificadores  $c_i$  e  $c_j$ . O valor  $a$  indica a porcentagem de dados que ambos os classificadores acertaram. Valores  $b$  e  $c$  são as porcentagens de dados que  $c_i$  acertou e  $c_j$  errou, ou vice-versa, enquanto  $d$  representam a porcentagem de dados que foram classificadas erroneamente pelos dois classificadores.

A partir da matriz  $M$ , as medidas de diversidade representadas nas Equações 1.30 a 1.32, denominadas medidas de dupla falha (DFM), Estatística-Q (QSTAT) e Acordo iterativo (IA) podem ser extraídas (Kuncheva and Whitaker, 2003).

$$DFM_{i,j} = d \quad (1.30)$$

$$QSTAT_{i,j} = \frac{ad - bc}{ad + bc} \quad (1.31)$$

$$IA_{i,j} = \frac{2(ac - bd)}{(a + b)(c + d) + (a + c)(b + d)} \quad (1.32)$$

## Matriz de Confusão

A matriz de confusão é utilizada para avaliar a eficácia, robustez e confiabilidade do classificador (Machado, 2007). A ordem da matriz de confusão depende da quantidade de classes envolvidas no classificador analisado. No caso de um classificador de duas classes, a matriz de confusão é formada por duas linhas e duas colunas, como indicado na Tabela 1.2. Os campos desta tabela indicam o número de verdadeiros negativos, falsos positivos, falsos negativos e verdadeiros positivos estimados pela comparação do resultado do modelo com os valores reais dos dados de validação. Valores positivos se referem à correta estimação do modelo, enquanto os valores falsos correspondem aos resultados estimados de maneira incorreta. A taxa de erro é quantificada pela soma dos valores falsos positivos e falsos negativos, enquanto a taxa de acerto é a soma dos valores verdadeiros positivos e verdadeiros negativos.

Tabela 1.2: Representação tabular da matriz de confusão

		Real	
		Positivo	Negativo
Estimado	Positivo	Verdadeiro Positivo (VP)	Falso Positivo (FP)
	Negativo	Falso Negativo (FN)	Verdadeiro Negativo (VN)

## Metodologia

A metodologia adotada neste projeto de pesquisa iniciou-se com a etapa de aquisição das imagens, a partir do equipamento de ressonância magnética do Hospital das Clínicas da Faculdade de Ciências Médicas na UNICAMP, seguida pela etapa de pré-processamento com o objetivo de uniformizar os níveis de cinza das regiões analisadas. A etapa adotada posteriormente foi a segmentação manual das regiões de interesse realizada por especialistas, seguida pela análise de textura, etapa na qual são extraídos os atributos de textura das regiões de interesse segmentadas anteriormente. Este conjunto de atributos de textura, por sua vez, foi submetido a um processo de seleção de atributos, com o objetivo de excluir atributos irrelevantes ou redundantes. Por fim, foi executada a etapa de identificação e classificação das lesões na substância branca do cérebro, através da aplicação dos diversos métodos de classificação de padrões descritos no capítulo anterior.

Neste capítulo, portanto, faremos uma descrição detalhada destas etapas também apresentadas na Figura 2.1, considerando questões de implementação, escolha dos parâmetros, ambiente de desenvolvimento, entre outros.

### **2.1 Aquisição das imagens, pré-processamento e segmentação manual das regiões de interesse**

O conjunto de dados que será utilizado foi adquirido no período entre janeiro de 2003 até dezembro de 2006 de 49 pacientes com esclerose múltipla, 4 pacientes com AVC e 19 controles com idades correspondentes aos pacientes. As imagens de RM ponderadas em T2 foram obtidas no plano axial (3mm espessura, ângulo *flip* de 120 graus, tempo de repetição de 6800 ms e tempo echo de 129 ms) em formato DICOM de um equipamento de 1.5 Tesla. O conjunto de imagens apresenta informações sobre o paciente como idade, sexo e data do diagnóstico. As imagens 3D foram salvas como um conjunto de fatias 2D, utilizando o *software* de imagens médicas MRÍcro <sup>1</sup>.

---

<sup>1</sup><http://www.cabiati.com/mricro/mricro/index.html>

Depois de adquirida a base de imagens, foi aplicado o método de normalização da imagem na faixa de valores  $[\mu \pm 3\sigma]$ , descrita no capítulo anterior, a fim de remover possíveis variações nos níveis de cinza de imagens adquiridas sob o mesmo protocolo de aquisição, utilizando o mesmo equipamento.

A seguir, foi realizada a etapa de segmentação manual de um conjunto de 283 ROIs contendo apenas um dos seguintes tecidos: substância branca normal, lesões de etiologia isquêmica ou lesões de etiologia desmielinizante, como representado na Figura 2.2. As regiões de interesse (ROIs) extraídas de fatias da imagem de RM apresentam diferentes tamanhos e formatos e possuem a seguinte distribuição: 76 ROIs representam substância branca normal, 64 ROIs representam lesões de substância branca de etiologia isquêmica, enquanto 143 ROIs representam lesões de etiologia desmielinizante.

## 2.2 Extração de atributos

Existem na literatura descrições de atributos de cor, forma e textura (Schwartz et al., 2012), mas dado que as lesões em substância branca não possuem tamanhos ou formatos específicos, e por se tratar de uma imagem de nível de cinza, os atributos de forma e cor não foram utilizados. Dessa forma, apenas atributos de textura serão extraídos.

Um conjunto de 87 atributos foi computado para cada ROI usando o *software* Mazda<sup>2</sup>. Os atributos de textura selecionados são baseados nas seguintes abordagens de análise de textura:

- Abordagem estatística baseada no histograma (H): medidas baseadas na distribuição de níveis de cinza (histograma) da região de interesse, como média do histograma (MédiaH), variância do histograma (VariânciaH), assimetria do histograma (AssimetriaH), curtose do histograma (CurtoseH), entre outros.
- Matriz de Co-Ocorrência de Nível de Cinza (MC): foram extraídas métricas a partir da MC calculada em diferentes distâncias (1, 2, 3, 4 e 5) e em diferentes orientações ( $0^0$ ,  $45^0$ ,  $90^0$  e  $135^0$ ). Além disso, também realizamos testes utilizando diferentes níveis de quantização da MC, considerando 4, 5 ou 6 *bits/pixel*.
- Matriz de comprimento de corrida (CC): assim como no caso da matriz de co-ocorrência, as métricas extraídas da matriz de comprimento de corrida foram calculadas sob diferentes orientações ( $0^0$ ,  $45^0$ ,  $90^0$  e  $135^0$ ) e considerando diferentes níveis de quantização (considerando 4, 5 ou 6 *bits/pixel*.)
- Abordagem baseada em gradiente (G): foram extraídos os atributos média do gradiente (MédiaG), variância do gradiente (VariânciaG), assimetria do gradiente (AssimetriaG), curtose do gradiente (CurtoseG) e percentagem de pixels não nulos no gradiente.

---

<sup>2</sup>[www.eletel.p.lodz.pl/programy/mazda/](http://www.eletel.p.lodz.pl/programy/mazda/)

- Abordagem baseada em modelo autoregressivo (AR): os parâmetros do modelo AR,  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$ ,  $\theta_4$  e  $\sigma$ , extraídos através da solução numérica das equações apresentadas no capítulo anterior.
- Abordagem baseada em Wavelet-Haar (W): diferentemente da MC e da CC, tem a sensibilidade para identificar alterações de textura mais grosseiras. Os atributos extraídos representam a energia de cada subimagem gerada pela decomposição Wavelet. Neste estudo, fizemos uso de 2 níveis de decomposição.

É importante salientar que nos casos onde os atributos são computados em diferentes orientações, como os MC e CC, foi realizada uma média daquele atributo em todas as orientações possíveis, dado que não existe orientação preferencial da textura, e portanto todos os atributos utilizados devem ser invariantes a orientação.

## 2.3 Seleção de atributos

A etapa de seleção de atributos realizada depois da extração dos atributos tem por objetivo excluir descritores irrelevantes e tornar o processo de classificação mais simples e direto. Fizemos uso de duas diferentes técnicas de seleção de atributos: árvore de decisão e análise de componentes principais (PCA).

O método da árvore de decisão, além de selecionar um conjunto de atributos, também possibilita a geração de uma lista com todos os atributos de acordo com o seu grau de discriminação na solução de determinada tarefa de classificação. Além disso, os atributos selecionados representam um subconjunto do conjunto original e não a combinação dos atributos originais, como é verificado no caso da análise das componentes principais.

A análise das componentes principais converte os atributos originais de forma a remover a correlação entre eles, e o conjunto selecionado de atributos não tem correspondência direta com o conjunto original. Além disso, este método requer que seja selecionada pelo usuário a quantidade final de atributos, enquanto no caso da árvore de decisão a quantidade de atributos final é encontrada automaticamente pelo próprio método. Neste trabalho ajustamos a mesma quantidade de atributos resultante da árvore de decisão para executar o método de PCA.

## 2.4 Identificação e Classificação de Lesões

Este projeto tem como principal proposta diferenciar tecidos saudáveis de lesionados, processo denominado de identificação de lesões, e caracterizar as lesões de acordo com sua etiologia, processo chamado de classificação de lesões. Dessa forma, a utilização dos diferentes métodos de classificação de padrões foram utilizados de acordo com as seguintes abordagens:

- Identificação de Lesões: distinguir as regiões de interesse de acordo com duas classes: lesão ou não lesão.

- Classificação de Lesões: distinguir as regiões de interesse representando as lesões de acordo com duas classes: lesão desmielinizante ou lesão isquêmica.
- Identificação e Classificação de Lesões em uma etapa única: distinguir as regiões de interesse dentre uma das seguintes classes: não lesão, lesão desmielinizante ou lesão isquêmica.

O processo de seleção de atributos, por sua vez, foi aplicado a cada uma das abordagens separadamente, selecionando o melhor conjunto de atributos para identificar as lesões e um outro conjunto otimizado para classificar as lesões. Foi também realizada a seleção de atributos para a terceira abordagem, na qual a identificação e a classificação das lesões foi realizada em uma etapa única, considerando as ROIs das três classes. A partir do conjunto de atributos de textura, foram projetados os classificadores. Um dos métodos de classificação utilizado foi o da OPF considerando diferentes funções de distâncias como a distância Euclidiana, Manhattan e Canberra. Outro classificador utilizado nos experimentos foi o SVM, considerando os kernels linear e rbf, e buscando o conjunto ótimo de parâmetros através da técnica de *grid-search* (Shawe-Taylor and Cristianini, 2000). Além disso, também foram realizados estudos comparativos com os classificadores LDA e kNN, usando  $k = 1$ ,  $k = 3$  e  $k = 5$ . As diferentes abordagens foram comparadas de acordo com as suas respectivas matrizes de confusão e taxas de acurácia.

A partição dos dados em conjuntos de treinamento e teste foi realizada utilizando o método de validação cruzada *k-fold* (Kohavi, 1995), utilizando folds de tamanho 10. Dessa forma, o conjunto de dados foi dividido aleatoriamente em 10 partes disjuntas, denominadas *folds*, uma parte foi selecionada para teste, e as remanescentes foram utilizadas para treinar o classificador. A seguir, o procedimento foi repetido dez vezes, de forma que cada um dos *folds*, foi utilizado uma única vez como conjunto de teste e nas outras instâncias foi incluído no conjunto de treino. O resultado final do classificador pode ser verificado através da média e desvio-padrão dos resultados de cada um dos classificadores modelados durante o processo.

### 2.4.1 Fusão de Classificadores

A fusão de classificadores foi realizada em cada uma das abordagens descritas acima, combinando as informações geradas por cada um dos classificadores, com o objetivo de obter melhores resultados. A fusão foi realizada tanto pelo método do *bagging*, no qual o resultado final foi gerado através de uma votação igualitaria entre os classificadores, como pela fusão por SVM, na qual os resultados dos classificadores treinam e testam um classificador SVM e este, por sua vez, gera o resultado da fusão. Além disso, foram realizados experimentos através da combinação de todos os classificadores, assim como através da utilização apenas daqueles selecionados através das métricas de diversidade.

## 2.5 Ambiente de Desenvolvimento

Os testes foram realizados no Adessowiki<sup>3</sup>. Adessowiki é um ambiente de desenvolvimento *web* colaborativo usado para escrita e programação científica. É adequado para comparação de soluções e *benchmarking* de algoritmos, visto que todos os programas são executados sob a mesma configuração (Lotufo et al., 2009). Foi utilizado o pacote *Scikits-learn* para rodar o SVM e o kNN. *Scikits-learn*<sup>4</sup> é um módulo *Python* que integra uma grande variedade de algoritmos de aprendizado de máquina para problemas de classificação supervisionada ou não supervisionada. Esse pacote é quase inteiramente escrito em *Python*, mas também incorpora bibliotecas C++, como a *libSVM* (Chang and Lin, 2011). Foi utilizada também a *libOPF3* para rodar a OPF. *LibOPF3*<sup>5</sup> é a implementação atual da OPF que apresenta um *binding* para linguagem *Python*.

---

<sup>3</sup>[www.adessowiki.org](http://www.adessowiki.org)

<sup>4</sup><http://scikit-learn.org/stable/>

<sup>5</sup><http://parati.dca.fee.unicamp.br/adesso/wiki/toolboxOPF/view/>

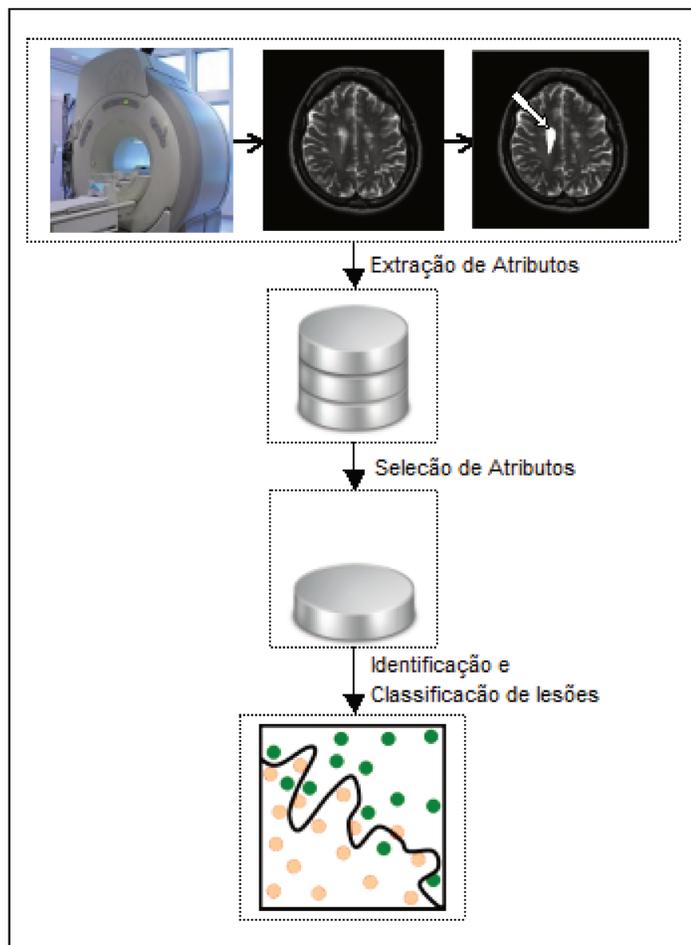


Figura 2.1: Passo a passo da metodologia: A etapa inicial é representada pela aquisição das imagens através do scanner de ressonância magnética, pré-processamento e segmentação manual das regiões de interesse; a segunda etapa é a análise de textura através da extração de atributos de textura; a terceira etapa é representada pela seleção de um conjunto otimizado de atributos; a última etapa é apresentada no fluxograma como identificação e classificação das lesões

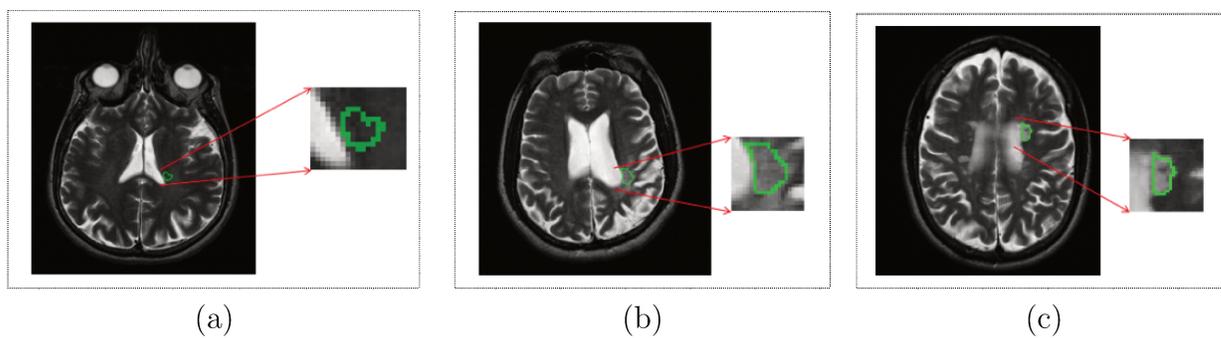


Figura 2.2: Exemplo do conjunto de dados e da extração das regiões de interesse. (a) Amostra de substância branca normal. (b) Amostra de lesão com etiologia isquêmica. (c) Amostra de lesão com etiologia desmielinizante.

## Experimentos, Resultados e Discussões

Neste capítulo, serão apresentados os resultados obtidos através da aplicação da metodologia proposta no capítulo 2 para a análise de lesões em substância branca do cérebro em imagens de ressonância magnética. Os resultados ilustram a capacidade de representação das regiões de interesse que os atributos de textura possuem, a efetividade dos métodos de seleção de atributos, assim como o desempenho dos métodos de classificação.

### 3.1 A análise de lesões em substância branca

Como dito anteriormente, a análise de lesões em substância branca do cérebro compreende duas tarefas principais: identificação das lesões e classificação das lesões de acordo com sua etiologia, isquêmica ou desmielinizante. Dessa forma, os experimentos conduzidos tiveram por objetivo comparar a acurácia dos classificadores quando aplicados a uma das tarefas propostas e descritas a seguir:

- Identificação de Lesões (tarefa Id): distinguir as regiões de interesse de acordo com duas classes: lesão ou não lesão.
- Classificação de Lesões (tarefa Cl): distinguir as regiões de interesse contendo lesões, de acordo com sua etiologia: lesão isquêmica ou lesão desmielinizante. Esta tarefa pressupõe que a tarefa Id tenha sido realizada.
- Identificação e Classificação de Lesões em uma etapa única (tarefa IdCl): distinguir as regiões de interesse dentre uma das seguintes classes: não lesão, lesão desmielinizante ou lesão isquêmica. Esta tarefa é independente das tarefas Id e Cl.

Todas as tarefas propostas foram realizadas pelos classificadores SVM, kNN, LDA e OPF utilizando diferentes configurações de parâmetros, de modo a se compararem as performances obtidas. Além disso, também foram realizados testes utilizando o método de fusão de classificadores *bagging*. Para o classificador SVM, por exemplo, foram testados os kernels linear e rbf, e os parâmetros foram selecionados através da técnica de *grid-search*. Também foram

simuladas diferentes funções de distância para o treinamento e teste do classificador OPF, assim como diferentes valores de  $k$  ( $k = 1$ ,  $k = 3$  e  $k = 5$ ) para o classificador kNN. Também foram aplicados diferentes métodos de seleção de atributos, como PCA e árvore de decisão, de modo a buscar a melhor combinação de técnicas para a realização de cada uma dessas tarefas. Além disso, os experimentos foram realizados com e sem a etapa de pré-processamento apresentada na seção 2.1 e utilizando diferentes quantizações para o cálculo da matriz de co-ocorrência e da matriz de comprimento de corrida.

Dentre esses diferentes métodos e configurações possíveis, a Tabela 3.1 mostra os resultados obtidos pela melhor combinação de técnicas na solução de cada uma das tarefas propostas. O classificador SVM com kernel linear com parâmetro  $C = 10$ , sem nenhum método de seleção de atributos, apresentou a melhor acurácia de 99.35% na solução da tarefa Id. O classificador LDA apresentou na execução desta tarefa acurácia de 99% quando combinado com o método de seleção de atributos PCA, enquanto o classificador kNN alcançou 98.61% utilizando duas diferentes configurações, tanto com  $k = 1$  combinado com o método de seleção de atributos da árvore de decisão, como com  $k = 3$  combinado com o método de seleção de atributos PCA. O classificador OPF com distância euclidiana e combinado com a árvore de decisão obteve a menor taxa de acerto na execução dessa tarefa: 98.28%.

Para tarefa Cl, a melhor acurácia (88.65%) foi obtida utilizando-se a fusão de classificadores através do método de *bagging*, enquanto o melhor classificador individual para essa tarefa foi o LDA combinado com o seletor de atributos da PCA alcançando uma acurácia de aproximadamente 87%. O classificador SVM obteve acurácia similar (86.76%) utilizando kernel *rbf*, com parâmetros  $C = 100$  e  $\gamma = 0.1$ . Resultados idênticos foram obtidos pelos classificadores OPF com distância euclidiana e o kNN com  $k = 3$  quando combinados ao PCA.

Por fim, utilizando o classificador SVM com kernel *rbf* e com parâmetros  $C = 10$  e  $\gamma = 0.1$ , sem método de seleção de atributos, foi obtida uma acurácia de 91.70% na execução da tarefa IdCl. O classificador LDA apresentou uma acurácia de 89.60%, enquanto o classificador kNN com  $k = 3$  obteve acurácia de 88.42% ambos combinados com o PCA. Por fim o classificador OPF obteve uma acurácia de 87.92% utilizando a distância manhattan combinado ao seletor de atributos da árvore de decisão. Os resultados indicados na tabela 3.1 foram obtidos sem a etapa de pré-processamento e utilizando a quantização de 5 bits/pixel no cálculo da matriz de co-ocorrência e da matriz de comprimento de corrida.

Tabela 3.1: Melhores acurácias obtidas para as tarefas Id, Cl e IdCl

Tarefa	SVM	kNN	OPF	LDA	<i>Bagging</i>
Tarefa Id	<b>99.35±0.02</b>	98.61±0.02	98.28±0.01	99.00±0.01	98.96±0.03
Tarefa Cl	86.76±0.04	83.65±0.02	83.65±0.04	86.91±0.06	<b>88.65±0.06</b>
Tarefa IdCl	<b>91.70±0.04</b>	88.42±0.04	87.92±0.04	89.60±0.04	90.60±0.02

Além do estudo comparativo das acurácias, analisamos também as matrizes de confusão obtidas em cada uma das tarefas por diferentes combinações de métodos de seleção de

atributo e de classificadores. As Tabelas 3.2 e 3.3 apresentam as matrizes de confusão dos classificadores que atingiram a melhor e a pior acurácia para a tarefa Id, respectivamente. A matriz de confusão do classificador SVM (Tabela 3.2) apresenta apenas duas ROIs classificadas erroneamente, enquanto a matriz de confusão do classificador OPF combinado com a árvore de decisão (Tabela 3.3), indica que apenas cinco amostras são classificadas erroneamente, três ROIs de lesão foram classificadas como não lesão, enquanto duas ROIs foram classificadas como lesão, mas representavam regiões não lesionadas. Esta foi a tarefa cuja solução apresentou a melhor acurácia dentre as tarefas propostas, e como pode visto através das matrizes de confusão apresentadas, a diferença entre o melhor e o pior classificador não é representativa.

Tabela 3.2: Matriz de confusão normalizada da tarefa de identificação de lesões realizada pelo classificador SVM

	Lesão	Não Lesão
Lesão	0.7244 (205/283)	0.007 (2/283)
Não Lesão	0 (0/283)	0.2686 (76/283)

Tabela 3.3: Matriz de confusão normalizada da tarefa de identificação de lesões realizada pelo classificador OPF

	Lesão	Não Lesão
Lesão	0.7244 (205/283)	0.0106 (3/283)
Não Lesão	0.007 (2/283)	0.258 (73/283)

As Tabelas 3.4 e 3.5 apresentam, respectivamente, as matrizes de confusão dos classificadores LDA e OPF com distância *manhattan*, quando combinados ao método de seleção de atributos PCA na execução da tarefa Cl. A Tabela 3.4, que representa o melhor resultado obtido para a tarefa Cl, indica que os erros do classificador LDA ocorrem de maneira quase igualitária entre as classes, de forma que a quantidade de lesões desmielinizantes classificadas como isquêmicas é similar à quantidade de lesões isquêmicas classificadas como desmielinizantes. A matriz de confusão do classificador OPF ilustrada na Tabela 3.5, por outro lado, apresenta uma diferente distribuição de erros, de forma que existe uma maior concentração de ROIs de lesões desmielinizantes classificadas como isquêmicas, do que o contrário.

Tabela 3.4: Matriz de confusão da tarefa de classificação de lesões realizada pelo classificador LDA

	Lesão Desmielinizante	Lesão Isquêmica
Lesão Desmielinizante	0.6328 (131/207)	0.058 (12/207)
Lesão Isquêmica	0.0773 (16/207)	0.2319 (48/207)

Por fim, a Tabela 3.6 apresenta a matriz de confusão do classificador SVM na execução da tarefa IdCl, sem a utilização de métodos de seleção de atributos. Esta matriz de confusão

Tabela 3.5: Matriz de confusão da tarefa de classificação de lesões realizada OPF utilizando distância manhattan

	Lesão Desmielinizante	Lesão Isquêmica
Lesão Desmielinizante	0.6377 (132/207)	0.0531 (11/207)
Lesão Isquêmica	0.1111 (23/207)	0.1981 (41/207)

apresenta uma maior concentração dos erros entre as classes de lesão desmielinizante e lesão isquêmica, indicando que a tarefa de classificação de lesões é mais complexa do que a tarefa de identificação de lesões.

Tabela 3.6: Matriz de confusão das tarefas de identificação e classificação de lesões em uma única etapa realizadas pelo classificador SVM

	Não lesão	Lesão Desmielinizante	Lesão Isquêmica
Não Lesão	0.2685 (76/283)	0 (0/283)	0 (0/283)
Lesão Desmielinizante	0.0035 (1/283)	0.4700 (133/283)	0.0318 (9/283)
Lesão Isquêmica	0.0035 (1/283)	0.0459 (13/283)	0.1768 (50/283)

A combinação de técnicas de processamento de imagens e de reconhecimento de padrões apresentou uma alta acurácia na solução da tarefa Id, indicando que os atributos de textura extraídos das ROIs são representativos e tornam a etapa de classificação de padrões mais fácil de ser resolvida. O classificador com o pior resultado na solução desta tarefa apresentou uma acurácia de 98.28%, cerca de apenas 1% menor do que a acurácia obtida pelo melhor classificador. Dessa forma, em um sistema real de identificação de lesões baseado em atributos de textura, qualquer um dos métodos de classificação propostos seria indicado, pois todos apresentaram altas taxas de acurácia. Neste caso, a escolha de qual método utilizar deve considerar outros critérios, como complexidade de implementação do algoritmo, tempo de processamento, entre outros.

A tarefa Cl, por outro lado, mesmo apresentando uma taxa de acurácia aceitável, foi a que obteve o pior resultado dentre as tarefas propostas, sugerindo que é mais fácil distinguir tecido normal de tecido lesionado, do que classificar corretamente diferentes tipos de lesões de acordo com sua etiologia. A diferença de valor de acurácia entre o melhor e o pior classificador foi de 3%, de forma que o melhor classificador indicado para essa tarefa é o LDA combinado ao método de seleção de atributos PCA. O classificador SVM, por sua vez, apresenta acurácia similar (menos de 0.2% inferior a acurácia da LDA), e também possui uma grande variedade de implementações e documentação disponíveis na literatura, sendo portanto tão indicado quando o LDA para a solução tarefa Cl.

Os resultados também confirmaram que os classificadores OPF e 1NN apresentam acurácias similares (Souza et al., 2012). Foi verificada na tarefa IdCl, por exemplo, uma acurácia de 84.10% para o classificador OPF e 84.46% para o classificador 1NN quando utilizando a distância euclidiana e o método de seleção de atributos PCA. Este resultado confirma a similaridade entre os classificadores, sendo a acurácia do 1NN ligeiramente superior. Modificando

a distância utilizada pelo classificador OPF para a distância manhattan a acurácia chega a 86.92%, indicando que para a tarefa realizada, o uso desta distância tornou o classificador mais discriminante. O mesmo teste realizado no classificador 1NN provavelmente deve gerar resultados superiores, mas por questões de implementação, não foi possível realizá-lo.

Além disso, pode-se verificar que a integração das tarefas de identificação e classificação de lesões, proposta na tarefa IdCl, é indicada em sistemas reais de análise de lesões em substância branca no cérebro, visto que apresenta uma quantidade de erros similares a realização das tarefas Id e Cl concatenadas, mas soluciona o problema com uma abordagem mais direta e compacta.

Por fim, não há como indicar um melhor método de seleção de atributos para todas as tarefas propostas, visto que cada uma delas apresentou seus melhores resultados utilizando diferentes métodos de seleção. Podemos destacar, por outro lado, os classificadores SVM e LDA, que apresentaram as melhores acurácias na análise de lesões em substância branca, seja na tarefa de identificação, ou na tarefa de classificação. Além disso, podemos verificar que a fusão de classificadores, utilizada com o objetivo de aumentar a acurácia, pode em alguns casos, ter efeito oposto e piorar os resultados, dependendo da tarefa proposta e da correlação entre os classificadores.

## 3.2 Influência das etapas de pré-processamento, extração de atributos e seleção de atributos nos resultados

Para que fossem obtidos os resultados indicados na tabela 3.1, foram executados outros experimentos relacionados às etapas preliminares da metodologia, apresentadas na Figura 2.1, a fim de avaliar a relevância de cada uma dessas etapas na solução das tarefas propostas. Os resultados apresentados nas próximas sub-seções referem-se ao pré-processamento das imagens, à escolha da quantização para o cálculo da matriz de co-ocorrência e da matriz de comprimento de corrida, assim como a análise dos métodos de seleção de atributos para cada uma das tarefas. Além disso, também foi realizado um estudo do grau de importância dos atributos de textura extraídos.

### 3.2.1 Etapa de pré-processamento

O método de pré-processamento selecionado para este trabalho foi o de normalização das ROIs para a faixa de valores  $[\mu \pm 3\sigma]$ , onde  $\mu$  representa a média dos níveis de cinza, e  $\sigma$  o seu desvio-padrão e ambos são calculados dentro da ROI analisada. Este tipo de normalização modifica a distribuição dos níveis de cinza dos pixels dentro da ROI e não preserva a variação relativa entre dois níveis de cinza, provocando variações relevantes nos atributos de textura calculados posteriormente.

Como é possível visualizar na Tabela 3.7, os melhores resultados são obtidos quando não

Tabela 3.7: Acurácias obtidas para as tarefas de identificação de lesões (tarefa Id), classificação de lesões (tarefa Cl) e identificação e classificação de lesões em uma única etapa (tarefa IdCl) com ou sem a etapa de pré-processamento (normalização das ROIs)

Tarefa/Classificador	Sem normalização	Com normalização
Id/SVM	<b>99.35±0.01</b>	98.25±0.01
Cl/LDA	<b>86.91±0.04</b>	81.52±0.04
IdCl/SVM	<b>91.70±0.02</b>	85.17±0.02

é utilizada a normalização das ROIs, indicando que a variação realizada nos níveis de cinza dos pixels dentro das ROIs modifica os atributos de textura, e os torna menos eficientes na diferenciação entre tecido normal e tecido lesionado, assim como na distinção de diferentes tipos de lesão de acordo com sua etiologia. Este resultado pode ser explicado pelo fato que essa normalização gera redução de sensibilidade (Lerski et al., 1993), causando, em aplicações como esta, perda de informação textural relevante para o posterior processo de classificação. Este comportamento é ainda agravado pelos truncamentos realizados durante a execução da operação de normalização, visto se tratar de imagens em formato .bmp (8 bits).

### 3.2.2 Quantização das matrizes de co-ocorrência e de comprimento de corrida

A dimensão da matriz de co-ocorrência e da matriz de comprimento de corrida está intimamente relacionada à quantização da imagem analisada. As imagens quantizadas a uma taxa de 8 bits/pixel, por exemplo, apresentam 256 possíveis níveis de cinza e portanto uma matriz de co-ocorrência de 256 x 256. Em casos como este, esta matriz será repleta de zeros, visto que a maioria das combinações de nível de cinza não ocorrem. Logo os atributos estatísticos extraídos dessa matriz perdem a capacidade de representar os padrões de textura analisados, visto que a matriz de co-ocorrência aproxima uma distribuição de probabilidade conjunta de dois pixels e com essa quantidade extrema de zeros, passa a ser uma aproximação ruim (Clausi, 2002). A mesma análise pode ser verificada para a matriz de comprimento de corrida.

Como alguns atributos estatísticos extraídos dessas matrizes não são eficientes na classificação quando utilizando uma grande quantidade de níveis de cinza, foi realizado um estudo da quantização ideal para a identificação de lesões, classificação de lesões e para a identificação e classificação de lesões em uma única etapa. Para tal fim, realizamos experimentos variando a quantização, iniciando com uma taxa de 6 bits/pixel. Para fins de comparação, a taxa de quantização foi modificada para 5 bits/pixel. Como a variação de 6 bits/pixel para 5 bits/pixel gerou ganho na acurácia, também foram realizados experimentos utilizando a taxa de 4 bits/pixel. Dessa forma, como pode ser visto na Tabela 3.8, a taxa de quantização que gera os melhores resultados é a de 5 bits/pixel.

Este resultado indica que quantizações menores do que 5 bits/pixel geram perdas de informações relevantes para a correta diferenciação dos padrões de textura analisados, enquanto

Tabela 3.8: Acurácias obtidas para as tarefas Id, Cl e IdCl variando a quantização utilizada para o cálculo da matriz de co-ocorrência e da matriz de comprimento de corrida

Tarefa/Classificador	4bits/pixel	5bits/pixel	6bits/pixel
Id/SVM	98.28±0.01	<b>99.35±0.01</b>	98.64±0.02
Cl/LDA	86.28±0.02	<b>86.91±0.04</b>	82.41±0.04
IdCl/SVM	90.63±0.04	<b>91.70±0.04</b>	89.92±0.04

quantizações maiores geram atributos de textura com menos capacidade de representação destes padrões de textura.

### 3.2.3 Seleção de atributos

Neste trabalho, foram realizados experimentos utilizando dois diferentes métodos de seleção de atributos: árvore de decisão e PCA. A Tabela 3.9 indica os valores obtidos para os melhores classificadores de uma das tarefas propostas utilizando um dos métodos de seleção de atributos, ou sem utilizar método algum de seleção. Os resultados indicam que o uso de um método de seleção de atributos não aumenta a acurácia do SVM. Este comportamento pode ser explicado pelo fato de que SVM atribui pesos aos atributos mais relevantes, de forma que os atributos menos relevantes têm menos influência no classificador gerado.

Tabela 3.9: Acurácias obtidas nas tarefas Id, Cl e IdCl quando variando o método de seleção de atributos e utilizando o melhor método de classificação para cada tarefa

Tarefa/Classificador	Sem Seleção	Árvore de Decisão	PCA
Id/SVM	<b>99.35±0.01</b>	98.96±0.01	99.35±0.01
Cl/LDA	80.52±0.04	81.52±0.04	<b>86.91±0.06</b>
IdCl/SVM	<b>91.70±0.04</b>	88.74±0.07	90.27±0.03

Na tarefa IdCl, por sua vez, apresentada na Tabela 3.10, a melhor acurácia foi obtida sem que nenhum método de seleção de atributos fosse aplicado, mas o resultado apresentado pelo PCA, em geral, é melhor do que o alcançado pela árvore de decisão.

Tabela 3.10: Acurácias obtidas na tarefas IdCl quando variando o método de seleção de atributos

Classificador	Sem Seleção	Árvore de Decisão	PCA
SVM	<b>91.70±0.05</b>	88.74±0.06	90.27±0.05
3NN	88.06±0.06	86.89±0.05	<b>88.42±0.05</b>
OPF	85.95±0.04	<b>87.92±0.04</b>	86.92±0.04
LDA	87.09±0.04	87.78±0.05	<b>89.60±0.03</b>

Por fim, não há como indicar o melhor método de seleção de atributos pois os resultados variam de acordo com o classificador utilizado e com a tarefa executada. Os resultados da

tarefa IdCl, por exemplo, indicam que o classificador OPF alcança a melhor acurácia com a árvore de decisão, enquanto o classificador kNN com  $k = 3$  e o classificador LDA apresentam os melhores resultados utilizando o seletor de atributos PCA, não sendo possível portanto indicar o melhor método de seleção de atributos.

Além de comparar os diferentes métodos de seleção de atributos, também realizamos experimentos com a finalidade de analisar com maior profundidade os atributos de textura extraídos das regiões de interesse. Para tal fim, fizemos um estudo do grau de importância dos atributos. O grau de importância pode ser computado através da análise do erro de classificação gerado por aquele atributo em específico. Quanto mais discriminativo for o atributo, maior será seu grau de importância.

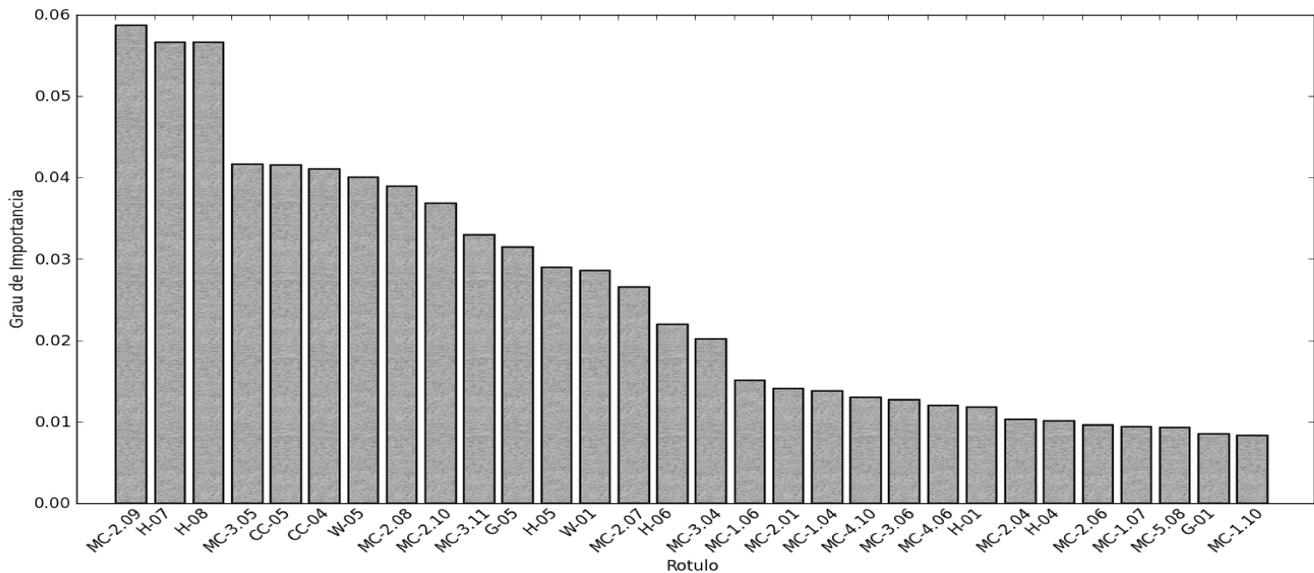


Figura 3.1: Representação decrescente do grau de importância dos atributos de textura na execução da tarefa Id gerado através do método da árvore de decisão. A Figura 3.4 relaciona o identificador do atributo, eixo x do gráfico, com o seu nome

As Figuras 3.1, 3.2 e 3.3 apresentam gráficos contendo os atributos de textura em ordem decrescente de grau de importância para as tarefas Id, Cl e IdCl respectivamente. É importante salientar que o conjunto de atributos de textura apresenta uma quantidade total de 87 atributos, mas apenas os 30 atributos mais relevantes são apresentados nos gráficos, a fim de facilitar a sua visualização. Por fim, a Figura 3.4 relaciona o identificador dos atributos (indicado no eixo x dos gráficos) com o seu nome.

Em todas as figuras é possível identificar um grupo de dois ou três atributos que se destacam visto seus altos graus de importância. No caso da tarefa Id, correspondente à Figura 3.1, 3 atributos são evidenciados: entropia extraída matriz de co-ocorrência (rótulo MC-2.09), e os percentis do histograma em 50% (rótulo H-06) e 90% (rótulo H-07). Já para a tarefa Cl, apresentada na Figura 3.2, apenas dois atributos foram destacados, cujos rótulos são MC-4.06, MC-3.07 e MC-3.08, todos extraídos da matriz de co-ocorrência. Por fim, a

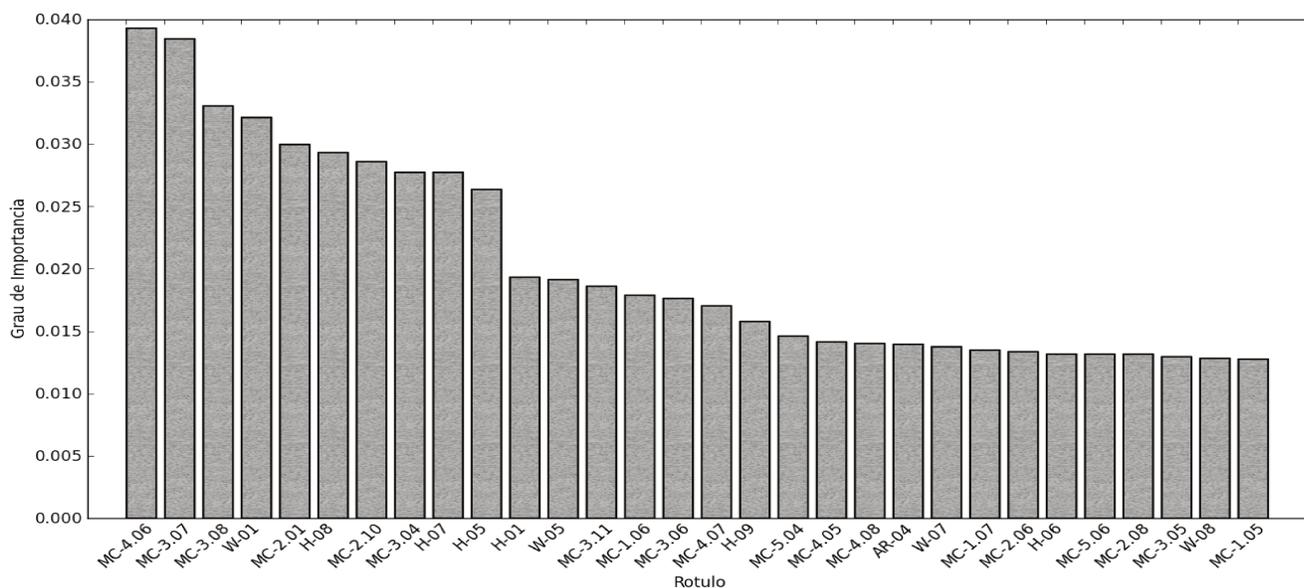


Figura 3.2: Representação decrescente do grau de importância dos atributos de textura na execução da tarefa Cl gerado através do método da árvore de decisão. A Figura 3.4 relaciona o identificador do atributo, eixo x do gráfico, com o seu nome

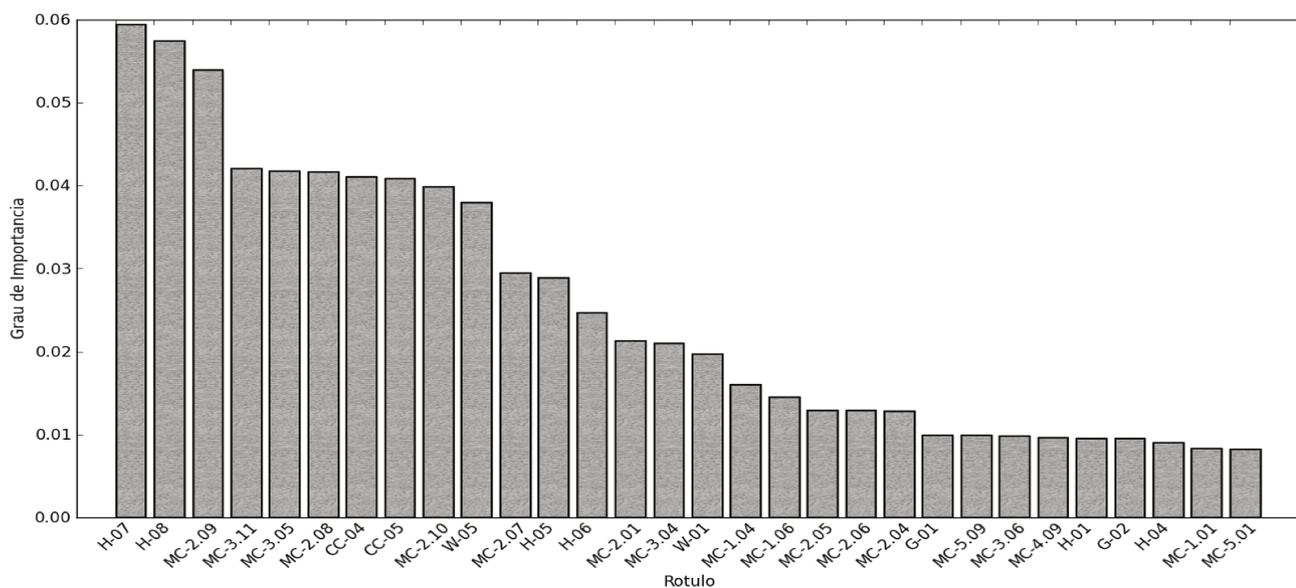


Figura 3.3: Representação decrescente do grau de importância dos atributos de textura na execução da tarefa IdCl gerado através do método da árvore de decisão. A Figura 3.4 relaciona o identificador do atributo, eixo x do gráfico, com o seu nome

tarefa IdCl (Figura 3.2) apresenta comportamento similar ao da tarefa Id, com os mesmos três atributos (rótulos H-06, H-07 e MC-2.09) em destaque, mas em ordem de grau de importância diferente.

Os gráficos 3.1, 3.2 e 3.3 indicam que os atributos extraídos da matriz de co-ocorrência destacam-se em todas as tarefas, indicando a sua relevância na identificação e classificação das lesões em substância branca. O segundo grupo de atributos que também apresenta um alto grau de relevância são os extraídos do histograma das regiões de interesse.

### 3.3 Fusão de Classificadores

A aplicação da fusão de classificadores teve por objetivo obter melhores acurácias através da combinação dos resultados gerados por cada um dos classificadores utilizando o método do *bagging*, no qual o resultado final foi gerado através de uma votação igualitaria entre os classificadores, e da fusão por SVM, na qual os resultados dos classificadores formam um conjunto de dados, utilizados para treinar e testar um classificador SVM. Além disso, os experimentos foram realizados tanto através da fusão de todos os classificadores, como através da utilização apenas daqueles selecionados através das métricas de diversidade.

Os experimentos da fusão de classificadores foram centrados na tarefa Cl, visto se tratar da tarefa mais complexa de ser solucionada dentre as propostas. A métrica de diversidade de dupla falha (DFM) indicou o par de classificadores LDA e 1NN como os mais diversos. A estatística-Q (QSTAT) apresentou como classificadores mais diversos o OPF e o LDA, o acordo iterativo (IA), por sua vez, indicou como mais diverso o par de classificadores LDA e SVM. Podemos verificar, portanto, que todos os classificadores foram apontados por pelo menos uma métrica como o mais diverso. O classificador LDA, por sua vez, destacou-se por apresentar alto grau de diversidade em todas as métricas.

Os resultados gerados são apresentados na tabela 3.11 e indicam que a fusão dos classificadores atinge a acurácia máxima de 88.65% utilizando os classificadores SVM, LDA e OPF com a sua melhor configuração individual de parâmetros e o método do *bagging*. Além disso, apresentamos resultados da fusão por *bagging* com todos os classificadores e outras possíveis combinações de classificadores utilizando os atributos de diversidade.

Tabela 3.11: Acurácias obtidas na tarefa Cl quando realizando a fusão dos classificadores utilizando uma das técnicas variando o método de seleção de atributos e utilizando o melhor método de classificação para cada tarefa

Método de Fusão	LDA	OPF	1NN	SVM	Acurácia
<i>Bagging</i>	x	x	x	x	88.01±0.06
SVM	x	x	x	x	85.99±0.03
<i>Bagging</i>	x	x	o	x	<b>88.65±0.08</b>
<i>Bagging</i>	x	o	x	x	88.14±0.08
<i>Bagging</i>	x	x	x	o	84.15±0.08

A fusão com o SVM gerou resultados inferiores aos apresentados pelo *bagging*, possivelmente devido ao fato da quantidade limitada de classificadores utilizados. Além disso, é importante salientar que mesmo combinando classificadores diversos, pode haver queda no

valor da acurácia, como no caso da combinação do LDA, OPF e 1NN, onde a fusão apresentou uma acurácia menor do que os classificadores isolados. Por outro lado, no caso da fusão entre os classificadores LDA, OPF e SVM, foi gerada uma acurácia maior do que a dos classificadores individualmente, motivando o uso da fusão). A combinação dos classificadores LDA, OPF e SVM também apresentou uma acurácia um pouco maior do que a combinação com todos os classificadores, indicando que a inclusão do 1NN degrada a performance da fusão. Isto provavelmente se deve ao fato que, por se tratar de voto majoritário, a inclusão de um quarto classificador (número par de classificadores) pode gerar decisões inconsistentes.

### 3.4 Considerações Finais

Os resultados obtidos indicam que a combinação de técnicas de processamento de imagens e de reconhecimento de padrões é adequada para a análise de lesões em substância branca, representada pelas tarefas de identificação e classificação das lesões. Dentre as tarefas propostas, a de classificação (tarefa Cl) revelou-se como a tarefa mais complexa de ser resolvida, indicando que é mais fácil distinguir ROIs contendo tecido normal de ROIs contendo lesões, do que distinguir corretamente diferentes tipos de lesões de acordo com sua etiologia.

Cada uma das etapas adotadas na metodologia deste trabalho foram analisadas neste capítulo, iniciando pelo pré-processamento. A normalização das ROIs gerou uma queda na performance dos classificadores, indicando que os atributos de textura calculados sobre as ROIs normalizadas são menos eficientes na diferenciação dos tipos de tecidos, de forma que esta etapa foi descartada. A etapa de extração dos atributos de textura, por sua vez, fez necessária a escolha da quantização das imagens para o cálculo da matriz de co-ocorrência e da matriz de comprimento de corrida. Para tal fim, realizamos experimentos variando esta quantização e os resultados obtidos apontaram como a melhor taxa de quantização a de 5 bits/pixel.

A etapa seguinte de seleção de atributos, produziu ganhos no desempenho de todos os classificadores, exceto o SVM, que é intrinsecamente menos influenciado por atributos irrelevantes. Além disso, dentre os classificadores utilizados, podem ser descartados o SVM e o LDA por apresentar as melhores acurácias na execução das tarefas propostas. Por fim, o estudo do grau de importância dos atributos apontou os atributos extraídos da matriz de co-ocorrência e do histograma como os mais relevantes na identificação e classificação de lesões na substância branca do cérebro.

Rótulo	Nome	Rótulo	Nome	Rótulo	Nome	Rótulo	Nome
H-01	MédiaH	MC-2.03	Correlação d=2	MC-4.03	Correlação d=4	CC-03	Enfase em Corrida Longa
H-02	VariânciaH	MC-2.04	Soma dos quadrados d = 2	MC-4.04	Soma dos quadrados d = 4	CC-04	Enfase em Corrida Curta
H-03	AssimetriaH	MC-2.05	Homogeneidade d=2	MC-4.05	Homogeneidade d=4	CC-05	Porcentagem de Corrida
H-04	CurtoseH	MC-2.06	Soma média d=2	MC-4.06	Soma média d=4	G-01	MédiaG
H-05	Percentil 01%	MC-2.07	Soma variância d = 2	MC-4.07	Soma variância d = 4	G-02	VariânciaG
H-06	Percentil 10%	MC-2.08	Soma entropia d=2	MC-4.08	Soma entropia d=4	G-03	AssimetriaG
H-07	Percentil 50%	MC-2.09	Entropia d = 2	MC-4.09	Entropia d = 4	G-04	CurtoseG
H-08	Percentil 90%	MC-2.10	Variância Diferencial d = 2	MC-4.10	Variância Diferencial d = 4	G-05	Porcentagem de Pixels com Gradiente não-nulo
H-09	Percentil 99%	MC-2.11	Entropia diferencial d = 2	MC-4.11	Entropia diferencial d = 4	AR-1	$\theta_1$
MC-1.01	Energia d = 1	MC-3.01	Energia d = 3	MC-5.01	Energia d = 5	AR-2	$\theta_2$
MC-1.02	Contraste d=1	MC-3.02	Contraste d=3	MC-5.02	Contraste d=5	AR-3	$\theta_3$
MC-1.03	Correlação d=1	MC-3.03	Correlação d=3	MC-5.03	Correlação d=5	AR-4	$\theta_4$
MC-1.04	Soma dos quadrados d = 1	MC-3.04	Soma dos quadrados d = 3	MC-5.04	Soma dos quadrados d = 5	AR-5	$\sigma$
MC-1.05	Homogeneidade d=1	MC-3.05	Homogeneidade d=3	MC-5.05	Homogeneidade d=5	W-01	Energia LL d=1
MC-1.06	Soma média d=1	MC-3.06	Soma média d=3	MC-5.06	Soma média d=5	W-02	Energia LH d=1
MC-1.07	Soma variância d = 1	MC-3.07	Soma variância d = 3	MC-5.07	Soma variância d = 5	W-03	Energia HL d=1
MC-1.08	Soma entropia d=1	MC-3.08	Soma entropia d = 3	MC-5.08	Soma entropia d=5	W-04	Energia HH d=1
MC-1.09	Entropia d = 1	MC-3.09	Entropia d = 3	MC-5.09	Entropia d = 5	W-05	Energia LL d=2
MC-1.10	Variância Diferencial d = 1	MC-3.10	Variância Diferencial d = 3	MC-5.10	Variância Diferencial d = 5	W-06	Energia LH d=2
MC-1.11	Entropia diferencial d = 1	MC-3.11	Entropia diferencial d = 3	MC-5.11	Entropia diferencial d = 5	W-07	Energia HL d=2
MC-2.01	Energia d = 2	MC-4.01	Energia d = 4	CC-01	Não-uniformidade Comprimento de Corrida	W-08	Energia HH d=2
MC-2.02	Contraste d=2	MC-4.02	Contraste d=4	CC-02	Não-uniformidade Nível de Cinza		

Figura 3.4: Lista com os rótulos e os nomes correspondentes dos atributos de textura utilizados e extraídos nos seguintes grupos: momentos do histograma (M), matriz de co-ocorrência (MC), matriz de comprimento de corrida (CC), gradiente (G), autoregressivos (AR) e wavelet(W)

## Conclusões e Perspectivas

Neste trabalho, foi proposta uma metodologia de análise de lesões em substância branca, visto a importância dessa tarefa no diagnóstico e acompanhamento de pacientes neurológicos, assintomáticos e sintomáticos. A análise dessas lesões nas imagens de ressonância magnética, no entanto, é uma tarefa não-trivial, o que fez necessária a combinação de técnicas de diferentes áreas, como processamento de imagens e reconhecimento de padrões.

Dessa forma, desenvolvemos uma metodologia para distinção entre ROIs de tecidos saudáveis e lesionados, tarefa chama de identificação de lesões, e também para a classificação dos diferentes tipos de lesão de acordo com sua etiologia, isquêmica ou desmielinizante, tarefa chamada de classificação de lesões. Além de desenvolver uma metodologia adequada para a solução do problema em questão, também foi realizada uma análise crítica da influência de cada uma das etapas utilizadas na acurácia final de cada uma das tarefas propostas.

Os resultados apontam que é possível realizar a identificação e classificação das lesões através da combinação da análise da textura com a utilização de métodos de classificação de padrões. Dentre os atributos de textura utilizados, aqueles extraídos da matriz de co-ocorrência e do histograma mostraram os maiores graus de importância, produzindo os menores erros de classificação. Além disso, pôde ser verificado que a inserção de uma etapa de seleção de atributos gerou melhores resultados, com exceção do classificador SVM, com aumentos de acurácia na faixa de valores entre 1% a 4%. Foram realizados ainda experimentos inserindo uma etapa de pré-processamento através da normalização nas regiões de interesse segmentadas, mas os resultados indicaram que, para esta aplicação específica, não deve ser utilizada essa etapa de pré-processamento, dada a perda de sensibilidade, e questões de truncamento, por se tratar de uma imagem em formato bmp.

Dentre as tarefas propostas, a de identificação de lesões apresentou os melhores resultados, indicando que é mais fácil distinguir tecidos lesionados de tecidos saudáveis do que diferenciar diferentes tipos de lesão de acordo com sua etiologia. O melhor resultado da tarefa de classificação de lesões (88.65%), que foi alcançado através da utilização do método de fusão de classificadores através da utilização do *bagging*, representa um valor aceitável no ambiente clínico.

Além disso, não há como indicar um melhor método de seleção de atributos dado que dependendo da tarefa proposta e do classificador utilizada, melhores resultados foram obtidos com diferentes métodos de seleção. Por outro lado, podemos destacar os classificadores SVM e LDA, por apresentar as melhores acurácias na análise de lesões em substância branca, seja na tarefa de identificação, ou na tarefa de classificação. Em sistemas reais de análise de lesões, além da acurácia obtida, a escolha de qual classificador deve ser utilizado deve considerar outras questões, como disponibilidade de documentação e dificuldade de implementação.

Por fim, podemos concluir que o uso de atributos de textura e dos classificadores SVM, kNN, OPF e LDA é adequado para a análise de lesões, visto que a metodologia proposta alcançou taxas de acurácia aceitáveis e tempos de execução não proibitivos no ambiente clínico, além de ter lidado de maneira eficiente com lesões de diferentes etiologias, com diferentes tamanhos e formatos.

## Perspectivas

Este trabalho será continuado como tema de doutorado e terá como objetivo a identificação automática de lesões causadas por diferentes doenças na substância branca do cérebro em imagens volumétricas de ressonância magnética de várias modalidades. Este projeto visa minimizar o trabalho manual dos especialistas, tornando a análise de lesões em substância branca menos subjetiva.

Além disso, serão propostas medidas quantitativas das lesões para análise de sua severidade e seu acompanhamento da evolução ao longo do tempo (estudos longitudinais). O método a ser desenvolvido deve ser generalista o suficiente para analisar corretamente lesões de diferentes etiologias, com tamanhos e formas variáveis.

## Publicações

Os resultados obtidos neste trabalho de mestrado possibilitaram publicações em eventos na área de imagens médicas. O trabalho intitulado *Analysis of Brain White Matter Hyperintensities using Pattern Recognition Techniques* (Bento et al., 2013) foi submetido e aceito no evento *SPIE Medical Imaging 2013* e indica os resultados finais obtidos e apresentados na dissertação.

Outro artigo, que foi apresentado em 2012 no evento *XXV SIBGRAPI - Conference on Graphics, Patterns and Images*, sob o título *Analysis of Brain White Matter Hyperintensities* (Bento et al., 2012), apresenta uma comparação entre os classificadores utilizados de acordo com as acurácias obtidas e os seus tempos de processamentos, além de uma análise da acurácia quando variando a quantidade de atributos utilizados. Neste artigo, ainda não tinha sido aplicada a etapa de seleção de atributos.

O artigo *Support vector machines classification of texture parameters of white matter lesions in Systemic Lupus Erythematosus. Possible mechanism to distinguish between demyelination and ischemia* (Bento et al., 2011), apresentado em 2011 no evento da área médica

---

*Annual Scientific Meeting /Arthritis and Rheumatism* , mostra um estudo inicial da análise das lesões em substância branca do cérebro, limitado ao uso do classificador SVM.

# Bibliografia

- P. Anbeek, K. L. Vincken, M. J. P. Osch, R. H. C. Bisschops, and J. Grond. Probabilistic segmentation of whitematter lesions in mr imaging. *NeuroImage*, 21(3):1037–1044, 2004.
- S. Appenzeller, A. V. Faria, L. M. Li, L. T. Costallat, and F. Cendes. Quantitative magnetic resonance imaging analyses and clinical significance of hyperintense white matter lesions in systemic lupus erythematosus patients. *Annals of Neurology*, 64(6):635–643, 2008.
- M. Bento, L. Rittner, R. Lotufo, and S. Appenzeller. Support Vector Machines Classification of Texture Parameters of White Matter Lesions in Systemic Lupus Erythematosus. Possible Mechanism to Distinguish Between Demyelination and Ischemia. In *Annual Scientific Meeting /Arthritis and Rheumatism*, Chicago, EUA, 2011.
- M. Bento, L. Rittner, R. Lotufo, and S. Appenzeller. Analysis of Brain White Matter Hyperintensities. In *Workshop of Works in Progress (WIP) in SIBGRAPI 2012 (XXV Conference on Graphics, Patterns and Images)*, pages 3–4, MG, BR, 2012.
- M. Bento, L. Rittner, S. Appenzeller, A. Lapa, and R. Lotufo. Analysis of Brain White Matter Hyperintensities using Pattern Recognition Techniques. In *SPIE Medical Imaging*, pages 1–7, Flórida, USA, 2013.
- J. Bergstra and Y. Bengio. Random search for hyper-parameter optimization. *Journal Machine Learning Research*, 13(1):281–305, 2012.
- N. Bhatia. Survey of Nearest Neighbor Techniques. *International Journal of Computer Science and Information Security*, 8(2):302–305, 2010.
- B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 927–929, NY, USA, 1992.
- G. Castellano, L. Bonilha, and F. Cendes. Texture analysis of medical images. *Clinical Radiology*, 59(12):1061–1069, 2004.

- C. C. Chang and C. J. Lin. Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):1 – 27, 2011.
- D. A. Clausi. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of Remote Sensing*, 28(1):45–62, 2002.
- C. A. Cocosco, A. P. Zijdenbos, and A. C. Evans. A fully automatic and robust brain mri tissue classification method. *Medical Image Analysis*, 7(4):513–527, 2003.
- G. Collewet, M. Strzelecki, and F. Mariette. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magnetic resonance imaging*, 22(1):81–91, 2004.
- A. Compston and A. Coles. Multiple sclerosis. *The Lancet*, 372(9648):1502 – 1517, 2008.
- C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.
- T. M. Cover and P. E. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21 – 27, 1967.
- Dionísio Doering. O uso da transformada Wavelet Bi-Dimensional no conceito do Espaço de Escala. Mestrado em engenharia elétrica, Universidade Federal do Rio Grande do Sul, 2005.
- R. O. Duda, P. E. Hart, and D. H. Stork. *Pattern Classification*. Wiley Interscience, 2000.
- A. S. Dudani. The distance-weighted k-nearest-neighbor rule. *IEEE Transactions on Man and Cybernetics*, 6(4):325–327, 1976.
- A. X. Falcão, J. Stolfi, and R. A. Lotufo. The image foresting transform: Theory, algorithms, and applications. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 26(1):19 – 29, 2004.
- F. A. Faria, J. A. Santos, A. Rocha, and R. S. Torres. Automatic classifier fusion for produce recognition. In *XXV SIBGRAPI - Conference on Graphics, Patterns and Images*, pages 252 – 259, MG, BR, 2012.
- F. Fazekas, F. Barkhof, M. Filippi, R. I. Grossman, D. K. Li, W. I. McDonald, H. F. Mcfarland, D. W. Paty, J. H. Simon, J. S. Wolinsky, and D. H. Miller. The contribution of magnetic resonance imaging to the diagnosis of multiple sclerosis. *Neurology*, 53(3):448 – 456, 1999.
- R. A. Fisher. The statistical utilization of multiple measurements. *Annals of human genetics*, 8(4):376 – 386, 2011.

- P. A. Freeborough and N. C. Fox. Mr image texture analysis applied to the diagnosis and tracking of alzheimer's disease. *IEEE Transactions on Medical Imaging*, 11(6):873–887, 1998.
- K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 1991.
- M. D. Ginsberg. Neuroprotection for ischemic stroke: Past, present and future. *Neuropharmacology*, 55(3):363 – 389, 2008.
- J. O. Glass, E. N. Cook, T. D. Elkin, and R. J. Deaton. Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks. *IEEE Transactions on Medical Imaging*, 16(9):911–918, 1997.
- J. Han and M. Kamber. *Data Mining: Concepts and Techniques*. Elsevier, 2006.
- R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(6):610–621, 1973.
- A. L. Hunt, W. Orrison, R. A. Yeo, K. Y. Haaland, R. L. Rhyne, P. J. Garry, and G. A. Rosenberg. Clinical significance of MRI white matter lesions in the elderly. *Neurology*, 39(11):1470–1474, 1989.
- K. W. Kim, J. R. MacFall, and M. E. Payne. Classification of white matter lesions on magnetic resonance imaging in the elderly. *Biological Psychiatry*, 64(4):273–280, 2008.
- S. Klöppel, A. Abdulkadir, S. Hadjideometriou, S. Issleib, L. Frings, T. N. Thanh, I. Mader, S. J. Teipel, M. Hüll, and O. Ronneberger. A comparison of different automated methods for the detection of white matter lesions in MRI data. *NeuroImage*, 57(2):416–422, 2011.
- R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *14th international joint conference on Artificial intelligence*, pages 1137–1143, CA, USA, 1995.
- V. A. Kovalev, F. Kruggel, H. J. Gertz, and D. Y. Cramon. 3d texture analysis of mri brain datasets. *IEEE Transactions Medical Imaging*, 20(5):424–433, 2001.
- F. Kruggel, J. S. Paul, and H. J. Gertz. Texture-based segmentation of diffuse lesions of the brain's white matter. *NeuroImage*, 39(3):987–996, 2008.
- L. I. Kuncheva and C. J. Whitaker. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine Learning*, 51(2):181–207, 2003.
- A. Lapa, M. Bento, L. Rittner, G. Castellano, H. Ruocco, B. Damasceno, L. Costallat, R. Lotufo, F. Cendes, and S. Appenzeller. Support vector machines classification of texture parameters of white matter lesions in systemic lupus erythematosus. possible mechanism to distinguish between demyelination and ischemia. In *Annual Scientific Meeting of the American College of Rheumatology*, page 2257, Chicago, USA, 2011.

- R. A. Lerski, K. Straughan, L. R. Schad, D. Boyce, S. Blüml, and I. Zuna. Mr image texture analysis - an approach to tissue characterization. *Magnetic Resonance Imaging*, 11(6): 873–887, 1993.
- R. Lotufo, R. Machado, A. Körbes, and R. Ramos. Adessowiki: on-line collaborative scientific programming platform. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, pages 10:1–10:6, NY, EUA, 2009.
- E. L. Machado. Um estudo de limpeza em base de dados desbalanceada com sobreposição de classes. Mestrado em computação, Universidade de Brasília, 2007.
- F. Maes, D. Vandermeulen, and P. Suetens. Automated model-based tissue classification of mr images of the brain. *IEEE Transactions on Medical Imaging*, 18(10):897–908, 1999.
- B. Ovbiagele and J. L. Saver. Cerebral White Matter Hyperintensities on MRI: Current Concepts and Therapeutic Implications. *Cerebrovascular Diseases*, 22(2):83–90, 2006.
- J. P. Papa, A. X. Falcão, and C. T. N. Suzuki. Supervised pattern classification based on optimum-path forest. *International Journal of Imaging Systems and Technology*, 19(2): 120–131, 2009.
- K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(6):559–572, 1901.
- M. Pechenizkiy, A. Tsymbal, and S. Puuronen. Pca-based feature transformation for classification: Issues in medical diagnostics. In *IEEE Symposium on Computer-Based Medical Systems*, pages 535–540, MD, USA, 2004.
- W. R. Schwartz. Segmentação de Imagens baseada em Dependência Espacial utilizando Campo Aleatório de Markov Associado com Características de Texturas. Mestrado em informática, Universidade Federal do Paraná, 2005.
- W. R. Schwartz, F. R. Siqueira, and H. Pedrini. *Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações*. Thomson Learning Ltda., 2008.
- W. R. Schwartz, F. R. Siqueira, and H. Pedrini. Evaluation of Feature Descriptors for Texture Classification. *Journal of Electronic Imaging*, 21(2):1–17, 2012.
- J. Shawe-Taylor and N. Cristianini. *Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, 2000.
- F. Y. Shih. *Image Processing and Pattern Recognition*. John Wiley & Sons, 2010.
- L. I. Smith. A tutorial on principal components analysis. Technical report, Universidade de Otago, Nova Zelandia, 2002.

- R. Souza, L. Rittner, and R. Lotufo. A Comparison between Optimum-Path Forest and k-Nearest Neighbors Classifiers. In *XXV SIBGRAPI - Conference on Graphics, Patterns and Images*, pages 260–267, MG, BR, 2012.
- A. M. Tiehuis, K. L. Vincken, W. P. T. M. Mali, L. J. Kappelle, P. Anbeek, A. Algra, and G. J. Biessels. Automated and Visual Scoring Methods of Cerebral White Matter Hyperintensities: Relation with Age and Cognitive Function. *Cerebrovascular Diseases*, 25(1):59–66, 2008.
- S. Tulyakov, S. Jaeger, V. Govindaraju, and D. Doermann. Review of Classifier Combination Methods. *Studies in Computational Intelligence*, 90(1):361–386, 2008.
- J. Walecki, S. Sierakowski, A. Lewszuk, A. Sulik, E. Tarasów, and U. Ebkowska. MR in neurological syndromes of connective tissue diseases. *Medical Science Monitor*, 8(6):105 – 111, 2002.
- Andrew R. Webb. *Statistical Pattern Recognition*. John Wiley & Sons, 2002.
- R. Woods and R. C. Gonzalez. *Processamento de Imagens Digitais*. Edgard Blucher, 2000.
- M. Wu, C. Rosano, M. Butters, E. Whyte, M. Nable, R. Crooks, C. C. Meltzer, C. F. Reynolds, and H. J. Aizenstein. A fully automated method for quantifying and localizing white matter hyperintensities on MR images. *Psychiatry Research: Neuroimaging*, 148(2-3):133–142, 2006.
- D. G. Zimring, A. Achiron, S. Miron, M. Faibel, and H. Azhari. Automatic detection and characterization of multiple sclerosis lesions in brain mr images. *Magnetic Resonance Imaging*, 16(3):311–318, 1998.