Métodos de apoio ao diagnóstico médico por imagens usando regras de associação e redes complexas

Carolina Yukari Veludo Watanabe

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura:

# Métodos de apoio ao diagnóstico médico por imagens usando regras de associação e redes complexas

Carolina Yukari Veludo Watanabe

Orientadora: Profa. Dra. Agma Juci Machado Traina

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Doutor em Ciências - Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*.

USP – São Carlos Abril de 2013

#### Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi e Seção Técnica de Informática, ICMC/USP, com os dados fornecidos pelo(a) autor(a)

| W324m | Watanabe, Carolina Yukari Veludo<br>Métodos de apoio ao diagnóstico médico por<br>imagens usando regras de associação e redes<br>complexas / Carolina Yukari Veludo Watanabe;<br>orientadora Agma Juci Machado Traina São<br>Carlos, 2013.<br>128 p. |
|-------|--|
|       | Tese (Doutorado - Programa de Pós-Graduação em<br>Ciências de Computação e Matemática Computacional)<br>Instituto de Ciências Matemáticas e de Computação,<br>Universidade de São Paulo, 2013.   |
|       | 1. mineração de imagens. 2. extração de<br>características. 3. classificador associativo. 4.<br>redes complexas. 5. CAD. I. Traina, Agma Juci<br>Machado, orient. II. Título.  |

# Agradecimentos

Em primeiro lugar agradeço à Deus, quem me deu a oportunidade e as condições necessárias para desenvolver este trabalho. Só Ele sabe por todas as coisas que passei para chegar até aqui. A Ele toda a honra, glória e louvor.

À minha orientadora Profa. Dra. Agma Juci Machado Traina, por tudo o que me ensinou e pela amizade desde a iniciação científica até o doutorado. Com você aprendi muito mais do que técnicas computacionais. Aprendi a trabalhar em grupo, a escrever artigos, a orientar alunos, a preparar aula e a tratar as pessoas como um todo, com respeito. Obrigada pelos conselhos referentes ao meio acadêmico e também aos que ultrapassaram este meio. Muito obrigada por me ajudar a me transformar na pesquisadora e professora que sou hoje.

Ao Prof. Dr. Caetano Traina Júnior, que apoiou todo o meu trabalho durante o período em que estive na USP. Sempre disposto a ajudar, revisando os artigos, discutindo métodos e resultados, e também preocupado com o auxílio financeiro.

Aos professores Dr. Antônio Carlos dos Santos, Dr. Tonicarlo Velasco, Dr. Fernando Cendes e Dra. Clarissa Yasuda que se dispuseram a me auxiliar no processo de aquisição de imagens e entendimento do problema de epilepsia no lobo temporal mesial. Muito obrigada pelo tempo dedicado, pelas imagens cedidas e pelas discussões.

Aos meus queridos tios, Maria Stela e João Paiva, que me acolheram em sua residência sempre que precisei, desde a graduação. Já no doutorado, acolheram não só a mim como também à minha filha Luiza. Obrigada por todo o carinho e cuidado conosco.

A minha mãe, sempre em constante oração, com conselhos riquíssimos, e por todo o amor. Mamis, valeu todo o nosso esforço! A melhor forma que tenho de lhe agradecer e lhe retribuir por tudo o que fez por mim é compartilhar com você as conquistas que tenho alcançado.

A minha pequena Luiza, que foi compreensiva com a mãe que ficou horas e horas ausente para realizar este trabalho.

Ao meu irmão Renato Seiki e aos meus amigos Luciana Maraldi e Rodolfo Jacarandá por ficarem vários fins de semana cuidando da Luiza enquanto eu terminava a escrita deste trabalho.

Ao meu marido João Almeida, por todo o apoio e compreensão. Por cuidar da minha saúde, o que foi crucial para o término deste trabalho, por cuidar da Luiza nos momentos em que tive que estar ausente, por todo o amor.

Aos meus colegas de pesquisa Marcela Ribeiro, Paula Diniz, Cícero Hildenberg e Jonathan Ramos, que me ajudaram no desenvolvimento deste trabalho, propondo ideias, discutindo resultados, literalmente pondo a 'mão na massa' comigo. Muito obrigada pelas discussões, pelo incentivo e pela realização deste trabalho junto comigo.

Aos meus colegas Mônica Ferreira, William Denner, Sérgio Silva, Marcelo Ponciano por toda a ajuda, desde ao auxílio em formatação de computador, ajuda com as questões administrativas, favores diversos até às conversas, risadas e momentos de discussões de ideias.

Às minhas amigas Mônica e Marcela e à minha tia Maria Stela, pela ajuda na revisão do texto desta tese.

A todos os meus colegas do GBDI pela amizade.

Ao jornal Rondoniagora (http://www.rondoniagora.com/) por compartilhar a internet comigo durante a escrita deste trabalho.

Aos meus colegas da UNIR, pelo apoio.

À USP, por toda a infraestrutura e oportunidades.

Ao CNPq, pelo auxílio financeiro.

# Resumo

Com o desenvolvimento e barateamento dos equipamentos de aquisição de imagens, principalmente na área médica, tem sido geradas muitas imagens, as quais devem ser analisadas pelos especialistas. Esta tarefa pode ser muitas vezes cansativa e demorada, levando a possíveis erros no diagnóstico, pois a leitura das imagens depende da experiência e do estado físico e emocional do médico. Assim, sistemas de auxílio ao diagnóstico por computador (*Computer-aided diagnosis* - CAD) têm se tornado grandes aliados no processo de diagnóstico, realizando uma segunda leitura da imagem, servindo como uma segunda opinião ao especialista. Por isso, é necessário o desenvolvimento de técnicas de mineração de imagens para o aumento da precisão e da velocidade da análise das imagens.

Assim, o objetivo deste trabalho foi desenvolver métodos de representação de imagens e de classificação associativa para aumentar a precisão da classificação de sistemas de auxílio ao diagnóstico médico por imagens.

Para a representação de imagens foram desenvolvidas técnicas para reduzir a lacuna que há entre a representação numérica das imagens e seu significado semântico, a qual é chamada de 'gap semântico'. Para isso, foi usada a teoria das redes complexas para modelar as imagens em redes livres de escala, e os descritores das imagens foram compostos pelas medidas topológicas extraídas rede modelada. Os vetores de características gerados foram bem compactos, o que possibilitou também evitar o problema da 'maldição da alta dimensionalidade'.

Para a classificação, foi desenvolvido o classificador associativo SACMiner, por meio do uso de regras de associação estatísticas, o qual evita a fase de discretização de dados, lidando diretamente com dados contínuos. Este foi um passo importante, já que a discretização pode causar a perda de informações e gerar inconsistência na base de dados. Além do SACMiner, foi desenvolvido o classificador MinSAR, o qual, além de não demandar a fase de discretização, também evita que o usuário tenha que fornecer parâmetros de entrada ao algoritmo responsável por gerar as regras. As técnicas até aqui listadas foram aplicadas em um sistema de auxílio ao diagnóstico de mama e comparadas com técnicas descritas na literatura, e os resultados mostram que as técnicas aqui propostas sobrepujaram as atuais da literatura.

E por fim, foram sugeridas novas medidas para caracterizar imagens de pacientes com epilepsia no lobo temporal mesial, por meio do uso de medidas de espessura cortical, as quais melhoraram a precisão do sistema para este tipo de diagnóstico.

# Abstract

The complexity of medical images and the high volume of exams per radiologist in a screening program can lead to a scenario prone to mistakes. Hence, it is important to inforce double reading and effective analysis, but those are costly measures. The computer-aided diagnosis (CAD) technology offers an alternative to double reading, because it can provide a computer output as a 'second opinion' to assist radiologists in interpreting images. Using this technology, the accuracy and consistency of radiological diagnoses can be improved, and also the image reading time can be reduced. Therefore, the need of classification and image representation methods and to speed-up and to assist the radiologists in the image analysis task has been increased. These methods must be more accurate and demand low computational cost, in order to provide a timely answer to the physician.

The aim of this thesis was to developed image representation and associative classifiers methods to improve the classification of computer-aided diagnosis systems.

Considering the image representation, in this work, we present some approaches to reduce the gap between the numeric representation of the images and their semantic, which is called 'semantic gap'. For this, we used the complex network theory to produce an image model based on scale-free networks. The image descriptors were composed of topological measures of the modeled network. The feature vectors produced were quite compact, which also allowed to avoid the problem called as 'curse of dimensionality'.

Considering the classification task, we proposed the SACMiner classifier, which uses statistical association rules in order to avoid the discretization step when working with continuous attributes. It is important because the discretization step can disturb the dataset and cause lost of information. We also proposed de MinSAR classifier, which mines the rules not requiring a discretization step neither input thresholds, as most of the other association rules methods do. These approaches were applied in a breast cancer computer-aided diagnosis system.

And finally, we developed an automatic technique which can aid in distinguishing between controls and patients with mesial temporal lobe epilepsy, based on cortical thickness, and potentially identifying abnormalities in tissue integrity in cases where atrophy cannot be visualized.

# Sumário

| $\mathbf{Li}$ | Lista de Figuras xi |   |                 |  |  |
|---------------|---------------------|---|-----------------|--|--|
| Li            | sta d               | le Tabelas  | xiii            |  |  |
| 1             | Intr                | odução  | 1               |  |  |
|               | 1.1                 | Considerações iniciais                                    | 1               |  |  |
|               | 1.2                 | Motivação e definição do problema                         | 2               |  |  |
|               | 1.3                 | Objetivos   | 4               |  |  |
|               | 1.4                 | Principais contribuições                                  | 5               |  |  |
|               | 1.5                 | Organização do trabalho                                   | 6               |  |  |
| <b>2</b>      | Pro                 | cessamento e caracterização de imagens                    | 9               |  |  |
|               | 2.1                 | Segmentação de imagens                                    | 10              |  |  |
|               |                     | 2.1.1 Contextualização e desafios                         | 10              |  |  |
|               |                     | 2.1.2 Classificação dos métodos de segmentação de imagens | 12              |  |  |
|               | 2.2                 | Extração de características                               | 14              |  |  |
|               |                     | 2.2.1 Características de cor                              | 15              |  |  |
|               |                     | 2.2.2 Textura   | 16              |  |  |
|               |                     | 2.2.3 Forma   | 17              |  |  |
|               | 2.3                 | Considerações finais                                      | 21              |  |  |
| 3             | Rec                 | les complexas   | 23              |  |  |
| Ŭ             | 3.1                 | Definição   | 24              |  |  |
|               | 3.2                 | Modelos de redes complexas                                | 24              |  |  |
|               | 0.2                 | 3.2.1 Grafos aleatórios                                   | 24              |  |  |
|               |                     | 3.2.2 Redes pequeno-mundo                                 | 25              |  |  |
|               |                     | 3.2.3 Redes livres de escala                              | 26              |  |  |
|               | 33                  | Medidas das redes complexas                               | $\frac{20}{28}$ |  |  |
|               | 3.4                 | Aplicação de redes complexas na representação de imagens  | 30              |  |  |
|               | 3.5                 | Considerações finais                                      | 32              |  |  |

| 4        | Min            | neração de imagens  | 33        |
|----------|----------------|---|-----------|
|          | 4.1            | Mineração de dados  | 35        |
|          |                | 4.1.1 Fase de preprocessamento dos dados  | 35        |
|          |                | 4.1.2 Fase de mineração de dados  | 37        |
|          | 4.2            | Mineração de regras de associação   | 38        |
|          |                | 4.2.1 Mineração de regras de associação estatísticas  | 43        |
|          | 4.3            | Classificação   | 46        |
|          |                | 4.3.1 Classificação associativa   | 50        |
|          | 4.4            | Avaliação e interpretação dos resultados  | 53        |
|          | 4.5            | Considerações finais  | 57        |
| <b>5</b> | Des            | envolvimento de extratores de características de forma usando redes   |           |
|          | com            | plexas  | 59        |
|          | 5.1            | Descrição do método: modelando imagens como uma rede livre de escala $\ .$                                  | 60        |
|          |                | 5.1.1 Descritores $\ldots$ | 62        |
|          | 5.2            | Experimentos  | 63        |
|          |                | 5.2.1 Estudo de caso 1: Base ALOI   | 65        |
|          |                | 5.2.2 Estudo de caso 2: Objetos   | 68        |
|          |                | 5.2.3 Estudo de caso 3: Imagens diversas  | 72        |
|          |                | 5.2.4 Estudo de caso 4: desenvolvimento de um sistema de auxílio ao   |           |
|          |                | diagnóstico de câncer de mama   | 73        |
|          | 5.3            | Considerações finais  | 77        |
| 6        | $\mathbf{Des}$ | envolvimento de classificadores baseados em regras de associação  |           |
|          | esta           | tísticas  | <b>79</b> |
|          | 6.1            | O classificador SACMiner  | 80        |
|          |                | 6.1.1 Descrição do SACMiner   | 80        |
|          |                | 6.1.2 Experimentos  | 84        |
|          |                | 6.1.3 Considerações sobre o SACMiner  | 89        |
|          | 6.2            | O classificador MinSAR  | 90        |
|          |                | 6.2.1 Descrição do MinSAR   | 90        |
|          |                | 6.2.2 Experimentos  | 92        |
|          |                | 6.2.3 Considerações sobre o MinSAR  | 94        |
|          | 6.3            | Considerações finais  | 95        |
| 7        | Det            | ecção automática de epilepsia no lobo temporal  | 97        |
|          | 7.1            | Descrição do método   | 98        |
|          |                | 7.1.1 Participantes   | 100       |
|          |                | 7.1.2 Aquisição de imagens e extração de características  | 100       |
|          |                | 7.1.3 Seleção de características  | 102       |
|          |                | 7.1.4 Classificação   | 102       |
|          | 7.2            | Experimentos  | 102       |
|          | 7.3            | Resultados  | 103       |
|          |                | 7.3.1 Espessura cortical de regiões cerebrais do hemisfério esquerdo  | 103       |
|          |                | 7.3.2 Espessura cortical de regiões cerebrais do hemisfério direito   | 104       |
|          |                | 7.3.3 Espessura cortical de regiões cerebrais de ambos hemisférios  | 104       |
|          | 7.4            | Discussão dos resultados  | 105       |
|          | 75             | Considerações finais  | 107       |
|          |                |   | -01       |

| 8  | Con    | clusão                        | 109 |
|----|--------|-------------------------------|-----|
|    | 8.1    | Principais contribuições      | 109 |
|    | 8.2    | Publicações                   | 112 |
|    | 8.3    | Proposta de trabalhos futuros | 113 |
|    | 8.4    | Considerações finais          | 114 |
| Re | eferêr | ncias Bibliográficas          | 115 |

#### ix

# Lista de Figuras

| 2.1        | Desafios em segmentação: imagens não-homogêneas, baixo contraste, e  |                 |
|------------|--|-----------------|
|            | oclusões (Figura adaptada de (Elnakib et al., 2011))   | 11              |
| 2.2        | Exemplo da aplicação do método de Otsu. (a) Imagem original em níveis de   |                 |
|            | cinza, (b) histograma da imagem original e limiar encontrado pelo método   |                 |
|            | de Otsu, (c) imagem segmentada. $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$  | 13              |
| 2.3        | Exemplo de segmentação realizada pelo método EM/MPM adaptado. (a)  |                 |
|            | imagem original; (b) imagem segmentada com $L = 5$ ; (c) região de classe 1  |                 |
|            | (fluido cérebro-espinhal); (d) região de classe 2 (massa cinzenta); (e) região   |                 |
|            | de classe 3 (massa branca); (f) região de classe 4 (dura, medula óssea,  |                 |
|            | gordura) (Retirado de (Balan, 2007)) $\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$  | 15              |
| 2.4        | Exemplos de matrizes de co-ocorrência (Balan, 2007). (a) imagem; (b) ma-   |                 |
|            | triz de co-ocorrência para o ângulo 0° e $d = 1$ ; (c) matriz de co-ocorrência   |                 |
|            | para o ângulo 135° e $d = 1$   | 17              |
| 2.5        | (a) Uma forma simples,(b) seu esqueleto e (c) pequeno distúrbio na borda   |                 |
|            | adiciona novos ramos ao esqueleto (Figura extraída de (Costa & Cesar Jr,   |                 |
|            | 2009)  | 19              |
| <b>9</b> 1 | (a) Evenanda da uma rada alastária da Endia Dánui (h) gráfico da dia   |                 |
| 5.1        | (a) Exemplo de uma rede aleatoria de Erdos-Reilyi, (b) granco da dis-  |                 |
|            | tribuição da media dos grads de 10 redes aleatorias formadas por 10.000<br>vártices usando uma probabilidado $n = 0.2$ . Exemplo retirado do Costa           |                 |
|            | vertices usando uma probabilidade $p = 0.2$ . Exemplo retirado de Costa<br>ot al (2007) p. 176   | 25              |
| 30         | Construção do rodo complexa poqueno mundo com $n = 30$ vérticos o $k = 4$  | 20              |
| 0.2        | Construção de rede comprexa pequeno mundo com $n = 50$ vertices e $k = 4$<br>prostas (a) $n = 0$ (b) $n = 0.01$ (c) $n = 0.1$ (d) $n = 1$ (Adaptado do Watts |                 |
|            | arestas. (a) $p = 0$ , (b) $p = 0.01$ , (c) $p = 0.1$ , (d) $p = 1$ (Adaptado de Watts<br>ls Strogatz (1008))  | 26              |
| 22         | Cráfico de uma rede complexe livre de escale de distribuição des ligrações   | 20              |
| 0.0        | dos vártices seguindo a lei de potência. Figura adaptada de Barabasi l   |                 |
|            | Bonaheau (2003)  | $\overline{27}$ |
| 34         | Exemplo de geração de uma rede livre de escala Figura adaptada de  | 21              |
| 0.1        | Barabasi & Bonabeau (2003)   | 28              |
| 3.5        | Representação de um contorno modelado como uma rede complexa (reti-  | 20              |
| 0.0        | rado de (Backes et al. 2009)   | 31              |
|            | Tado do (2000)   | 91              |
| 4.1        | Fases da mineração de imagens para a obtenção de conhecimento  | 34              |

| $4.2 \\ 4.3$                              | Distribuição Normal com média 0 e variância 1   |                      |  |
|---|---|----------------------|--|
| $5.1 \\ 5.2 \\ 5.3 \\ 5.4$                | Rede complexa livre de escala   | 62<br>65<br>66<br>66 |  |
| 5.5                                       | Imagem da classe 2. (a) Modelagem da rede complexa e (b) gráfico na escala log  | 67                   |  |
| 5.6                                       | Imagem da classe 3. (a) Modelagem da rede complexa e (b) gráfico na escala log  | 67                   |  |
| 5.7                                       | Exemplo de objetos da base MPEG7  | 69                   |  |
| $5.8 \\ 5.9$                              | Exemplos de figuras de algumas classes  | 69                   |  |
| 5.10                                      | 994 nós   | 70                   |  |
| 5.11                                      | <ul> <li>(a) Modelagem da rede complexa e (b) gráfico na escala log. A imagem binarizada possui uma resolução de 352 × 288 pixels e a rede gerada possui</li> </ul> | 70                   |  |
| 5.12                                      | 964 nos   | 71                   |  |
| 5.13                                      | (a) Modelagem da rede complexa e (b) gráfico na escala log. A imagem original possui uma resolução de $320 \times 320$ pixels e a rede gerada possui 5184 pós       | 72                   |  |
| 5 1/                                      | Exemple de uma mamografia digital de tumor (a) benigne e (b) maligne  | 73<br>74             |  |
| 5.14<br>5.15                              | (a) Imagem original; (b) Imagem segmentada em 5 regiões; (c) Máscara da   | 75                   |  |
| 5.16                                      | (a) Modelagem da rede complexa e (b) gráfico na escala log. A imagem original possui uma resolução de $72 \times 72$ pixels e a rede gerada possui 217              | 75                   |  |
| 5.17                                      | nós   | 75<br>76             |  |
|   |   | .0                   |  |
| $\begin{array}{c} 6.1 \\ 6.2 \end{array}$ | Pipeline do método SACMiner   | 81<br>91             |  |
| 7.1                                       | Método proposto   | 99                   |  |

# Lista de Tabelas

| 2.1 | Características de textura de Haralick   | 18 |
|-----|--|----|
| 4.1 | Relações de itens por transação  | 40 |
| 4.2 | Conjunto $C_1$ .   | 41 |
| 4.3 | Conjunto $L_1$   | 41 |
| 4.4 | Conjunto $C_2$   | 41 |
| 4.5 | Conjunto $L_2$   | 41 |
| 4.6 | Conjunto $C_3$   | 41 |
| 4.7 | Conjunto $L_3$   | 41 |
| 4.8 | As principais diferenças entre classificação associativa e descobrimento de  |    |
|     | regras de associação (Thabtah, 2007)   | 51 |
| 4.9 | Matriz de confusão   | 56 |
| 5.1 | Descritor de texturas de Haralick e posições dos valores no vetor de carac-  |    |
|     | terística (Felipe et al., 2003; Ribeiro, 2008).  | 65 |
| 5.2 | Base ALOI.   | 68 |
| 5.3 | Dados das imagens processadas  | 70 |
| 5.4 | Base OBJETOS. Legenda: PT (Precisão Total); Ci (Classe i, $1 \le i \le 8$ ).   | 71 |
| 5.5 | Base mama 250 imagens  | 76 |
| 6.1 | Comparação entre SACMiner e outros classificadores usando a abordagem  |    |
|     | holdout.   | 86 |
| 6.2 | Comparação entre SACMiner e outros classificadores usando a abordagem  | 96 |
| ເາ  | Companyação entre SACMiner o entres elegrificadores usando o elegrificadores usando o elegrificadores usando e elegrificadores usando elegrificadores usan | 00 |
| 0.3 | <i>holdout.</i>  | 87 |
| 6.4 | Comparação entre SACMiner e outros classificadores usando a abordagem  |    |
|     | leave-one-out.   | 87 |
| 6.5 | Comparação dos valores de acurácia entre SACMiner e outro classificadores usando a abordagem <i>holdout</i>  | 89 |
| 6.6 | Resultados comparando MinSAR com outros classificadores para a base de   |    |
|     | 250 imagens  | 92 |
|     | $\sim$   |    |

| 6.7 | Resultados comparando MinSAR com outros classificadores para a base de |  |     |
|-----|--|--|-----|
|     | 569 imagens  |  | 94  |
|     |  |  |     |
| 7.1 | Espessura cortical de estruturas cerebrais do hemisfério esquerdo      |  | 103 |
| 7.2 | Espessura cortical de estruturas cerebrais do hemisfério esquerdo      |  | 104 |
| 7.3 | Espessura cortical de estruturas cerebrais do hemisfério esquerdo      |  | 105 |
|     |  |  |     |

# 1 Introdução

Capítulo

## 1.1 Considerações iniciais

Com o aperfeiçoamento dos sistemas computacionais e dos equipamentos eletrônicos, muitas áreas de pesquisa têm sido desenvolvidas, sendo a medicina uma das áreas que mais tem sido beneficiada. Por causa deste avanço, foi possível desenvolver equipamentos modernos que geram imagens de alta resolução e precisão do corpo humano (Webb, 2012; Shah & Parikh, 2012). Com isso, mais recentemente, os métodos de diagnóstico por imagem, associados aos exames de análises clínicas, história clínica e exame físico, têm se tornado os pilares que a medicina utiliza para a prevenção, diagnóstico, planejamento terapêutico e controle de tratamento dos pacientes (Muralidhar et al., 2011).

Além dos exames de imagens, outros tipos de exames, diagnósticos e procedimentos de tratamento devem ser armazenados por no mínimo 20 anos, segundo leis vigentes em nosso país (Medicina, 2007). Assim, há um grande volume de dados médicos armazenados, que são uma valiosa fonte de conhecimento, o qual pode ser usado para o auxílio ao diagnóstico médico e para o ensino da medicina. Entretanto, devido à complexidade da análise dos dados que incluem imagens, os profissionais da área da saúde ainda não se beneficiam de grande parte dessa fonte de conhecimento. Por exemplo, ainda não existe uma técnica totalmente eficiente que recupere imagens de exames antigos com o mesmo tipo de anomalia mostrado em uma imagem recém obtida. Uma outra situação é a classificação correta de uma imagem recém obtida. Uma das principais causas para a ocorrência destes fatos é o grande número de características extraídas das imagens que podem ser usadas para sua busca ou classificação, mas são desconhecidas quais delas são as mais relevantes para cada tipo de aplicação. Além disso, o uso de muitas características para representar as imagens pode levar ao problema da "maldição da alta dimensionalidade" (Allen et al., 2012; Lei et al., 2012), que degrada a precisão e o tempo de busca e também a performance da classificação. Devido a tais desafios, as técnicas de recuperação de imagens por conteúdo e sistemas de classificação têm sido muito pesquisados nos últimos anos.

#### 1.2 Motivação e definição do problema

A detecção de anomalias em imagens médicas por um radiologista é, em geral, um procedimento demorado, que pode tender a erros e a algum grau de subjetividade devido a várias razões, dentre as quais: a complexidade das imagens médicas; a existência de estruturas complexas e de grande número de imagens normais; a grande variação na aparência dos tecidos (mesmo os normais); a sutileza das anormalidades; a superposição dos tecidos; a necessidade de minimizar o retorno desnecessário dos pacientes; o grande volume de exames por radiologista em um programa de rastreio (Pires et al., 2006).

Por isso, é importante realizar a leitura das imagens e a análise por dois radiologistas, processo este conhecido como dupla leitura (Muralidhar et al., 2011). Entretanto, este processo aumenta os custos para a geração do diagnóstico. A tecnologia de diagnóstico auxiliado por computador (Computer-aided diagnosis - CAD) oferece uma alternativa a dupla leitura, pois pode fornecer uma saída no computador que atua como uma "segunda opinião" para auxiliar radiologistas na interpretação das imagens (Azavedo et al., 2012; Nishikawa et al., 2012). Segundo Shiraishi et al. (2011), os algoritmos que compõem os sistemas CAD geralmente consistem em várias etapas, que incluem processamento de imagens, análise de características de imagens e classificação de dados, usando ferramentas da área de inteligência artificial e aprendizado de máquina. Usando esta tecnologia, a precisão e a consistência do diagnóstico radiológico podem ser melhorados, e também o tempo de leitura da imagem é reduzido. Portanto, a necessidade de métodos de representação de imagens e de métodos de classificação para acelerar e para auxiliar os radiologistas na tarefa de análise de imagens tem aumentado. Estes métodos devem fornecer um resultado de classificação mais preciso e demandar baixo custo computacional para que forneçam um tempo de resposta aceitável ao médico.

Uma abordagem promissora para o desenvolvimento de sistemas CAD é a mineração de imagens, que é uma área multidisciplinar que abrange técnicas de processamento de imagens e técnicas de mineração de dados.

Dentre as técnicas de processamento de imagens, metodologias de extração de características analisam objetos que são representativos às várias classes de objetos. Essas características, se escolhidas cuidadosamente, representam o máximo de informação relevante que a imagem tem a oferecer para uma caracterização completa de uma lesão

(Mohanty et al., 2012). As características extraídas são usadas como entrada aos classificadores que as relacionam à classe que elas representam.

Geralmente, os extratores de características são baseados nas características de cor, textura e forma das imagens. Valores numéricos são extraídos das imagens e são armazenados em um vetor, o qual passa a ser a representação da imagem correspondente. Considerando então a geração desses vetores de características, podemos ressaltar dois desafios desta área:

- A redução do gap semântico que existe entre a representação numérica das imagens, por meio das características de baixo nível, e a representação visual de acordo com a compreensão humana (Akgül et al., 2011).
- O problema da maldição da alta dimensionalidade dos dados, proveniente da grande quantidade de valores numéricos gerados pelos extratores de características da literatura para representar as imagens. A alta dimensionalidade dos dados degrada a precisão e o tempo de busca e também a performance da classificação (Allen et al., 2012; Lei et al., 2012).

É importante ressaltar que uma importante característica de imagens médicas pode ser incluída na representação das imagens: o conhecimento sobre as características gerais e sobre o local esperado das estruturas a serem analisadas (Rebelo et al., 2007).

Já a mineração de dados refere-se a extrair ou "minerar" conhecimento de um grande volume de dados (Kantardzic, 2011; Han & Kamber, 2006). Esta área de pesquisa surgiu do esforço conjunto de pesquisadores de diversas áreas, dentre elas, banco de dados, inteligência artificial, estatística, aprendizado de máquina e visualização, e pode ser vista como um resultado de uma evolução natural da tecnologia da informação (Kantardzic, 2011). Considerando imagens, é importante ressaltar que a mineração deste tipo de dado é uma tarefa muito importante para a área médica, já que a mineração de imagens pode automaticamente descobrir informações e padrões implícitos das imagens envolvidas, comparando um grande número de características. Dessa forma, essa tarefa rapidamente ganha a atenção da comunidade em mineração de dados.

Uma abordagem que tem se mostrado muito apropriada para o desenvolvimento de sistemas CAD é a mineração de classificação associativa. Esta abordagem usa técnicas de descobrimento de regras de associação para construir sistemas de classificação. No domínio da imagem, geralmente as imagens são submetidas a algoritmos de processamento para produzir um vetor de características (geralmente com valores contínuos) que as represente. Os vetores de características são introduzidos aos algoritmos de mineração de regras de associação para revelar suas dependências intra e inter-classe. Estas regras são então empregadas na classificação. Em geral, as abordagens baseadas em regras de associação alcançam altos valores de precisão quando comparados a outros métodos de classificação (Shooshtari et al., 2012; Rodda & Mogalla, 2011; Dua et al., 2009). Outra vantagem dos

classificadores associativos é que por seu modelo de aprendizado ser baseado em regras, permite um fácil entendimento e interpretação do conhecimento obtido (Rodda & Mogalla, 2011). Esse fato é importante, já que o modelo de aprendizado precisa ser fornecido para o especialista validar o padrão minerado.

Entretanto, classificadores associativos desenvolvidos para a área de mineração de dados apresentam alguns desafios quando aplicados no domínio das imagens. Dentre eles, podemos citar:

- Algoritmos tradicionais de regras de associação adotam um método iterativo para descobrir um conjunto de itens frequentes (um conjunto de valores de atributos disjuntos contidos em um objeto de treinamento), o qual requer muitos cálculos e um processo de transação complicado (Deshpande, 2011).
- No domínio de imagens, a representação das imagens geralmente ocorre na forma de valores contínuos. A maioria das abordagens tradicionais de regras de associação que usam valores contínuos requer a discretização dos dados antes de aplicar a mineração de regras de associação. Entretanto, o processo de discretização pode causar perda de informação durante o processo de mineração e gerar inconsistências nos dados (Lustgarten et al., 2011).
- Outro desafio de classificadores associativos é o grande número de regras geradas, o que dificulta o entendimento do conhecimento obtido e pode gerar ambiguidades (Christopher, 2011).
- O *framework* clássico de classificadores associativos requer que o usuário decida quais regras são significantes pela **especificação de limiares**. Este fator muitas vezes demanda um conhecimento específico da informação que está sendo avaliada; caso contrário, a escolha de tais limiares é geralmente arbitrária e pode levar a um grande número de descobertas espúrias que podem enfraquecer o sucesso da análise subsequente (Kirsch et al., 2012).

# 1.3 Objetivos

O objetivo deste trabalho foi definir, modelar e desenvolver técnicas e métodos de representação de imagens e de classificação associativa para aumentar a precisão da classificação de sistemas de diagnóstico médico por imagens auxiliado por computador. Neste propósito, como objetivos específicos, esta tese buscou:

• Desenvolver métodos de extração de características de imagens baseados em redes complexas para diminuir o *gap* semântico existente entre a representação das características de baixo nível das imagens e sua interpretação humana;

- Lidar com o problema da "maldição da alta dimensionalidade", que degrada as estruturas de índice e diminui o poder de discriminação das características extraídas das imagens;
- Analisar características de alto nível e encontrar relações entre elas automaticamente por meio de técnicas de mineração de dados para o auxílio ao diagnóstico médico por imagens;
- Usar regras de associação estatística para desenvolver classificadores associativos a fim de (i) reduzir o custo da fase de geração de itens frequentes, (ii) minimizar a perda de informações e inconsistências nos dados geradas pela fase de discretização dos atributos, (iii) diminuir o número de regras geradas para a construção do modelo de classificação, e (iv) diminuir a redundância e a ambiguidade das regras geradas;
- Desenvolver um classificador associativo que requeira poucos parâmetros de entrada do usuário para o algoritmo gerador das regras;
- Aplicar técnicas e métodos de extração de características e de classificação em sistemas de auxílio ao diagnóstico médico por imagens;
- Aumentar a precisão da classificação, que servirá como segunda opinião ao especialista médico.

## 1.4 Principais contribuições

O trabalho desenvolvido e apresentado nesta tese traz contribuições para a área de sistemas de auxílio ao diagnóstico e mineração de imagens. As principais contribuições envolvem o desenvolvimento de métodos e técnicas de extração de características de imagens e de classificadores associativos e também o desenvolvimento de um *framework* de sistema de auxílio ao diagnóstico. Assim, nesta seção é apresentado o resumo das principais contribuições, que estão listadas a seguir e detalhadas no Capítulo 8:

- Redução do *gap* semântico existente entre a representação numérica das imagens e sua interpretação humana por meio do desenvolvimento de extratores de características baseados em redes complexas, os quais geram vetores com poucos atributos e alto poder de discriminação de imagens.
- Desenvolvimento de dois novos classificadores associativos SACMiner e MinSAR

   os quais evitam a fase de discretização de dados, geram regras simples e diretas e
   não requerem a geração de itens candidatos. Além dessas características, o MinSAR
   não requer que o usuário forneça parâmetros de entrada ao algoritmo de geração de
   regras. Essas características possibilitaram um aumento na precisão da classificação
   de imagens.

- Desenvolvimento de um *framework* de sistema de auxílio ao diagnóstico de câncer de mama usando os classificadores SACMiner e MinSAR, o qual possui a vantagem de apresentar o conhecimento obtido, tornando o sistema mais confiável para ser utilizado pelo especialista.
- Redução do gap semântico e minimização do problema da alta dimensionalidade de dados em um sistema de auxílio ao diagnóstico de epilepsia no lobo temporal mesial (ELTM), por meio de aplicação de algoritmos de classificação e de seleção de atributos, analisando as relações entre as medidas de espessura de estruturas corticais, gerando uma sugestão de diagnóstico mais precisa e relevante ao especialista.

## 1.5 Organização do trabalho

Este trabalho está organizado em oito capítulos, sendo que os capítulos 2 a 4 trazem os conceitos relacionados a este trabalho, e os capítulos 5 a 7 apresentam os trabalhos desenvolvidos. Estes capítulos estão organizados da seguinte maneira.

Capítulo 2 - Processamento e caracterização de imagens. Este capítulo traz alguns conceitos de processamento de imagens relacionados ao trabalho desenvolvido, como os métodos de segmentação usados na literatura e métodos de extração de características baseados na cor, textura e forma das imagens.

Capítulo 3 - Redes Complexas. Neste capítulo é realizada uma pequena introdução à teoria das redes complexas. Nele são apresentados os modelos de redes que propulsionaram o desenvolvimento desta área de pesquisa e também são apresentados trabalhos que usam redes complexas aplicadas à modelagem de imagens.

Capítulo 4 - Mineração de imagens. Este capítulo aborda os conceitos de mineração de imagens e de dados, focando principalmente nas regras de associação estatísticas e nos classificadores associativos, que foram usados no desenvolvimento deste trabalho.

Capítulo 5 - Desenvolvimento de extratores de característica de forma usando redes complexas. Este capítulo apresenta um novo extrator de características que modela a forma de uma imagem em uma rede complexa livre de escala, a partir do contorno e do esqueleto da imagem, e extrai medidas topológicas da rede modelada para compor o vetor de características. São descritos estudos de caso com diversos tipos de imagens para a validação do modelo de rede criada e para a verificação do poder de discriminação das imagens dos descritores propostos. Os descritores também são usados para caracterizar imagens de exames de mamografia para o auxílio ao diagnóstico de câncer de mama.

Capítulo 6 - Desenvolvimento de classificadores baseados em regras de associação estatística. Este capítulo apresenta os dois novos classificadores associativos

desenvolvidos neste trabalho: o SACMiner e o MinSAR. Também é apresentado um *framework* de um sistema de auxílio ao diagnóstico de câncer de mama para avaliar o poder de classificação destes métodos.

Capítulo 7 - Detecção automática de epilepsia no lobo temporal. Este capítulo traz uma aplicação de métodos tradicionais da literatura para o desenvolvimento de um *framework* para detecção de epilepsia no lobo temporal mesial (ELTM) e apresenta o uso de medidas de espessura de estruturas corticais do cérebro para a representação das imagens. São apresentados experimentos sobre a análise dessas medidas usando algoritmos de seleção de atributos e classificadores bem difundidos na literatura.

**Capítulo 8 - Conclusão.** Este capítulo apresenta as conclusões e as principais contribuições do trabalho desenvolvido, e sugere trabalhos futuros.

Capítulo 2

# Processamento e caracterização de imagens

Um sistema de apoio ao diagnóstico (*Computer-Aided Diagnosis*, CAD) permite que o médico use o resultado de análises quantitativas de imagens e textos de laudos para decidir sobre o diagnóstico de um exame (Jiang et al., 2001). A primeira etapa na análise de uma imagem médica é a extração de suas características e segmentação. A seleção e extração das características das imagens são os passos mais críticos para a identificação de objetos na imagem. O principal problema é como identificar um conjunto reduzido de características que descrevam a imagem com precisão suficiente para sua identificação e para a extração de padrões desta.

A segmentação da imagem é considerada uma etapa fundamental na interpretação e representação da imagem, pois ela permite a extração de características de regiões de interesse da imagem, ao contrário dos métodos que buscam uma representação global e direta da mesma, que é inadequada para a maior parte dos domínios.

Neste capítulo é feita uma breve revisão dos métodos de segmentação e de extratores de características de imagens necessários para a compreensão e para o desenvolvimento deste trabalho.

## 2.1 Segmentação de imagens

#### 2.1.1 Contextualização e desafios

Segmentação de imagens é um dos problemas mais interessantes e desafiadores em visão computacional em geral e, especificamente, em aplicações de imagens médicas. A segmentação de imagens médicas é de considerável importância, pois permite adquirir informação não-invasiva sobre estruturas do corpo humano que ajuda os radiologistas a visualizar e estudar a anatomia das estruturas (Pu et al., 2011), simular processos biológicos (Johnson, 2012; Prastawa et al., 2009), localizar patologias (Gao et al., 2012), trilhar o progresso das doenças (Ye et al., 2009; Roy et al., 2011) e avaliar a necessidade de radioterapia ou cirurgias (Pasquier et al., 2010; Lu, 2012). Por estas razões, a segmentação é uma parte essencial de qualquer sistema de diagnóstico auxiliado por computador (Computer-aided diagnosis system - CAD), e a funcionalidade do sistema depende muito da precisão da segmentação. Entretanto, para que seja realizada uma segmentação precisa de imagens médicas, vários desafios devem ser superados. Primeiramente, muitas estruturas anatômicas de imagens médicas são não-homogêneas com respeito à repetição espacial da intensidade individual de pixel/voxel ou suas co-ocorrências agrupadas. Dois exemplos típicos são mostrados na Figura 2.1: ambos, pulmões e rins são difíceis de separar por causa da similaridade dos vasos dos pulmões e aparência visual dos tecidos do tórax, assim como a sobreposição do rim com a aparência do fundo, respectivamente.

Imagens médicas com baixo contraste apresentam desafios adicionais: por exemplo, as bordas de objetos em imagens típicas de MRI de rim e cardíaca na Figura 2.1 são difíceis de serem determinadas, inclusive manualmente. A segmentação se torna até mais complexa com oclusões ou desvios de sinal e ruído causados por diferentes modalidades de formação de imagens e processos de captura de imagem. Alguns exemplos de MRI e fotos digitais de objetos naturais são mostrados na Figura 2.1.

Para superar estes desafios, muitas técnicas de segmentação têm sido desenvolvidas já há algum tempo e divulgadas na literatura. Geralmente, nenhuma técnica de segmentação é adequada para todas as aplicações, e várias abordagens de diferentes precisões, rapidez e grau de complexidade têm sido exploradas para cada problema CAD na prática. Sua caracterização está baseada em finalidades e aspectos específicos do processamento do dado envolvido. Em particular, técnicas baseadas em contorno, região e pixel/voxel visam encontrar fronteiras do objeto em estudo, formando regiões conectadas ocupadas por estes objetos, ou realizando apenas classificação inteligente de pixel/voxel, respectivamente. Um exemplo bastante comum de técnicas baseadas em região ou pixel, por meio de contornos, são os modelos deformáveis (fronteiras), isto é, curvas fechadas 2-D ou superfícies 3-D envolvendo o contorno do objeto procurado, ou força interna à superfície



**Figura 2.1:** Desafios em segmentação: imagens não-homogêneas, baixo contraste, e oclusões (Figura adaptada de (Elnakib et al., 2011))

refletindo o modelo corrente da forma, e/ou força externa, dependendo da homogeneidade do sinal dentro e fora do modelo.

Com relação à interferência do usuário, a segmentação pode ser interativa ou iniciada pelo usuário, ou totalmente automatizada. Muitos problemas de segmentação são determinados em termos de custo específico ou funções de energia dependendo da homogeneidade das regiões de interesse, e suas soluções são encontradas minimizando estes custos ou funções de energia sobre todos os conjuntos de todos os possíveis mapas de regiões ou seu subconjunto provável. Considerando as funções de energia, estas funções geralmente são multimodais no espaço de busca, e métodos exatos ou de otimização global aproximada tais como corte de grafos interativo, "propagação de crenças", programação dinâmica, ou algoritmos genéticos são usados para executar a melhor segmentação.

Em contraste com a segmentação não supervisionada, as técnicas supervisionadas envolvem um conjunto de treinamento adicional (imagens com mapas de regiões obtidos manualmente) para adaptar tipos particulares de objetos e imagens. Segmentação baseada em atlas simplifica a adaptação devido ao uso de um ou mais mapas anatômicos médicos padrões, ou atlas de seções do corpo humano, onde cada ponto do mapa é legendado com a classe do objeto correspondente. A técnica mais simples baseada em atlas alinha uma imagem a um mapa anatômico relevante para associar os labels dos objetos com a imagem (Jia et al., 2012).

#### 2.1.2 Classificação dos métodos de segmentação de imagens

Segundo Gonzalez & Woods (2008), os métodos para segmentação de imagens são classificados em três categorias básicas: limiarização (*thresholding*), segmentação baseada em bordas e segmentação baseada em regiões.

#### Limiarização

A limiarização define um valor de nível de cinza T e, para pixels com valores menores que T, atribui-se valor 0 (valor de intensidade mínima) e, para valores maiores que T, atribui-se valor 1 (valor de intensidade máxima). Assim, as principais vantagens da limiarização são sua simplicidade e baixo custo computacional (Huang et al., 2012). O principal problema desta técnica é definir o valor do limiar T, sendo a técnica que mais tem sido usada é a segmentação de Otsu (Otsu, 1979; Huang et al., 2012).

Na técnica de Otsu, considerando uma imagem em níveis de cinza, primeiramente o histograma normalizado da imagem é considerado como uma função discreta de densidade de probabilidade (Dougherty, 2009):

$$p_r(r_q) = \frac{n_q}{n}, q = 01, 2, 3..., L - 1$$
(2.1)

onde n é o número total de pixels na imagem,  $n_q$  é o número de pixels que tem nível de cinza igual a  $r_q$ , e L é a quantidade possível de níveis de cinza da imagem. Imagine que um limiar k é escolhido de forma que o conjunto  $C_0$  contém todos os pixels com níveis de cinza menores que k e  $C_1$  contém todos os maiores, i.e.,  $C_0 = 0, 1, ..., k - 1$  e  $C_1 = k, k + 1, ...L - 1$ . O método de Otsu escolhe o valor de limiar k que maximiza a variância entre classes  $(G_B^2)$ , definida como (Gonzalez & Woods, 2008):

$$G_B^2 = w_0(\mu_0 - \mu_r)^2 + \omega_1(\mu_1 - \mu_r)^2$$
(2.2)

onde

$$\omega_0 = \sum_{q=0}^{k-1} p_q(r_q) \qquad (2.3) \qquad \mu_0 = \sum_{q=0}^{k-1} q p_q(r_q) / \omega_0 \quad (2.5) \qquad \mu_T = \sum_{q=0}^{L-1} q p_q(r_q). \tag{2.7}$$
$$\omega_1 = \sum_{q=k}^{L-1} p_q(r_q) \qquad (2.4) \qquad \mu_1 = \sum_{q=k}^{L-1} q p_q(r_q) / \omega_1 \quad (2.6)$$

O método de segmentação de Otsu usa a maximização da variância interclasse, que equivale à minimização da variância intraclasse, sendo, no entanto, mais facilmente calculada. A maximização da variância intraclasse consiste em dividir em duas classes  $C_1$  e  $C_2$  e obter a melhor separação entre elas, o que significa a obtenção da maior diferença de contraste entre as classes. Procura-se de uma parte distanciar uma classe da outra,

ou seja, maximizar a variância interclasse, e de outra parte agrupar cada classe em torno de seu centro, ou seja, minimizar a variância intraclasse. Um exemplo da aplicação da técnica de Otsu é apresentada na Figura 2.2.



Figura 2.2: Exemplo da aplicação do método de Otsu. (a) Imagem original em níveis de cinza, (b) histograma da imagem original e limiar encontrado pelo método de Otsu, (c) imagem segmentada.

#### Segmentação baseada em bordas

A segmentação baseada em bordas envolve localizar na imagem a ser segmentada, regiões onde ocorre uma variação abrupta dos níveis de cinza. Essas regiões podem ocorrer na forma de pontos, linhas e segmentos e, a partir delas, são construídos contornos dos objetos presentes na imagem. Algoritmos de enlace são então utilizados para conectar os fragmentos evidenciados para gerar contornos que estejam associados a objetos presentes na imagem. A detecção de descontinuidades é feita por meio de operadores de derivação local. O conceito de gradiente é usado para a diferenciação de imagens, do qual o vetor gradiente aponta para a direção de mudança mais rápida na posição (x, y) da imagem. A derivação pode ser implementada de diferentes formas, e, os operadores mais conhecidos são o operador de Sobel, Prewitt, Canny e Laplaciano. O que é feito na prática, é a convolução destes operadores, que são representados como uma máscara, com a imagem a ser segmentada (Gonzalez & Woods, 2008).

#### Segmentação baseada em regiões

Por fim, a **segmentação baseada em regiões** leva em consideração a similaridade das propriedades dos pixels. O crescimento de região por agregação de pixel realiza o agrupamento de pixels ou grupo de pixels em regiões maiores. Os pixels a serem agrupados devem ter propriedades similares, como nível de cinza, textura, entre outros. Primeiramente, inicia-se o algoritmo com um conjunto de "sementes" em torno do qual as regiões crescem. Os passos desta técnica são: (a) a seleção das sementes, que depende da aplicação do problema; (b) a seleção das propriedades que estabelecem os critérios de

similaridade, que depende do tipo de dados disponíveis; (c) a utilização de conectividade e adjacência; e (d) a formulação de uma regra de parada, como a utilização de critérios de tamanho, semelhança entre um pixel candidato e os pixels da região, e formato de uma dada região.

Um modelo robusto, automático e eficaz para análise de textura, e que tem sido utilizado em técnicas de segmentação de imagens por regiões, são os Campos Aleatórios de Markov (Markov Random Fields) (Gerhardinger, 2006). Um método com esta abordagem que se destaca é o EM/MPM (Comer & Delp, 2000), um algoritmo iterativo que combina a técnica EM (Expectation and Maximization) para realizar uma estimação de parâmetros por máxima verossimilhança com a MPM (Maximizer of the Posterior Marginals), utilizada para segmentação pela minimização do número de pixels erroneamente classificados. Ou seja, o algoritmo EM é usado para ajustar um Modelo de Mistura Gaussiana no histograma da imagem com um classificador baseado nos Campos Aleatórios de Markov que, essencialmente, incorpora o conceito de coerência espacial entre os pixels. Uma extensão desta técnica foi proposta por Balan (2007), a qual faz com que o valor do parâmetro de interação espacial do modelo de Markov aumente gradativamente durante a segmentação de uma imagem. A Figura 2.3 apresenta um exemplo do resultado da segmentação obtida de uma imagem em cinco classes (incluindo o *background*), usando o método proposto por Balan. As regiões de textura obtidas também são mostradas separadamente para uma melhor visualização.

## 2.2 Extração de características

Uma questão importante em sistemas de reconhecimento de padrões é a extração de características apropriadas que caracterizem eficientemente padrões diferentes. A extração de características é um passo importante na detecção e classificação de padrões (Akilandeswari et al., 2012).

Os objetos são representados como uma coleção de pixels em uma imagem. Por isso, para fins de reconhecimento e categorização, é preciso descrever as propriedades dos grupos de pixels. A descrição é geralmente um conjunto de números - o descritor do objeto, ou assinatura, o qual também é chamado de vetor de características. A partir destes números, é possível comparar e reconhecer objetos simplesmente fazendo a correspondência dos descritores dos objetos em uma imagem com os descritores de objetos conhecidos. Os descritores devem ter quatro propriedades importantes. Primeiro, devem definir um *conjunto completo*, i.e., dois objetos devem ter os mesmos descritores se e somente se eles têm a mesma forma. Segundo, eles devem ser *congruentes*. Dessa maneira, é possível reconhecer objetos semelhantes quando possuírem descritores semelhantes. Terceiro, é conveniente que eles tenham propriedades *invariantes* à rotação, escala e translação. E por fim, os descritores devem ser um *conjunto compacto*. Ou seja, um descritor deve



Figura 2.3: Exemplo de segmentação realizada pelo método EM/MPM adaptado. (a) imagem original; (b) imagem segmentada com L = 5; (c) região de classe 1 (fluido cérebro-espinhal); (d) região de classe 2 (massa cinzenta); (e) região de classe 3 (massa branca); (f) região de classe 4 (dura, medula óssea, gordura) (Retirado de (Balan, 2007))

representar a essência de um objeto de maneira eficiente, contendo apenas informação sobre o que torna o objeto único, ou diferente dos outros objetos. A quantidade de informação usada para descrever esta categorização deve ser menos que a informação necessária para ter uma descrição completa do próprio objeto. Infelizmente não existe um conjunto completo e compacto de descritores para categorizar objetos de maneira genérica. Por isso, os melhores resultados são alcançados selecionando cuidadosamente as propriedades. Assim, o processo de reconhecimento está fortemente relacionado a cada aplicação particular com um tipo de objeto particular (Nixon & Aguado, 2008).

Devido a esta particularidade, uma grande variedade de métodos de extração de características têm sido desenvolvidos nas últimas décadas. Neste capítulo, o conteúdo visual de uma imagem é analisado em termos de características de baixo-nível, incluindo cor, textura e forma.

#### 2.2.1 Características de cor

As características de cor são as características visuais mais usadas em recuperação de imagens por serem geralmente simples de serem extraídas quando comparadas com a informação de textura e forma. A característica de cor é relativamente robusta quanto ao embaraço do fundo da imagem e independe do tamanho ou da orientação da imagem (Singh & Hemachandran, 2012).

Uma das técnicas mais comuns usadas para representar uma imagem quanto ao nível de cinza (cor) é o **histograma tradicional**. O histograma fornece a frequência de ocorrências de um histograma de cor específico obtido diretamente dos pixels da imagem. Sua popularidade em sistemas de imagens é principalmente devido à sua distinta propriedade de custo linear, assim como de ser invariante a rotação, translação e escala, para histogramas normalizados. Esta técnica pode ser usada como um primeiro passo, de baixo custo, antes de aplicar um extrator de características com custo maior (Traina et al., 2011).

Entretanto, algumas das desvantagens dos histogramas tradicionais são que não indicam a localização espacial dos pixels na imagem e imagens diferentes podem ser representadas por um mesmo histograma. Além do problema da ambiguidade, o histograma também apresenta alta dimensionalidade, geralmente com valores indo de 100 a mais de 4.000 níveis de cinza para cada imagem de raio-X, em unidades de Hounsfield em tomografia computadorizada, por exemplo. Isso porque um histograma com 100 cores distintas representa uma imagem como um ponto em um espaço com dimensionalidade 100. A maioria das estruturas de indexação de base de dados espaciais sofre com a "maldição da alta dimensionalidade" (Jeong et al., 2009), perdendo sua capacidade de agilização do processamento de consultas.

Para contornar o problema da maldição da alta dimensionalidade dos histogramas de cor, algumas abordagens têm sido desenvolvidas. Traina et al. (2003) propõe o histograma métrico, o qual considera a correlação entre os *bins* adjacentes dos histogramas extraídos das imagens. Já Bugatti et al. (2008) usa a entropia de Shannon para reduzir a dimensionalidade de histogramas em nível de cinza tradicionais.

Ainda em pesquisas recentes, os histogramas de cor tradicionais têm mostrado que, apesar de simples, são um dos mais eficientes para representar imagens em recuperação por conteúdo (Deselaers et al., 2008). Além disso, eles ainda são base para comparação com outros métodos propostos (Zhang, 2012) e muitos métodos desenvolvidos ainda são baseados neles (Pattanaik & Bhalke, 2012; Weinland et al., 2011).

#### 2.2.2 Textura

Uma importante abordagem para descrever uma região é quantificar seu conteúdo de textura. Apesar de não existir uma definição formal para textura, intuitivamente, este descritor fornece medidas relacionadas à suavidade, aspereza e regularidade. Dentre as características de baixo-nível usadas para representar imagens médicas, os extratores de textura se destacam por sua efetiva discriminação de tecidos. As medidas de textura capturam principalmente a granularidade e a repetição de padrões na distribuição dos pixels.

Algumas das técnicas mais conhecidas de extração de características de textura baseiam-se nas wavelets (Silva et al., 2008; Stollnitz et al., 1996; Fan, 2012), nos filtros de Gabor (Gabor, 1946; Saremi et al., 2013) e sumarizações das matrizes de co-ocorrência (conhecidas como descritores de Haralick) (Haralick et al., 1973; Shahbahrami et al., 2012).

As matrizes de co-ocorrência, também denominadas matrizes SGLD (Spatial Gray Level Dependence) (Haralick et al., 1973), são uma das mais populares fontes de características de textura para imagens. Dado uma imagem f com um conjunto discreto de tons de cinza I, define-se a matriz de co-ocorrência  $P_{d,\phi}(i, j)$ , onde cada elemento (i, j)é um número inteiro que indica quantas vezes um pixel  $p_1$  de nível de cinza i aparece distante de um pixel  $p_2$  de intensidade j por uma distância d e um ângulo  $\phi$ . As Figuras 2.4 (b) e (c) ilustram duas matrizes de co-ocorrência para a imagem em tons de cinza representada na Figura 2.4 (a). As matrizes de co-ocorrência são matrizes quadradas e simétricas em relação à diagonal principal, ou seja,  $P_{d,\phi}(i, j) = P_{d,\phi}(j, i)$ .

**Figura 2.4:** Exemplos de matrizes de co-ocorrência (Balan, 2007). (a) imagem; (b) matriz de co-ocorrência para o ângulo  $0^{\circ} e d = 1$ ; (c) matriz de co-ocorrência para o ângulo  $135^{\circ} e d = 1$ 

Diversas medidas podem ser extraídas de uma matriz de co-ocorrência. Haralick et al. (1973) propõe 14 medidas. Dessas, as mais usadas na literatura são contraste, entropia, energia e homogeneidade, que estão apresentadas na Tabela 2.1. Como este método necessita de computar várias vezes matrizes de grandes dimensões, pode-se citar como grande desvantagem o seu alto custo computacional. Além disso, a escolha dos parâmetros  $d e \phi$  depende muitas vezes de um conhecimento prévio da qualidade das imagens. No entanto, as características extraídas das matrizes de co-ocorrência promovem uma descrição satisfatória das imagens sem a necessidade de uma etapa prévia de segmentação.

#### 2.2.3 Forma

As características de forma são usadas para detectar, por exemplo, massas e calcificações em mamografias. São também chamadas de características geométricas ou morfológicas. Este tipo de característica é baseada na forma de regiões de interesse, e não consideram

| Característica | Equação  | Significado   |
|----------------|--|---------------|
| Variância      | $\sum_{i}\sum_{j}(i-j)^2 P(i,j)$                 | Contraste     |
| Entropia       | $\sum_{i}\sum_{j}^{i} P(i,j)log(P(i,j))$         | Suavidade     |
| Energia        | $\sum_{i}\sum_{j}^{j}P(i-j)^{2}$                 | Uniformidade  |
| Homogeneidade  | $\sum_{i} \sum_{j}^{i} \frac{P(i-j)}{(1+ i-j )}$ | Homogeneidade |

 Tabela 2.1: Características de textura de Haralick.

a intensidade dos pixels da região, tomando apenas a forma da região segmentada. (Akilandeswari et al., 2012).

Segundo Costa & Cesar Jr (2009), as representações de formas podem ser baseadas em três categorias: contorno, região e transformadas. Representações baseadas em contorno exploram as informações contidas na borda de um objeto. O contorno da forma pode ser representado em três classes: (i) como uma curva paramétrica, implicando então em uma ordem sequencial; (ii) como um conjunto de pontos, sem nenhuma ordem especial entre eles; e (iii) como um conjunto de primitivas geométricas(por exemplo, segmentos de reta) ajustadas ao contorno.

As abordagens baseadas em região, incluem: (i) decomposição de região, na qual a região da forma é particionada em formas mais simples (por exemplo, polígonos) e representada pelo conjunto de tais primitivas; (ii) fronteira da região, na qual a forma é aproximada por uma primitiva geométrica especial pré-definida (por exemplo, um retângulo); e (iii) características internas, onde a forma é representada por um conjunto de características relacionadas a sua região interna (por exemplo, um esqueleto).

Esqueletos estão relacionados à estrutura essencial de cada forma. O esqueleto constitui um conjunto de pontos no interior de um objeto de uma imagem, de modo a representá-lo. A definição de esqueleto diz que, um ponto pertence ao esqueleto se, ele é o centro de um círculo máximo, que toca a borda do objeto em pelo menos dois pontos distintos. Assim, o processo de esqueletização consiste em reduzir-se a figura analisada a uma cadeia simples, com a largura de apenas um pixel. O problema é a dificuldade de se implementar círculos no plano discreto e a sua aproximação é extremamente cara, computacionalmente falando. A maioria dos algoritmos de esqueletização encontrados na literatura utiliza o mesmo conceito aplicado ao centro de um losango, de um quadrado, ou uma combinação de ambos. Esta técnica além de ser bastante eficiente, consegue aproximar do conceito original de esqueletização. Alguns algoritmos de esqueletização são apresentados em (Souza, 2002). A Figura 2.5 mostra um exemplo de uma figura, seu esqueleto e de seu esqueleto quando é alterada a borda.

As abordagens baseadas em transformadas incluem as características a seguir. (i) Linear, na qual uma transformação linear T é tal que, dada duas formas  $A \in B$ , e dois escalares  $\alpha \in \beta$ , temos  $T(\alpha A + \beta B) = \alpha T(A) + \beta T(B)$ . Uma transformação linear representa uma forma A em termos de T(A), por exemplo, os coeficientes (ou


**Figura 2.5:** (a) Uma forma simples,(b) seu esqueleto e (c) pequeno distúrbio na borda adiciona novos ramos ao esqueleto (Figura extraída de (Costa & Cesar Jr, 2009)

energias) de suas transformações (por exemplo, Fourier). (ii) Não linear: são semelhantes à transformação linear, i.e., a forma é representada em termos de suas transformações, com a diferença que as transformações são não lineares.

Como mencionado, ambas as transformações são usadas para representação da forma e descrição. Dentre as principais técnicas que geralmente são aplicadas para obter descritores a partir de transformações pode-se citar como exemplo de transformações lineares as transformadas de Fourier, Laplace, Wavelets, filtros de Gabor, e para o caso de transformações não lineares a transformada Hough e Morfologia Matemática (Costa & Cesar Jr, 2009; Gonzalez & Woods, 2008).

Ainda segundo Zahn & Roskies (1972), pode-se descrever a forma e as características de contorno de objetos basicamente de três maneiras: por aproximação polinomial, por momentos invariantes e por descritores de Fourier. Backes et al. (2007) propõe uma aproximação poligonal de contornos baseada em Redes Complexas. Para momentos invariantes, tem-se os momentos de Zernike, que têm sido muito usados na literatura em diversas aplicações, como representar imagens mamográficas (Rosa et al., 2008; Wei & Li, 2008), e suas variações, como os momentos complexos de Zernike (Li et al., 2009) e os pseudo-momentos de Zernike (Yang & Guo, 2008), os quais são mais sensíveis a ruidos na imagem que os momentos de Zernike convencional. As principais vantagens dos momentos de Zernike é que são invariantes à transformação de rotação na imagem, e, se fizer que o centro de massa da imagem coincida com a origem do disco unitário no mapeamento, também torna-se invariante à translação. Além disso, não é necessário uma etapa prévia de segmentação das imagens (Felipe et al., 2006).

Antes de introduzir os descritores de forma, é interessante notar que um significado natural para produzir características interessantes é obter novas medidas a partir das que já foram computadas anteriormente, como o contorno ou o esqueleto, por exemplo, que são características por si mesmos, mas sendo possível considerar seu perímetro ou área como novas características.

Alguns descritores gerais, relacionados ao aspecto de métrica da forma podem ser considerados. Como exemplo, pode-se citar área, tamanho do eixo maior e do eixo menor, excentricidade, orientação, área convexa, área preenchida, número de Euler, solidez, amplitude e perímetro. Essas medidas são apresentadas a seguir com respeito a dimensão em pixels (Gonzalez & Woods, 2008):

- i) area: número atual de pixels da região.
- ii) tamanho do eixo maior: especifica o tamanho (em pixels) do maior eixo da elipse que tem o mesmo segundo momento central da região.
- iii) tamanho do eixo menor: o tamanho (em pixels) do menor eixo da elipse que tem o mesmo segundo momento central normalizado da região.
- iv) excentricidade: escalar que especifica a excentricidade da elipse que tem o mesmo segundo momento da região. É o raio da distância entre os focos da elipse e o tamanho de seu eixo maior. O valor está entre 0 e 1. Observe que os valores 0 e 1 são os casos degenerados; uma elipse com excentricidade 0 é um círculo, enquanto que uma elipse com excentricidade 1 é um segmento de reta.
- v) orientação: é o ângulo (em graus, de -90 a 90 graus) entre o eixo x e o eixo maior da elipse que tem o mesmo segundo-momento da região.
- vi) área convexa: escalar que especifica o número de pixels da imagem binarizada
- vii) área preenchida: escalar especificando o número de pixels da imagem binarizada.
- viii) número de Euler: Escalar que especifica o número de objetos na região menos o número de buracos nestes objetos.
- ix) solidez: especifica a proporção dos pixels no casco convexo que também estão na região. É calculado como área/áreaconvexa.
- x) extensão: escalar que especifica a quantidade de pixels na região limitada (textitbounding box). É calculada como área/área do textitbounding box.
- xi) perímetro: distância em torno da fronteira da região. É calculada a distância entre cada par de pixels contínuos em torno da borda da região.

Uma das maneiras de definir forma é do ponto de vista estrutural. A forma pode ser vista como as partes que a compõe e as relações entre tais partes. Assim, um modo de representar as formas é por meio de grafos, nos quais as 'partes' da forma estão associadas aos vértices do grafo e as relações entre as partes são representadas pelas arestas do grafo. Outra possibilidade introduzida recentemente é explorar grafos por meio de redes complexas, a qual está um pouco mais detalhada no Capítulo 3, já que esta foi uma das abordagens adotadas neste trabalho.

# 2.3 Considerações finais

Neste capítulo foram apresentadas as principais técnicas de processamento de imagens que são empregadas para a obtenção da representação das imagens a partir de vetores de características. O conhecimento do processamento de imagens é importante para entender como é realizado o processo de análise automática de imagens, a mineração de imagens e por conseguinte, o desenvolvimento de sistemas de auxílio ao diagnóstico, processos estes que são o objetivo deste trabalho. Os vetores de características obtidos nesta etapa são os dados de entrada no sistema de auxílio ao diagnóstico desenvolvido.

No próximo capítulo é apresentado o conceito de redes complexas e os modelos de pequeno mundo e de redes livres de escala, os quais foram usados neste trabalho para o desenvolvimento de um extrator de características baseado em redes complexas.

Capítulo 3

# **Redes complexas**

As informações e dados do mundo real são organizados em diversos tipos de estruturas. Um modo de organizá-los, é através de suas ligações como um modelo de redes. Assim, redes (ou grafos) têm sido objeto de estudo em várias áreas como matemática, sociologia, ciência da computação, geografia quantitativa, dentre outros. No caso de redes aleatórias, o primeiro e mais importante modelo foi proposto por Erdos e Rényi no fim dos anos de 1950's e foi a base da maioria dos estudos até recentemente (Barthélemy, 2011). O interesse em redes foi, entretanto, renovado em 1998 por Watts & Strogatz (1998), que extraíram informações interessantes de redes do mundo real e propuseram um simples e novo modelo de redes aleatórias. O interesse por esta área de estudo foi reforçado depois da publicação, um ano depois, de um artigo por Barabasi & Albert (1999) sobre a existência de forte heterogeneidade de graus. Estas fortes heterogeneidades foram muito contrastantes com os grafos aleatórios, que foram considerados por muito tempo, e a existência de fortes flutuações nas redes do mundo real desencadeou uma riqueza de estudos. Mais de uma década depois, é possível encontrar muitos livros (Caldarelli, 2007; Barrat et al., 2008; Newman, 2010) e revisões sobre este assunto (Albert & Barabasi, 2001; Dorogovtsev & Mendes, 2003; Newman, 2003; Boccaletti et al., 2006; Costa et al., 2007, 2011; Barthélemy, 2011). Logo, a razão para os estudos terem sido intensificados desde aquela época é que redes reais têm características que não podem ser explicadas pela conectividade aleatória uniforme. Ao invés disto, redes geradas a partir de dados reais podem envolver estrutura de comunidade, distribuição de graus de potência e hubs, entre outras características estruturais.

Os dois modelos de redes que contribuíram para o desenvolvimento dos estudos nesta área foram as redes de pequeno-mundo (Watts & Strogatz, 1998; Watts, 1999) e a caracterização de modelos livres de escala de (Barabasi & Albert, 1999), os quais são abordados no decorrer deste capítulo.

Neste capítulo é apresentada uma breve introdução à teoria das redes complexas. A estrutura deste capítulo inclui: a definição de redes complexas; a descrição dos modelos de rede pequeno mundo e do modelo livre de escala; medidas das redes; e aplicação das redes complexas para a caracterização de imagens.

# 3.1 Definição

Uma *rede* é um conjunto de itens, que são chamados *vértices* ou *nós*, com conexões entre eles, chamadas *arestas*. Redes também são chamadas de "grafos" na maior parte da literatura de matemática (Newman, 2003).

Uma maneira intuitiva de trabalhar com redes é utilizar a matriz de adjacência M[i, j]característica do sistema, onde  $i \in j$  representam os vértices e os elementos de matriz  $m_{ij}$ representam as arestas. As ligações representadas pelas arestas podem ser unidirecionais, bidirecionais ou sem direção (matriz simétrica). Podem ser simples, com M[i, j] = 1 ou M[i, j] = 0, ou ponderadas, ou seja, com pesos diferenciados, com  $M[i, j] \in \mathbb{R}$ .

# 3.2 Modelos de redes complexas

Com o objetivo de se estudar as propriedades topológicas de redes reais, diversos modelos de redes têm sido propostos. Em Costa et al. (2011) são apresentadas diversas aplicações e modelos de redes complexas em problemas reais, incluindo redes biológicas, internet, redes sociais, linguística, economia, química, dentre outras. Nesta seção são apresentados três modelos de redes que apoiaram o desenvolvimento do estudo da aplicação de redes complexas: as redes (grafos) aleatórias, redes de pequeno mundo e redes livres de escala.

## 3.2.1 Grafos aleatórios

Os grafos aleatórios desenvolvidos por Rapoport (1951, 1953) e independentemente por Erdös & Rényi (1959) podem ser considerados o modelo mais básico de redes complexas (Costa et al., 2007). Erdös & Rényi (1959) introduziram um modelo para gerar grafos aleatórios consistindo de N vértices e M arestas. Inicialmente, o grafo é iniciado com N vértices desconectados. A rede é então construída adicionando M arestas de maneira aleatória, mas evitando auto-conexões e conexões múltiplas. Outro modelo similar define N vértices e uma probabilidade p de cada par de vértices conectados. Este último modelo é conhecido como modelo Erdös-Rényi (ER). Para este modelo de rede, em uma rede com muitos pontos, tal que  $N \to \infty$ , o número médio de conexões de cada vértice  $\langle k \rangle$ , dado por

$$\langle k \rangle = p(N-1), \tag{3.1}$$

diverge se p for fixado (ou seja,  $\langle k \rangle \to \infty$ ). Ao invés disso, p é escolhido como uma função de N para tomar  $\langle k \rangle$  fixado:  $p = \langle k \rangle / (N - 1)$ . Para este modelo, P(k) (a distribuição de graus) é uma distribuição de Poisson. A Figura 3.1 mostra um exemplo deste tipo de rede e o gráfico da distribuição de graus de um grafo aleatório.



Figura 3.1: (a) Exemplo de uma rede aleatória de Erdös-Rényi, (b) gráfico da distribuição da média dos graus de 10 redes aleatórias formadas por 10.000 vértices usando uma probabilidade p = 0.2. Exemplo retirado de Costa et al. (2007), p. 176.

## 3.2.2 Redes pequeno-mundo

Em 1967, Milgram (1967) mostrou em um experimento de redes sociais que entre duas pessoas existem apenas seis contatos intermediários, mesmo que elas estejam em lados opostos do mundo. Essa propriedade foi chamada de pequeno mundo (*small world*). Em um grafo, isso significa que a maioria dos vértices podem ser alcançados a partir de outros através de um pequeno número de arestas.

Outra propriedade presente em muitas redes é a presença de um grande número de laços (*loops*) de tamanho três. Isto significa que se um vértice i está conectado aos vértices j e k, então há uma alta probabilidade dos vértices j e k estarem conectados. Esta propriedade é chamada de coeficiente de agrupamento (*clustering coefficient*).

As chamadas redes complexas de pequeno mundo, propostas por Watts & Strogatz (1998), possuem a propriedade de pequeno mundo definida por Milgram (1967) e também

a característica de possuir muitos laços pequenos. Logo, apenas possuir a propriedade de pequeno mundo não caracteriza a rede como sendo do modelo de pequeno mundo.

Por exemplo, as redes de ER apresentam a propriedade de pequeno mundo, mas possuem uma quantidade reduzida de laços pequenos. Por outro lado, redes regulares que possuem muitos laços pequenos são mais fáceis de serem construídas (Costa et al., 2007), mas a distância média entre os nós é grande (ou seja, não possuem a propriedade de pequeno mundo). Portanto, o modelo pequeno mundo está situado entre uma estrutura ordenada finita e um grafo aleatório que apresenta a propriedade de pequeno mundo e alto coeficiente de agrupamento.

Para construir uma rede de pequeno mundo, o modelo inicia com n vértices organizados na forma de um anel, com cada vértice ligado aos seus  $\kappa$  vizinhos mais próximos. No exemplo apresentado na Figura 3.2 a rede possui n = 30 vértices e  $\kappa = 4$  arestas para os vizinhos, inicialmente representada pela Figura 3.2(a). A seguir, para gerar o modelo pequeno mundo, é tomada uma fração das arestas, as quais são aleatoriamente re-ligadas com probabilidade p. Quando p = 0 (Figura 3.2(a)), tem-se uma estrutura ordenada regular com alto número de laços, mas grandes distâncias, e quando  $p \rightarrow 1$ (Figura 3.2(d)) a rede torna-se um grafo aleatório com distâncias curtas mas poucos laços. Logo, o parâmetro p interpola entre um comportamento completamente regular e um comportamento completamente aleatório. Assim, para valores intermediários de p (Figuras 3.2(b) e 3.2(c)), o modelo da rede é de pequeno mundo apresentando tanto distâncias curtas quanto um grande número de laços.

O grau de distribuição para uma rede de pequeno mundo é semelhante ao das redes aleatórias, tendo a forma de sino, mas com um pico em  $\langle k \rangle = 2\kappa$  (Costa et al., 2007).



**Figura 3.2:** Construção de rede complexa pequeno mundo com n = 30 vértices e  $\kappa = 4$  arestas. (a) p = 0, (b) p = 0.01, (c) p = 0.1, (d) p = 1 (Adaptado de Watts & Strogatz (1998)).

## 3.2.3 Redes livres de escala

Depois do modelo de redes de pequeno mundo de Watts & Strogatz (1998), Barabasi & Albert (1999) mostraram que muitos sistemas reais são caracterizados por uma

distribuição desigual. Ao invés dos vértices destas redes terem um padrão aleatório de conexões com um grau característico, como o modelo de pequeno mundo, por exemplo, alguns vértices estão altamente conectados enquanto outros têm poucas conexões, com a ausência de um grau característico. Mais especificamente, o grau de distribuição tem sido encontrado para seguir uma lei de potência para um k grande,

$$P(k) \sim k^{-\gamma}.\tag{3.2}$$

Estas redes são chamadas redes livres de escala. Uma característica deste tipo de rede é a existência de *hubs*, i.e., vértices que são ligados a uma fração significante do total de número de arestas da rede. Considerando então a distribuição das arestas, a característica que define tais redes é que, um gráfico que representa o número de arestas (na escala log) vs. o número de vértices (na escala log), resulta em uma reta (Barabasi & Bonabeau, 2003). A Figura 3.3 apresenta um exemplo do gráfico de distribuição das ligações dos vértices.



**Figura 3.3:** Gráfico de uma rede complexa livre de escala da distribuição das ligações dos vértices seguindo a lei de potência. Figura adaptada de Barabasi & Bonabeau (2003).

O modelo de rede de Barabási-Albert está baseado em duas regras básicas: crescimento e ligação preferencial. A rede é gerada iniciando com a adição de novos vértices. Para cada novo vértice, m novas arestas são inseridas entre os novos vértices e alguns vértices prévios. Os vértices que receberam as novas arestas são escolhidos seguindo uma regra de ligação preferencial linear, i.e., a probabilidade do novo vértice i estar conectado com um vértice existente j é proporcional ao grau de j,

$$P(i \to j) = \frac{k_j}{\sum_u k_u}.$$
(3.3)

Além disso, a maioria dos vértices conectados têm grande probabilidade de receber novos vértices. Isto é conhecido como o paradigma "o rico fica cada vez mais rico". A Figura 3.4 apresenta um exemplo de geração de uma rede livre de escala. A rede livre de escala cresce incrementalmente a partir de 2 vértices até 11 vértices neste exemplo. Um novo vértice (azul) prefere se conectar a um vértice existente (vermelho) que já possui muitas outras conexões. Estes dois mecanismos básicos (crescimento e ligação preferencial) eventualmente levará o sistema a ser dominado por hubs, vértices que possuem um grande número de arestas.



**Figura 3.4:** Exemplo de geração de uma rede livre de escala. Figura adaptada de Barabasi & Bonabeau (2003).

# 3.3 Medidas das redes complexas

Uma vez modelada a rede complexa, é possível calcular algumas características topológicas que caracterizam sua conectividade, tais como o grau e a forma de um vértice, e a junção do grau, que estão descritos a seguir.

O grau  $k_i$  de um vértice i é o número de arestas diretamente conectadas a ele, e é definido por meio da matriz de adjacências A como

$$k_i = \sum_{j=1}^{N} a_{ij},$$
 (3.4)

onde N é o número de vértices existentes na rede e  $a_{ij}$  são elementos da matriz de adjacências A.

Baseando-se no grau dos vértices é possível extrair outras medidas da rede. Duas destas medidas são o grau máximo  $(k_{\kappa})$ ,

$$k_{\kappa} = \max_{i} k_{i},\tag{3.5}$$

#### e o grau médio $(k_{\mu})$ ,

$$k_{\mu} = \frac{1}{N} \sum_{j=1}^{N} k_{i}.$$
(3.6)

As informações topológicas mais usadas são sobre os vértices da rede. Uma generalização natural no caso e redes ponderadas é a **força**  $s_i$ , definida como (Barrat et al., 2004)

$$s_i = \sum_j^N w_{ij},\tag{3.7}$$

onde o peso da rede é descrito por uma matriz W cuja entrada  $w_{ij}$  fornece o peso das arestas que conectam os vértices  $i \in j$  (e  $w_{ij} = 0$  se os nós  $i \in j$  não estão conectados). Na verdade, a força de um vértice combina a informação sobre sua conectividade e a intensidade dos pesos de suas ligações. Portanto, fornece característica relevante da rede. Além disso, baseando-se no grau do vértice, é possível extrair a **força máxima**  $(ms_i)$ ,

$$ms_i = \max_i s_i, \tag{3.8}$$

e a **força média**  $(as_i)$ ,

$$as_i = \frac{1}{N} \sum_{j=1}^{N} s_i.$$
 (3.9)

Muitas vezes é interessante analisar a correlação entre os graus de diferentes vértices, os quais têm sido encontrados para desempenhar um papel importante em muitas propriedades estruturais e dinâmicas da rede (Costa et al., 2007). A abordagem mais natural é considerar a correlação entre dois vértices conectados por uma aresta. Esta correlação pode ser expressada pela junção de distribuição de graus P(k, k'), i.e., como a probabilidade que uma aresta arbitrária conecta um vértice de grau k a um vértice de grau k'. A escolha de k' pode ser feita de forma arbitrária, e neste caso considera-se  $k_i = k'$ , i.e.,  $P(k_i, k')_i$  mostra a probabilidade de um vértice i ter um vizinho com o mesmo grau (Backes et al., 2009).

Analisando esta distribuição de probabilidade, é possível extrair diversas medidas, tais como entropia, energia e a média do grau de junção.

Historicamente, a **entropia** tem sido associada com a quantidade de ordem, desordem e/ou caos em um sistema, e é definida como

$$H = -\sum_{i=1}^{N} P(k_i, k')_i \log_2 P(k_i, k')_i.$$
(3.10)

A energia de um sistema é dada por

$$E = \sum_{i=1}^{N} (P(k_i, k')_i)^2.$$
(3.11)

A média do grau de junção denota a probabilidade de encontrar dois nós arbitrários na rede com o mesmo grau:

$$P = \frac{1}{N} \sum_{i=1}^{N} P(k_i, k')_i.$$
 (3.12)

Outras medidas podem ser encontradas em Costa et al. (2007) e Boccaletti et al. (2006).

# 3.4 Aplicação de redes complexas na representação de imagens

Como consequência de sua flexibilidade para representação e modelagem virtuais de qualquer sistema discreto, redes complexas também são potencialmente úteis na análise de forma (Costa, 2004; Costa et al., 2007).

Em Costa (2004) é dada uma definição de representação de uma imagem como uma rede complexa. Seja a imagem original em níveis de cinza de tamanho  $M \times M$  pixels, sendo que cada pixel pode ter valores de níveis de cinza entre 0 e G, que é proporcionalmente normalizado para o intervalo [0, 1]. Portanto, a rede que representa a imagem possui  $N = M^2$  pixels e n = N(N - 1)/2 arestas ponderadas, as quais são representadas pela matriz ponderada W de tamanho  $N \times N$ . As Equações 3.13 a 3.15 definem um possível mapeamento entre cada nó *i* da rede e os pixels (x, y) da imagem, onde  $1 \le x, y \le M$  e a função mod(a, b) representa o resto da divisão de *a* por *b*.

$$i = y + (x - 1)M \tag{3.13}$$

$$x = \lfloor (k-1)/M \rfloor + 1 \tag{3.14}$$

$$y = mod((k-1), M) + 1 \tag{3.15}$$

Ainda segundo Costa (2004), diversos tipos de interações entre dois pixels podem ser definidos respectivamente a cada atributo visual possível, tais como intensidade luminosa, componentes de cor, forma local, textura, *motion* e disparidade, assim como as posições dos pixels e a adjacência espacial entre os pares de pixels. Os valores escalares derivados de tais propriedades podem ser organizados em um vetor de características, de tal maneira que cada nó i da rede torna-se associado a um vetor de característica que descreve uma propriedade visual local ao redor do pixel respectivo da imagem.

Em (Backes et al., 2009; Backes, 2010) é feita uma aplicação do uso de redes complexas para análise de forma do contorno. A ideia principal é representar uma forma nos termos de uma rede de Watts & Strogatz (1998) seguida da análise de suas características topológicas e dinâmicas. Este modelo de rede apresenta uma propriedade que é chamada de pequeno mundo, na qual todos os vértices podem ser alcançados por qualquer outro através de um número pequeno de vértices. Outra propriedade importante é que esse tipo de rede apresenta um grande número de *loops* de tamanho três, i.e., se o vértice i está conectado ao vértice  $j \in k$ , existe uma grande probabilidade dos vértices  $j \in k$ também estarem conectados (alto coeficiente de aglomeração). O modelo dinâmico da rede pequeno mundo é obtido artificialmente por meio de limiares sequenciais aplicados nos vértices da forma modelada. A rede obtida está correlacionada com a estrutura das redes pequeno mundo em vários estágios do crescimento da rede. O estudo de suas propriedades dinâmicas (medidas derivadas da dinâmica de crescimento da rede, baseado na variação do número de componentes conectados) produzem uma forma única. Assim, esta forma pode ser usada para o processo de análise e classificação da imagem, por meio de extração de medidas que irão compor um vetor de características. A Figura 3.5 apresenta um exemplo de um contorno modelado como uma rede complexa.



**Figura 3.5:** Representação de um contorno modelado como uma rede complexa (retirado de (Backes et al., 2009)

No trabalho de Casanova et al. (2013) é realizada uma generalização do método proposto por Backes et al. (2009). Aqui, ao invés de modelar um contorno em um grafo e usar regras de redes complexas para caracterizar o contorno, o trabalho propõe uma ferramenta matemática para a caracterização de sinais, curvas e conjunto de pontos. O método proposto é usado para a caracterização de imagens de nervuras de folhas. As nervuras foram modeladas em um conjunto de pontos, que por sua vez, foram modelados em grafos. Os descritores foram compostos pelas medidas de grau máximo e grau médio em uma evolução dinâmica. Em Florindo et al. (2013) é proposto um método que consiste em modelar a textura da imagem por meio de redes complexas, representando cada valor da intensidade do pixel como um vértice, e conectando dois vértices se houverem pixels com intensidades correspondentes em uma vizinhança 8-conectada. Assim, a matriz de adjacência é analisada como um objeto de interesse em uma imagem binária e é estimada a dimensão fractal deste objeto. No passo seguinte, é aplicada uma transformação multi-escala sobre a dimensão obtendo os descritores de textura. Os descritores obtidos foram usados na tarefa de classificação de texturas e os resultados obtidos foram muito próximos aos resultados obtidos pelos descritores de Gabor, alcançando ambos mais de 90% de precisão na classificação das imagens de folhas.

# 3.5 Considerações finais

Neste capítulo foi apresentada uma pequena introdução ao conceito de redes complexas e a descrição dos principais modelos. Além disso, foram apresentadas aplicações de redes complexas na modelagem de imagens, para a representação das mesmas a partir da forma/contorno e textura.

Até aqui foram discutidas as técnicas para a representação de imagens. Uma vez obtidos os vetores de características das imagens, a próxima etapa agora é minerar esses dados, ou seja, descobrir padrões que forneçam informações relevantes ao especialista e classificar essas imagens. Para isto, são aplicadas técnicas de mineração de dados, no nosso caso, as tarefas de associação e classificação, a serem discutidas no próximo capítulo.

Capítulo 4

# Mineração de imagens

Os avanços no processo de aquisição e armazenamento de imagens tem proporcionado um grande crescimento das bases de dados de imagens. Estas imagens, se analisadas, podem revelar informações importantes para o usuário humano (Deshpande, 2011). A mineração de imagens lida com a extração do conhecimento que está implícito, com a relação dos dados da imagem, ou outros padrões que não estão explicitamente armazenados nas imagens. Assim, a mineração de imagens é mais que apenas uma extensão da mineração de dados para o domínio de imagem. É um campo interdisciplinar que conta com os conhecimentos de visão computacional, processamento de imagens, recuperação de imagens, mineração de dados, aprendizado de máquina, banco de dados e inteligência artificial (Becker et al., 2010; Zhang et al., 2001).

Por definição, mineração de imagens lida com a extração de padrões das imagens a partir de uma grande coleção de imagens (Zhang et al., 2001). Claramente mineração de imagens é diferente da representação de baixo-nível da visão computacional e das técnicas de processamento de imagens, pois o foco da mineração de imagens é em extrair padrões de uma grande coleção de imagens, enquanto que o foco das técnicas de visão computacional e de processamento de imagens é compreender e/ou extrair características específicas apenas de uma imagem.

Enquanto parece haver algumas partes que coincidem entre a mineração de imagens e a recuperação baseada em conteúdo (ambas abordagens lidam com grandes coleções de imagens), a mineração de imagens vai além do problema de recuperar imagens relevantes. Na mineração de imagens, o objetivo é descobrir os padrões das imagens que são significantes em uma dada coleção de imagens.

Talvez, o conceito errado mais comum da mineração de imagens é que a mineração de imagens não passa da aplicação de algoritmos existentes de mineração de dados em imagens. Isto não é verdade, pois há três diferenças importantes entre base de dados relacional e base de dados de imagens (Pan, 2012; Hsu et al., 2002). (1) A primeira é que enquanto na base de dados relacional os dados são semanticamente relevantes (por exemplo, todos compreendem a expressão "idade igual a 6"), na base de imagens os valores por eles mesmos podem não possuir um significado compreensível, a menos do contexto que estão (por exemplo, o nível de cinza 200 é mais claro que o valor 45). (2) A segunda diferença é que a informação espacial implícita é crítica para a interpretação de conteúdo de imagens, o que já não é problema para as bases de dados relacionais. Para tentar sobrepujar este problema, os mineradores de imagens extraem características que sejam independentes da posição antes de minerar os padrões das imagens. (3) E a terceira diferença é que as características das imagens têm múltiplas interpretações para o mesmo padrão visual. Logo, os algoritmos de mineração de dados tradicionais para associar um padrão a uma classe (interpretação) não funcionam bem com imagens, pois lidam apenas com uma única interpretação. Por causa dessas diferenças, é necessário que uma nova classe de algoritmos de mineração, para atender às necessidades especiais para minerar imagens, seja desenvolvida.

Para lidar com esses desafios, espera-se que um bom sistema de mineração de imagens forneça aos usuários um acesso efetivo ao repositório de imagens e generalização de conhecimento e padrões implícitos nas imagens. Para este fim, tal sistema tipicamente engloba as seguintes funções com imagens: armazenamento, processamento, extração de características, indexação e recuperação, descoberta de padrões e conhecimento, conforme está ilustrado na Figura 4.1. A fase de preprocessamento serve para aumentar a qualidade da imagem a ser analisada, ressaltando determinadas particularidades e atenuando ruídos. Na extração de características, valores são calculados a partir das características visuais de baixo nível das imagens, como cor, textura e forma, e são armazenados em vetores multidimensionais para a representação das imagens na mineração. Uma vez geradas as características, a mineração pode ser realizada usando técnicas da mineração de dados para descobrir padrões significantes. Os padrões resultantes são avaliados e interpretados para obter o conhecimento final relacionado ao domínio em questão (Min & Shuangyuan, 2010; Hsu et al., 2002).



Figura 4.1: Fases da mineração de imagens para a obtenção de conhecimento.

As duas primeiras fases do processo de mineração de imagens foram apresentadas e discutidas nos Capítulos 2 e 3. Neste capítulo são descritas as fases de mineração e de avaliação e interpretação dos dados.

## 4.1 Mineração de dados

Uma vez que as imagens já estão representadas por meio de vetores de características, a próxima etapa é a mineração dos dados propriamente dita. Segundo Han & Kamber (2006), mineração de dados (*data mining*) é o processo de explorar grandes quantidades de dados à procura de informações e padrões ocultos. É também conhecido como processo de Descoberta de Conhecimento em Base de Dados (*Knowledge Discovery in Databases* – KDD).

Fayyad et al. (1996) considera 9 fases no processo de KDD: conhecimento do domínio da aplicação, obtenção do conjunto de dados, limpeza e preprocessamento dos dados, redução do volume de dados, escolha da tarefa de mineração, escolha do algoritmo de mineração, mineração de dados, interpretação e avaliação dos padrões descobertos, e, utilização do conhecimento obtido, o que tem sido apoiado por outros autores (Ribeiro, 2008).

Considerando que o processo de representação dos dados já foi realizado, o processo de KDD pode ser compactado nas seguintes etapas: preprocessamento, mineração dos dados e apresentação, as quais estão descritas a seguir.

## 4.1.1 Fase de preprocessamento dos dados

O preprocessamento dos dados envolve a eliminação de ruídos e erros nos dados, além de também ser responsável, se necessário, pela conversão do formato dos dados. O principal objetivo desta etapa é fornecer dados de qualidade para a fase de mineração dos dados. Esta etapa envolve a discretização dos dados e a seleção de características.

## Discretização

O processo de *discretização* muda o domínio dos dados de contínuo para discreto, dividindo os valores contínuos em intervalos. Assim, o objetivo de um algoritmo de discretização é encontrar o melhor conjunto de pontos de corte (limite de um intervalo de valores reais) para discretizar os dados. A discretização pode ser usada para lidar com os seguintes problemas:

- (i) aplicar sobre dados contínuos algoritmos de mineração desenvolvidos para dados nominais;
- (ii) distribuição dispersa de valores;

(iii) muitos valores diferentes e poucas instâncias de dados.

Entretanto, como desvantagem, o processo de discretização causa a perda de informações no processo de mineração, o que pode causar uma queda significante na precisão dos resultados. Por isso, diversos trabalhos têm buscado desenvolver técnicas de discretização adequadas (Abraham et al., 2006; Ribeiro et al., 2008a; Tillander, 2012).

## Seleção de características

A seleção de características seleciona os atributos mais relevantes, ou seja, que possuem maior poder de representação dos dados, e remove os que são irrelevantes, redundantes e os que são ruído. A esse processo de reduzir o número de características (atributos) chamamos de redução de dimensionalidade. O intuito de se reduzir a dimensionalidade dos dados é aumentar a velocidade e a precisão dos algoritmos de mineração de dados. Além disso, a seleção de características ajuda a construir um modelo geral mais simples e conciso e a obter uma melhor compreensão da tarefa de mineração realizada (Tan, 2007; Dash & Liu, 1997).

E importante ressaltar que o conceito de seleção de características é diferente do de extração de características (ou transformação de características). Enquanto que a transformação de características cria novas características combinando as originais (por exemplo, análise de componentes principais - PCA, análise de discriminante linear - LDA), a seleção de características mantém os valores originais das características selecionadas, o que é desejável em muitos domínios.

Um dos algoritmos de seleção de características mais conhecidos é o Relief (Kira & Rendell, 1992), cujo pseudocódigo está descrito no Algoritmo 1. A ideia principal do algoritmo é estimar a qualidade das características (atributos) considerando quão bem seus valores distinguem entre as instâncias da mesma e de diferentes classes que estão próximas umas as outra. Em cada iteração, um padrão x é aleatoriamente selecionado, e então dois vizinhos mais próximos de x são encontrados: um da mesma classe, chamado *nearest hit NH*, e o outro de uma classe diferente, chamado *nearest miss NM*. O peso (ou estimador de qualidade) da i-ésima característica é atualizado:

$$w_i = w_i + |x^{(i)} - NM^{(i)}| - |x^i - NH^{(i)}(x)|.$$
(4.1)

O número de iterações é definido pelo usuário.

#### Algoritmo 1: Algoritmo Relief

- 1 Inicialização: dado  $D = \{(x_n, y_n)\}_{n=1}^N$ , faça  $w_i = 0, 1 \le i \le I$ , número de iterações T;
- 2 para t=1:T faça
- **3** Selecione aleatoriamente um padrão x de D;
- 4 Encontre o nearest hit NH(x) e nearest miss NM(x) de x;
- 5 | para i=1:I faça
- 6 Calcule:  $w_i = w_i + |x^{(i)} NM^{(i)}| |x^i NH^{(i)}(x)|;$
- 7 fim
- s fim

A complexidade de tempo do Relief é O(kN), onde N é o número de instâncias da base de treinamento e k é o número de características. O algoritmo retorna uma lista de características ordenadas de acordo com suas relevâncias, mas não fornece uma indicação do número de características que deve ser removido. Uma limitação do Relief é que trabalha apenas com conjuntos que envolvem classificação binária. O desenvolvimento de uma extensão do algoritmo Relief para trabalhar com dados cuja classificação pode assumir múltiplos valores foi feita, e o algoritmo proposto foi chamado de Relief-F, proposto por Kononenko (1994).

Outro algoritmo muito usado é o método baseado em árvore de decisão, ou *decision tree method - DTM* (Cardie, 1993). A DTM realiza uma busca para frente (*forward search*) para gerar subconjuntos de características usando o critério de entropia para avaliá-los. A DTM executa o algoritmo C4.5, proposto por Quinlan (1993), o qual constrói uma árvore de decisão, que é uma sequência de atributos que define o estado de uma instância. Desta forma, a DTM seleciona as características que aparecem na árvore de decisão como o melhor subconjunto de características (Ribeiro, 2008).

## 4.1.2 Fase de mineração de dados

Os primeiros *frameworks* de mineração de imagens procuraram usar técnicas existentes para extrair informações das imagens (Hsu et al., 2002). Dentre estas técnicas estão o reconhecimento de objetos (Mori et al., 2012), a indexação e recuperação por conteúdo (Xu et al., 2012; Feng et al., 2012), a classificação e agrupamento de imagens (Xu, 2012), a mineração de regras de associação (Pan et al., 2012; Mohanty et al., 2011) e as redes neurais (Ciresan et al., 2012).

Ainda segundo (Sousa, 2006), as tarefas de mineração mais comuns são:

- **Classificação:** atividade de aprendizado cujo objetivo principal é encontrar uma função que mapeie casos em classes pré-definidas, representadas por atributos discretos ou categóricos;
- **Regressão:** é conceitualmente similar à classificação com a diferença de que o atributo de interesse a ser predito é geralmente um atributo contínuo;
- **Regras de associação:** são regras do tipo "se L então R" que identificam situações em que os valores de um conjunto de atributos (L) determinam os valores assumidos por outro conjunto de atributos (R);
- Sumarização: atividade descritiva cujo propósito é encontrar uma descrição compacta para um subconjunto dos dados como, por exemplo, valores de média e desvio padrão dos atributos;
- **Detecção de Agrupamentos** (*clustering*): visa a identificação de agrupamentos (*clusters*) que descrevem os dados de maneira que os casos de um agrupamento sejam similares uns aos outros e o mais distintos possível dos casos de outros agrupamentos;
- Modelagem de Dependências: consiste em encontrar um modelo que descreva as dependências significativas entre atributos.

É importante observar que as tarefas de mineração não são completamente disjuntas, muitas vezes elas se inter-relacionam, onde técnicas desenvolvidas para uma tarefa podem ser aplicadas em outra e vice-versa. Um exemplo de inter-relacionamento são os classificadores associativos, onde técnicas de associação são utilizadas para a classificação de novos dados (Thabtah, 2007).

Das tarefas listadas na Seção 4.1.2, nas próximas seções são discutidas as tarefas de regras de associação e de classificação, que foram as exploradas neste trabalho para o desenvolvimento de sistemas de auxílio ao diagnóstico por computador (CAD).

# 4.2 Mineração de regras de associação

A mineração de regras de associação é uma das tarefas mais importantes no campo da mineração de dados devido a sua vasta aplicabilidade e a facilidade de compreensão dos padrões que gera. É uma tarefa que encontra relacionamentos entre a ocorrência de itens nas tuplas da base de dados (registros) e foi primeiramente apresentada em Agrawal et al. (1993), como descrito a seguir.

Seja  $I = i_1, ..., i_n$  um conjunto de literais chamados itens. Um conjunto  $X \subseteq I$  é chamado de *itemset*. Seja R uma tabela com transações t envolvendo elementos que são subconjuntos de I. Uma regra de associação é uma expressão na forma  $X \to Y$ , onde

 $X \in Y$  são *itemsets*. X é chamado de corpo ou antecedente da regra e Y é chamado de cabeça ou consequente da regra.

O Suporte é a razão entre o número de transações de R que contém o itemset  $X \cup Y$ e o número total de transações de R, ou seja, mede o quão frequente é uma determinada regra no conjunto de dados:

$$Suporte = \frac{|X \cup Y|}{|R|}.$$
(4.2)

A Confiança é a fração do número de transações contendo X que também contém Y, ou seja, mede a frequência com que os itens em Y ocorrem em transações que contém X:

$$Confiança = \frac{|X \cup Y|}{|X|}.$$
(4.3)

Assim, o problema de minerar regras de associação, como foi proposto inicialmente, consiste em encontrar regras de associação que satisfazem às restrições de suporte mínimo (denominadas regras frequentes) e confiança mínima, especificados pelo usuário. Se considerarmos um algoritmo de força bruta para resolver o problema, seria necessário gerar todas as regras possíveis e então calcular os valores de suporte e confiança para cada regra, descartando as que não atenderem às restrições de suporte e confiança mínimos. No entanto, esta não é uma maneira viável de resolver o problema, já que o total de regras que podem ser geradas cresce exponencialmente considerando o número de itens existentes na base de dados (Costa, 2012).

O descobrimento de regras de associação é tipicamente feito em dois passos (Deshpande, 2011): (i) descoberta dos *itemsets* frequentes e (ii) a geração de regras de associação. O segundo passo é bastante simples e é comum para a maioria dos algoritmos, e o primeiro passo domina o tempo de processamento.

O algoritmo mais conhecido e utilizado, por sua simplicidade, para a determinação de *itemsets* frequentes foi proposto por (Agrawal & Srikant, 1994), denominado Apriori (Wu et al., 2007). Ele encontra todos os conjuntos de *itemsets* frequentes  $(L_k)$ .

O Algoritmo 2 apresenta o algoritmo Apriori, como descrito em Agrawal & Shafer (1996). O algoritmo pode ser descrito em três passos:

- 1. Contagem de ocorrências: o primeiro passo do algoritmo é simplesmente contar as ocorrências dos itens para determinar a frequência de 1-*itemset*.
- 2. Geração de candidatos: no passo seguinte, digamos, k, os *itemsets* frequentes  $L_{k-1}$  encontrados no (k-1)-ésimo passo são usados para gerar os *itemsets* candidatos  $C_k$ , usando o procedimento de geração de candidatos descrito abaixo.
- 3. Suporte: ainda no passo k, a base de dados é percorrida e é contado o suporte dos candidatos em  $C_k$ .

| Algoritmo 2: Algoritmo Apriori.  |
|--|
| <b>Dados</b> : Tabela com tuplas $t$ , suporte mínimo $minsup$                       |
| <b>Resultado</b> : Conjunto de <i>itemsets</i> frequentes                            |
| 1 $L_1 = \{1 \text{-itemsets frequentes}\};$   |
| 2 k = 2;   |
| $3 \ \mathbf{enquanto} \ L_{k-1} \neq \emptyset \ \mathbf{faça}$                     |
| 4   $C_k$ = novos candidatos de tamanho k gerados a partir de $L_{k-1}$ ;            |
| 5 para cada tupla t na base de dados D faça  |
| <b>6</b> Incremente o contador de todos os candidatos em $C_k$ que estão contidos em |
| $t;$   |
| 7 fim  |
| <b>8</b> $L_k = \text{todos os candidatos em } C_k \text{ com suporte mínimo;}$      |
| $9  \mathbf{k} = \mathbf{k} + 1;$  |
| 10 fim   |
| 11 retorna $\cup_k L_k$  |

No passo de geração de candidatos, a ideia é que se um *itemset* X tem suporte mínimo, então todos os subconjuntos de X também o possuem. Logo, dado  $L_{k-1}$  o conjunto de todos os (k-1)-*itemsets* frequentes, um superconjunto do conjunto de todos os k-*itemsets* frequentes é gerado pela (i) junção de  $L_{k-1}$  com  $L_{k-1}$  e (ii) poda, na qual são deletados todos os *itemsets*  $c \in C_k$  tal que algum (k-1)-subconjunto de c não esteja em  $L_{k-1}$ .

A seguir é apresentado um exemplo da execução do Apriori para a geração de regras. Seja uma base de dados B que contém um conjunto de itens  $A = \{a, b, c, d, e\}$  e um conjunto de transações  $T = \{1, 2, 3, 4\}$  no qual a relação de itens ocorridos considerando cada transação  $t_i$  está apresentada na Tabela 4.1, e considerando o valor de suporte mínimo igual a duas transações (50%).

Tabela 4.1: Relações de itens por transação.

| Transação | Itens |
|-----------|-------|
| 1         | a,b,c |
| 2         | b,c   |
| 3         | d,c   |
| 4         | a,e   |

Ao aplicar o algoritmo Apriori na base de dados B, o algoritmo percorre a base gerando um conjunto  $C_1$  de 1-*itemsets* candidatos e em seguida gera um conjunto  $L_1$  de 1-*itemsets* frequentes, como é apresentado nas Tabelas 4.2 e 4.3.

| Itemset | Suporte |
|---------|---------|
| {a}     | 2       |
| {b}     | 2       |
| {c}     | 3       |
| {d}     | 1       |
| {e}     | 1       |

| Tabela 4.2: | Conjunto | $C_1$ . |
|-------------|----------|---------|
|-------------|----------|---------|

Tabela 4.3: Conjunto  $L_1$ .

| Itemset | Suporte |
|---------|---------|
| {a}     | 2       |
| {b}     | 2       |
| {c}     | 3       |

Depois, o algoritmo gera um conjunto  $C_2$  de 2-*itemsets* candidatos. O próximo passo é verificar quais dos 2-*itemsets* candidatos em  $C_3$  são frequentes, gerando, assim, um conjunto  $L_2$  de 2-*itemsets* frequentes, como é apresentado nas Tabelas 4.4 e 4.5.

Tabela 4.4: Conjunto  $C_2$ .

| Itemset   | Suporte |
|-----------|---------|
| $\{a,b\}$ | 1       |
| $\{a,c\}$ | 1       |
| $\{b,c\}$ | 2       |

#### Tabela 4.5: Conjunto $L_2$ .

| Itemset | Suporte |
|---------|---------|
| ${b,c}$ | 2       |

Os passos do algoritmo são executados pelo Apriori até não ser mais possível gerar um conjunto de *itemsets* frequentes.

Tabela 4.6: Conjunto  $C_3$ .

| Itemset         |  |
|-----------------|--|
| $\{\emptyset\}$ |  |

#### Tabela 4.7: Conjunto $L_3$ .

| Itemset         |  |
|-----------------|--|
| $\{\emptyset\}$ |  |

Muitos outros algoritmos baseados no Apriori têm sido propostos para lidar com a dificuldade de encontrar os *itemsets* frequentes. Park et al. (1997) propôs o algoritmo DHP (*Direct Hashing and Pruning*) usando diretamente *hashing* e poda. Holt & Chung (2002) propôs o IHP usando *hashing* invertido e poda. Li et al. (2003) estende o algoritmo Apriori com uma técnica efetiva de poda horizontal. Além de algoritmos baseados no Apriori, também existe o FP-growth (Han et al., 2004), o qual usa uma estrutura de árvore prefixa para minerar *itemsets* frequentes sem a geração de candidatos e percorre a base de dados apenas duas vezes. Tanto o Apriori quanto o FP-growth mineram padrões frequentes a partir de um conjunto de transação de dados de forma horizontal. Uma alternativa é executar a mineração de forma vertical, o que traz a vantagem de não precisar percorrer a base de dados para calcular o suporte dos (k + 1)-*itemsets* (para  $k \ge 1$ ). Um exemplo deste tipo de algoritmo é o Eclat (Zaki & Hsiao, 2000; Li et al., 2011). Extensões destes algoritmos podem ser encontrados em vários artigos (Zhu & Deng, 2011; Naganathan

et al., 2011; Tiwari et al., 2010; Li et al., 2011; Martins et al., 2007). Entretanto, mesmo com a presença de uma vasta literatura que floresceu desde Agrawal et al. (1993), abordando variantes do problema, estudando questões fundamentais, e apresentando novas estratégias de algoritmos inteligentes ou implementações de estratégias conhecidas, muitos problemas ainda permanecem abertos (Han et al., 2007; Kirsch et al., 2012). Em particular, a tarefa de avaliar a importância dos conjuntos de itens descobertos, ou equivalentemente, sinalizar descobertas estatisticamente significativas com um número limitado de resultados falsos positivos, ainda é pouco compreendida e continua sendo um dos problemas mais desafiantes nesta área. O *framework* clássico de mineração de regras de associação requer que o usuário decida o que é significativo, especificando os limiares de suporte. A menos que o domínio do conhecimento esteja disponível, a escolha de um limitar é muitas vezes arbitrária (Tan et al., 2006) e pode levar a um grande número de descobertas espúrias (falsos positivos), o que afetaria o sucesso da análise subsequente (Kirsch et al., 2012).

Além da determinação dos *itemsets* frequentes, um outro problema é o grande número de regras geradas. Algumas pesquisas têm proposto o uso do ganho de informação (Chen et al., 2006) e o uso do conjunto Rough (Yao, 2012) para reduzir o número de regras geradas. Em Christopher (2011), uma abordagem estatística é usada para gerar um conjunto compacto de regras usando apenas regras correlacionadas positivamente, sendo que as regras menos significativas são eliminadas. Ainda em Yamamoto et al. (2008) foram desenvolvidas técnicas de visualização de *itemsets* para a análise visual dos mesmos, que permitem ao usuário selecionar os *itemsets* que ele tem mais interesse que apareçam nas regras.

Como o interesse de estudo deste trabalho são imagens, é necessário considerar que na maioria das vezes os dados envolvidos são contínuos. Esse é um grande desafio na mineração de regras de associação em imagens, já que a maioria dos métodos desenvolvidos na literatura consideram dados de domínio discreto (categórico), sendo necessário, em muitos casos, a discretização dos valores contínuos. Entretanto, apesar de muitos métodos eficientes de discretização de dados serem desenvolvidos, a discretização pode não ser desejada, já que causa a perda de informação dos dados e pode distorcer os resultados do algoritmo de mineração (Ribeiro, 2008).

Uma abordagem para lidar com dados contínuos é a mineração de regras de associação estatísticas, que são regras de associação encontradas com base na distribuição dos valores dos atributos quantitativos (Aumann & Lindell, 1999). Esta abordagem está detalhada e discutida a seguir.

## 4.2.1 Mineração de regras de associação estatísticas

Definidas inicialmente por Aumann & Lindell (1999), as regras de associação estatísticas passam a noção de encontrar um "comportamento interessante", gerando regras que revelem um fenômeno diferenciado dos outros na base de dados. Para isso, são usados testes estatísticos aceitáveis para confirmar a validade das regras descobertas. Os algoritmos propostos com esta definição não usam o processo de discretização dos dados, mas vê os atributos quantitativos como contínuos. Assim, o lado esquerdo da regra representa uma descrição de um subconjunto da população, e o lado direito da regra é a descrição de um comportamento interessante particular da população descrita no lado esquerdo. Logo, é necessário descrever o que significa "comportamento interessante". Para atributos categóricos (discretos), o comportamento está descrito por uma lista de itens e sua probabilidade de ocorrência. Estatisticamente, esta descrição é a distribuição de probabilidade do conjunto de itens, para a população dada. Logo, segundo Aumann & Lindell (1999), para um conjunto de valores quantitativos (valores contínuos), a melhor descrição de seu comportamento é sua distribuição. Os autores ainda escolhem os valores de média e variância para descrever o comportamento de um conjunto de valores quantitativos.

Uma das distribuições de probabilidade mais interessantes é a Distribuição Normal. Uma importante característica da Distribuição Normal é que ela pode ser usada como aproximação para o cálculo de outras distribuições, quando há um grande número de observações. Esse fato é devido ao Teorema Central do Limite, que afirma que quando o tamanho da amostra aumenta, a distribuição amostral da sua média aproxima-se cada vez mais de uma distribuição normal (Devore, 2011). Um exemplo de curva de distribuição normal está na Figura 4.2.



Figura 4.2: Distribuição Normal com média 0 e variância 1.

Considerando uma média  $\mu$  e desvio padrão  $\sigma$ , todas as distribuições normais têm as seguintes propriedades: (i) 68.27% das observações estão a menos de  $\pm 1\sigma$  da média  $\mu$ ; (ii) 95.45% das observações estão a menos de  $\pm 2\sigma$  de  $\mu$ ; e (iii) 99.73% das observações estão a menos de  $\pm 3\sigma$  de  $\mu$ . Esta regra é conhecida como regra 68-95-99 ou regra empírica.

Assim, pode-se observar que na distribuição normal na maior parte das variáveis existe um valor central, representado pela média, em torno do qual se situam a maioria dos indivíduos e, à medida que toma-se valores mais afastados da média, observa-se que os indivíduos são menos frequentes.

O teste Z é um teste estatístico usado para inferir se a diferença entre as médias de uma amostra de uma população, que seguem a Distribuição Normal, é significante estatisticamente. Como as bases de imagens usadas neste trabalho são suficientemente grandes (n > 100) (Glaser, 2001), foi usado o teste Z para a mineração de regras de associação estatística.

Uma extensão das técnicas propostas por Aumann & Lindell (1999) para discriminação de imagens médicas é proposta por Ribeiro et al. (2009a), e está descrita a seguir.

#### O algoritmo StARMiner

Ribeiro et al. (2009a) propôs o algoritmo StARMiner (*Statistical Association Rule Miner*) para selecionar características. O StARMiner é um algoritmo supervisionado que usa regras de associação estatísticas para encontrar regras de associação que selecionem o conjunto mínimo de características que preservem a habilidade de diferenciar imagens de acordo com suas categorias (classes). Este algoritmo trabalha com regras de associação empregando diretamente os dados contínuos, sem discretizá-los. Isto é feito por meio da associação das classes a características com alto poder para discriminar imagens. A formalização deste algoritmo está descrita abaixo, conforme Ribeiro et al. (2009b).

Seja T a base de dados de imagens,  $x_j$  uma classe de imagem,  $T_{x_j} \in T$  o subconjunto de imagens da classe  $x_j$  e  $f_i$  a i-ésima característica do vetor de características F. Seja  $\mu_{f_i}(Z)$  e  $\sigma_{f_i}(Z)$ , respectivamente, a média e o desvio padrão dos valores de características  $f_i$  no subconjunto de imagens Z. O algoritmo usa três limitares definidos pelo usuário:

•  $\gamma_{min}$ : a confiança mínima para rejeitar o teste de hipótese

$$H_0: \mu_{f_i}(T_{x_j}) = \mu_{f_i}(T - T_{x_j}), \tag{4.4}$$

onde a média  $\mu_{f_i}(T_{x_i}) \in \mu_{f_i}(T - T_{x_i})$  são estatisticamente diferentes;

- $\Delta \mu_{min}$ : a diferença mínima permitida entre a média dos valores da característica  $f_i$  nas imagens da classe  $x_i$  e a média dos valores de  $f_i$  no restante da base;
- $\Delta \sigma_{max}$ : o desvio padrão máximo permitido de  $f_i$  em uma dada classe.

O algoritmo StARMiner minera regras na forma:

$$x_j \to f_i \tag{4.5}$$

se a hipótese  $H_0$  for rejeitada e se as condições dadas nas Equações 4.6 e 4.7 forem satisfeitas.

$$|\mu_{f_i}(T_{x_j}) - \mu_{f_i}(T - T_{x_j})| \ge \Delta \mu_{min}$$
(4.6)

$$|\sigma_{f_i}(T_{x_i})| \ge \Delta \sigma_{max} \tag{4.7}$$

Uma regra  $x_j \rightarrow f_i$  retornada pelo algoritmo, relaciona uma característica  $f_i$  com uma classe  $x_j$ , onde os valores de  $f_i$  têm um comportamento estatisticamente diferente em imagens da classe  $x_j$ . Esta propriedade indica que  $f_i$  é uma característica interessante para diferenciar imagens da classe  $x_j$  de outras classes. As características retornadas na regra minerada pelo StARMiner têm um comportamento particular e uniforme em imagens de uma dada categoria. Isto é importante, pois as características que apresentam comportamento uniforme para toda imagem na base de dados, independentemente da categoria da imagem, não contribuem para caracterizá-la e devem ser eliminadas. Assim, as regras geradas são úteis para revelar a relevância das características das imagens. Estas regras são empregadas para realizar seleção de características de valores contínuos.

O StARMiner é apresentado no Algoritmo 3. O algoritmo realiza duas varreduras na base de dados. Na primeira, o valor da média de cada característica é calculado (linhas 1 a 6). Na segunda varredura (linhas 7 a 16), os valores do desvio padrão e do Z-valor são calculados para cada característica. Nas linhas 11 e 12 são processadas as restrições de interesse. Se a regra satisfizer os limiares  $\Delta \mu_{min}$ ,  $\sigma_{max} e \gamma_{min}$  fornecidos como parâmetros de entrada pelo usuário, ela é retornada, caso contrário, ela é descartada. A complexidade do algoritmo StARMiner é  $\Theta(ckN)$ , onde N é o número de instâncias da base, k é o número de características, e c é o número de categorias (Ribeiro et al., 2009a).

É importante observar que este algoritmo é sensível aos limiares da entrada fornecidos pelo usuário, sendo que pequenas variações podem ocasionar resultados bem diferentes em cada execução do algoritmo.

A outra tarefa abordada neste trabalho é a de classificação, apresentada e discutida na próxima seção.

```
Algoritmo 3: Algoritmo StARMiner
   Dados: Base de dados T de tuplas de imagens estruturadas como \{x_j, f_1, f_2, ..., f_n\}
               onde x_i representa a categoria da imagem e f_i uma característica da
               imagem; limiares \Delta \mu_{min}, \sigma_{max} \in \gamma_{min}
   Resultado: As regras mineradas
 1 Percorra a base de dados T;
 2 para cada característica f_i faça
        para cada categoria x_i faça
 3
             calcule \mu_{f_i}(T_{x_i}) \in \mu_{f_i}(T - T_{x_i});
 \mathbf{4}
 \mathbf{5}
        fim
 6 fim
 7 Percorra a base de dados T;
   para cada característica f_i faça
 8
        para cada categoria x_i faça
 9
             calcule \sigma_{f_i}(T_{x_i}) \in \sigma_{f_i}(T - T_{x_i});
10
             calcule o valor Z_{i_i};
11
              se (\mu_{f_i}(T_{x_j}) - \mu_{f_i}(T - T_{x_j})) \ge \Delta \mu_{min} \ e \ \sigma_{f_i}(T_{x_j}) \le \sigma_{max} \ e \ (Z_{i_j} < Z_1 \ ou 
12
             Z_{i_i} > Z_2) então
                 escreva x_j \to f_i, \ \mu_{f_i}(T_{x_j}), \ \mu_{f_i}(T - T_{x_j}), \ \sigma_{f_i}(T_{x_j}), \ \sigma_{f_i}(T - T_{x_j});
\mathbf{13}
             fim
14
        fim
15
16 fim
```

## 4.3 Classificação

Data mining tem sido aplicado com sucesso em diferentes campos do esforço humano, incluindo marketing, operações bancárias, relacionamento com o cliente, engenharia e várias áreas da ciência. Entretanto, suas aplicações à análise de dados médicos tem sido limitada até recentemente. Isto é particularmente verdade em aplicações práticas em clínica médica, que pode se beneficiar de abordagens de mineração de dados específicas que permitem realizar modelagem preditiva, explorar o conhecimento avaliado no domínio clínico e explicar decisões propostas, uma vez que os modelos são usados para apoiar decisões clínicas. O objetivo de mineração de dados preditiva em clínica médica é derivar modelos que podem usar informações específicas do paciente para predizer a saída de interesse e assim apoiar a tarefa de tomada de decisão. Métodos de mineração de dados preditiva podem ser aplicados na construção de modelos de decisão para procedimentos tais como prognóstico, diagnóstico e planejamento de tratamento, o qual, uma vez avaliado e verificado, pode ser inserido em sistemas de informação clínica.

Modelos de predição em clínica médica são ferramentas para auxiliar a tomada de decisão que combina dois ou mais itens de dados do paciente para predizer saídas clínicas" (Wyatt & Altman, 1995; Bellazzi & Zupan, 2008). Tais modelos podem ser usados em diversos contextos clínicos por médicos e podem permitir uma reação de prontidão a situações desfavoráveis. Mineração de dados pode contribuir efetivamente ao desenvolvimento de modelos preditivos clinicamente úteis graças a pelo menos três aspectos inter-relacionados: (a) uma abordagem compreensiva e vantajosa à análise de dados que envolve a aplicação de métodos e abordagens vindas de diferentes áreas científicas; (b) a capacidade de exploração de tais modelos; (c) a capacidade de usar o domínio do conhecimento no processo de análise dos dados.

Modelos preditivos devem modelar probabilidades e, sempre que possível, devem informar intervalos de confiança. Muitos dos métodos de mineração de dados informam as probabilidades da saída. Entretanto, poucas das *toolboxes* de mineração de dados incluem implementações que permitam informar a confiança dos intervalos de probabilidade preditos.

Métodos de mineração de dados preditiva originam de diversos campos de pesquisa e geralmente usam abordagens de modelagem muito diversas. Devem ser comparados na base levando em consideração (Bellazzi & Zupan, 2008):

- tratamento de ruídos e falta de dados;
- tratamento de diferentes tipos de atributos (categórico, ordinal, contínuos);
- a apresentação de modelos de classificação que podem ou não permitir que o especialista no domínio examine e compreenda o trabalho interno;
- a redução do número de testes, i.e., a redução de atributos necessária para derivar a conclusão;
- o custo computacional para indução e o uso de modelos de classificação;
- sua habilidade em explicar as decisões alcançadas quando os modelos são usados na tarefa de tomada de decisão;
- generalização, i.e., a habilidade de executar bem com casos não analisados anteriormente.

Segundo Kantardzic (2011), a classificação de dados ocorre em duas etapas: modelagem e avaliação. Na etapa de modelagem é construído o modelo para descrever a base de dados e na fase de avaliação é verificado se a acurácia do modelo é aceitável para que o classificador possa ser usado para classificar futuros objetos cuja classe seja desconhecida. Maiores detalhes sobre a modelagem e a avaliação de classificadores estão descritos na seção 4.4.

A seguir estão descritos alguns classificadores que são bastante usados na literatura.

### C4.5

O C4.5 é um dos algoritmos mais usados baseado em árvore de decisão para resolver a tarefa de mineração de dados. O C4.5, desenvolvido por Quinlan (1993), basicamente é uma versão melhorada do algoritmo ID3, que também foi proposto por Quinlan (1986). O C4.5 tem características adicionais tais como lidar com valores ausentes, categorização de atributos contínuos (por meio da discretização dos dados), poda de árvores de decisão, derivação de regras e outros (Mazid et al., 2010). O algoritmo constrói, na fase de treinamento, uma grande árvore considerando os valores de todos os atributos da amostra, e finaliza a regra de decisão pela poda. O C4.5 usa uma abordagem de heurística para a poda, baseada na significância estatística das divisões.

A construção básica da árvore de decisão é: (1) Os nós que são raiz estão no topo da árvore, considerando todos as amostras e selecionado os atributos que são mais relevantes. (2) A informação da amostra é passada aos nós subsequentes, os nós filhos, os quais eventualmente terminam em nós folhas, que contém a informação de decisão (resposta final). (3) As regras são geradas pelo caminho que vai do nó raiz ao nó folha.

Um dos maiores desafios dos usuários do C4.5 é trabalhar com uma quantidade de dados muito grande e ter eficiência computacional. Na maioria das vezes é muito difícil lidar com os dados quando a dimensionalidade aumenta enormemente durante o processo de generalização de regra. É necessário considerar algumas questões como profundidade da árvore de decisão, tratamento de dados contínuos, métodos de seleção para escolher os atributos mais significativos, valores ausentes, etc (Mazid et al., 2010).

Os dados contínuos passam por um processo de discretização dos dados. Quanto às medidas usadas para a seleção de atributos, o C4.5 usa o conceito de Informação de Entropia (Wu et al., 2007).

## 1-R

O algoritmo 1R, proposto por Holte (1993), é um classificador baseado em regras que classificam um objeto/imagem em uma base com um único atributo. É considerado como uma árvore de decisão de um nível, e envolve discretização de atributos. No algoritmo de discretização, os limites dos intervalos (pontos de corte) são ajustados de acordo com a informação de classe das instâncias. Entretanto, este método produz um grande número de intervalos que ainda podem ser reduzidos, conforme é mostrado no trabalho de Ribeiro et al. (2008a). Assim, a vantagem deste classificador é que classifica um objeto baseado em uma única regra, a qual utiliza apenas um único atributo, e a desvantagem é que usa discretização dos dados contínuos, o que pode gerar inconsistências na regra gerada.

### Naive Bayes

O classificador *naive Bayes* (Rish, 2001) é um classificador que usa uma abordagem probabilística baseada no teorema de Bayes para predizer a classe. É baseado no princípio de Maximização à Posteriori (MAP). Apesar de sua simplicidade, mantém sua performance comparável com outras abordagens mais sofisticadas (Bellazzi & Zupan, 2008). Quando é superado por outros algoritmos mais sofisticados na performance da predição, isto geralmente indica a presença de interações não lineares entre os atributos.

### AdaBoost

O AdaBoost (*ADAptative BOOSTing*) é um classificador que combina de forma linear um conjunto de funções simples de classificação, chamadas de classificadores "fracos" para construir um classificador "forte". Um classificador forte é composto por um conjunto de classificadores fracos, os quais estão associados a pesos que classificam com maior precisão dois conjuntos de dados com a classe conhecida (pré-rotulados), nos quais as características com pesos maiores são mais significativas para a classificação de exemplos definidos como parte de um certo conjunto. Assim, o classificador AdaBoost (Freund & Schapire, 1997) basicamente seleciona várias amostras aleatoriamente da base e constrói um modelo de classificação para cada amostra. O conjunto de treinamento é mudado baseado no resultado dos modelos anteriores. A classificação final é a classe que tiver maior ocorrência nos outros modelos (Edelstein, 1999).

### 1-NN

O algoritmo dos **k vizinhos mais próximos** (k nearest neighbor - k-NN) é inspirado na abordagem frequentemente tomada pelo domínio dos especialistas que tomam decisões, baseado em casos semelhantes já vistos anteriormente (Hastie et al., 2011; Bellazzi & Zupan, 2008). Um objeto é classificado pela maioria de votos de seus vizinhos, ou seja, ele é classificado pela classe mais comum entre os seus k vizinhos. k deve ser um número inteiro positivo, geralmente pequeno. Quando k = 1, então o objeto simplesmente recebe a mesma classe de seu vizinho mais próximo. Em problemas de classificação binária (duas classes) é interessante que a escolha de k seja um número ímpar, para assim evitar o empate de votos (Phyu, 2009).

Assim, dada uma instância do dado, o classificador dos k-vizinhos mais próximos busca por k instâncias de treinamento mais semelhantes e classifica baseado em sua classe predominante. A busca pelas instâncias mais semelhantes pode ser lenta e requer a recuperação de um conjunto de treinamento completo no momento da classificação. Entretanto, é um dos algoritmos mais usados por causa de ser fácil de ser compreendido e fácil de implementar e, apesar de sua simplicidade, o k-NN gera resultados satisfatórios em muitas situações (Wu et al., 2007). Os métodos listados acima são frequentemente uma parte integral do mais moderno conjunto de mineração de dados e, sozinhos ou em combinação com pré-processamento, geralmente funcionam bem e suficientemente rápidos. Com relação a dados clínico, as maiores diferenças podem surgir na performance da predição e na interpretação dos resultados.

A seguir é discutido o classificador associativo, que é o interesse deste trabalho.

## 4.3.1 Classificação associativa

Recentemente, métodos de classificação associativa têm recebido bastante atenção (Zhang & Zhou, 2012). A classificação associativa integra o descobrimento de regras de associação e classificação para construir um classificador para o propósito de predição.

Classificação associativa é um caso especial de descoberta de regras de associação na qual apenas o atributo da classe é considerado no consequente (lado direito da regra); por exemplo, em uma regra tal que  $X \to Y$ , Y deve ser atributo da classe. Uma das principais vantagens de usar classificação baseada em regras de associação sobre abordagens de classificação clássica é que a saída de um algoritmo de classificação associativa é representada em regras simples "Se-Então", que facilitam a compreensão e interpretação do usuário final. Além disso, diferentemente de algoritmos de árvores de decisão (Mitchell, 1997), o primeiro pode atualizar ou ajustar uma regra em classificação associativa sem afetar o conjunto completo de regras, considerando que a mesma tarefa requer a reconstrução de toda a árvore na abordagem de árvore de decisão.

Logo, pode-se considerar que um classificador associativo é um mapeamento da forma  $H: A \to Y$ , onde A é um conjunto de *itemsets* e Y é o conjunto de classes. A principal tarefa de classificação associativa é construir um conjunto de regras (modelo) que é capaz de predizer as classes de dados não previamente vistos, conhecido como conjunto de teste, tão preciso quanto possível. Ou seja, o objetivo é encontrar um classificador  $h \in H$  que maximize a probabilidade que h(a) = y para cada objeto teste.

E importante observar que uma tarefa de classificação associativa é diferente do descobrimento de regras de associação. A diferença mais óbvia entre elas é que a classificação associativa considera apenas o atributo classe nas regras consequentes. Entretanto, o padrão permite valores de atributos múltiplos nas regras consequentes. A Tabela 4.8 mostra as principais diferenças entre classificação associativa e descoberta de regras de associação, onde a prevenção de super ajustamento é essencial em classificação associativa, mas não em descoberta de regras de associação. Super ajustamento geralmente ocorre quando o algoritmo de descoberta de regras executa bem o conjunto de treinamento e executa mal o conjunto teste. Isto pode ser devido a diversas razões tais como pequeno número de objetos no conjunto de treinamento ou ruído.

**Tabela 4.8:** As principais diferenças entre classificação associativa e descobrimento de regras de associação (Thabtah, 2007)

| Descoberta de regras de associação  | Classificação associativa   |
|---|---|
| Não há atributo classe envolvido (apren-  | Uma classe deve ser dada (aprendizado   |
| dizado não supervisionado)  | supervisionado)   |
| O objetivo é descobrir associações entre<br>itens em uma base de dados transacional | O objetivo é construir um classificador que<br>possa predizer as classes dos objetos do<br>conjunto teste |
| Pode haver mais que um atributo no con-<br>sequente de uma regra                    | Há apenas atributo (atributo classe) no consequente de uma regra  |
| Super ajustamento geralmente não é um problema                                      | Super ajustamento é um problema impor-<br>tante   |

Segundo Christopher (2011), as fases da classificação associativa são geração de regras, poda de regras, ranqueamento ou ordenação de regras, construção do modelo e predição. As duas últimas tarefas juntas são conhecidas como classificação. A geração de regras na classificação associativa é uma tarefa difícil e requer muito processamento. Um rico conjunto de regras é construído após a aplicação de estratégias adequadas de poda e ranqueamento. O conjunto de regras, o qual é gerado a partir de um conjunto de treinamento, é usado para construir o modelo. Este modelo é usado para predizer (classificar) casos de teste presentes na base de dados de treinamento (Christopher, 2011). A Figura 4.3 mostra um *framework* para a classificação associativa.



Figura 4.3: Fases da classificação associativa

Apesar de possuírem diferenças, a classificação associativa apresenta os mesmos desafios da mineração de regras de associação que são a desvantagem de requerer a discretização dos dados contínuos antes de aplicar a mineração das regras de associação; requerem que o usuário forneça limitares como parâmetro de entrada do algoritmo, o que pode ser uma tarefa muito árdua e exigir algum grau de *expertise* do usuário; e a geração de muitos *itemsets* candidatos. Pesquisas recentes têm proposto novas abordagens para lidar com estes problemas. Por exemplo, em Christopher (2011), uma abordagem estatística para classificação associativa tem sido empregada para reduzir o número de *itemsets* candidatos. Em Das & Vyas (2010) é apresentado um estudo de métodos de discretização para classificadores associativos.

Em Baralis & Garza (2012) é proposto o algoritmo I-prune, que é uma abordagem de poda de itens. O método seleciona os itens desinteressantes considerando o valor de uma medida interessante e os poda e deleta. Assim, o número de regras extraídas é reduzido e o tempo de geração do modelo diminui.

Outras variações de classificadores associativos estão descritas em Antonie et al. (2009). Essas variações consideram a mineração de conjuntos de dados com itens recorrentes, usando regras de associação negativa, e regras de poda usando técnicas baseadas em grafos. Neste trabalho os autores ainda apresentam um sistema, o ARC-UI que possui uma interface que apresenta ao usuário o relacionamento entre os atributos, regras e classes do conjunto de regras e analisa o conjunto de treinamento com respeito ao item a ser classificado.

Heravi & Zaïane (2010) apresentaram 53 medidas interessantes para classificadores associativos, além das medidas de suporte e confiança, e realizaram experimentos usando várias bases de dados do repositório UCI, observando o número de regras geradas e a acurácia do classificador. Eles chegaram à conclusão de que não há uma única medida que proporcione resultados interessantes para todas as bases de dados.

Na última década, diversos trabalhos têm aplicado ou desenvolvido classificadores associativos para sistemas de apoio ao diagnóstico. Para classificação e detecção de câncer de mama, Dua et al. (2009) usou regras de associação ponderadas para a classificação de mamogramas. Foram extraídas características de textura de regiões de interesse e os valores foram discretizados para a descoberta de regras. Regras de associação foram derivadas entre vários componentes de textura a partir da imagem segmentada e empregada para a classificação baseada em suas dependências intra e inter classe. Ribeiro et al. (2008b) propuseram o método IDEA (*Image Diagnosis Enhancement through Association Rules*) para classificar regiões de interesse de mamografias e para classificar os tecidos das mamas. Este método incorpora um algoritmo que realiza simultaneamente a seleção de características e a discretização, e um algoritmo de classificação associativa, chamado ACE (*Associative Classifier Engine*), o qual sugere múltiplas palavras-chave para compor o diagnóstico de uma imagem. Já Ganesan et al. (2011) apresentaram uma revisão de classificadores associativos desenvolvidos para a tarefa de auxílio ao diagnóstico de câncer de mama.

Jose et al. (2012) propuseram o uso de um classificador associativo para auxiliar no diagnóstico de rim, em rim normal, com cisto cortical ou *medical renal*. Primeiramente o

método seleciona e discretiza as características. Depois, o algoritmo Apriori-T (Agrawal & Srikant, 1994) é usado para gerar as regras. Por fim, as regras são fornecidas a um classificador bayesiano para a classificação das imagens.

Rajendran & Madheswaran (2012) desenvolveram um sistema de auxílio ao diagnóstico de tumor em cérebro, considerando as classes normal, benigno e maligno. Primeiramente as características são extraídas e discretizadas. Eles propuseram uma adaptação do algoritmo Apriori para gerar as regras, realizar a poda e construir o classificador. Os resultados experimentais de pré-diagnóstico mostraram até 97% de sensibilidade, 91% de especificidade e 98,5% de acurácia.

Pode-se observar que a maioria dos classificadores associativos propostos usam discretização de dados e muitos usam extensões ou o próprio algoritmo Apriori, o que necessita que os usuários forneçam como parâmetro de entrada limiares como suporte e confiança.

Um outro aspecto importante da mineração de imagens é a avaliação e interpretação dos resultados, que estão apresentadas na próxima seção.

# 4.4 Avaliação e interpretação dos resultados

Dentre os estudos realizados para a realização deste trabalho, observou-se que há uma interseção de métodos de avaliação usados tanto para avaliação, organização e seleção de sistemas de diagnóstico e/ou predição quanto na área de aprendizagem de máquina e mineração de dados, para as quais os mesmos métodos são usados como uma ferramenta útil e poderosa para a avaliação de modelos de classificação.

Uma importante regra em construção e avaliação de modelos preditivos (classificação) e de execução de sistemas CAD é que estes nunca devem ser construídos e testados sobre o mesmo conjunto de dados. Para isto, técnicas como validação cruzada (*cross-validation*) são usadas, mas também pode-se dividir o conjunto de dados em dois: o primeiro, geralmente chamado de **conjunto de aprendizagem**, ou treinamento, é usado para comparar diferentes algoritmos de mineração de dados, estimar sua performance usando algumas métricas estatísticas, encontrar o melhor conjunto de parâmetros para categorização de características, seleção e métodos de aprendizagem e selecionar a técnica de modelagem que obteve a melhor performance. Usando esta técnica, um modelo final é desenvolvido a partir de um conjunto de aprendizagem completo e testado em um segundo conjunto de dados, geralmente chamado **conjunto de validação** ou teste. A divisão dos dados pode ser arbitrária ou baseada em tempo ou na classe das instâncias dos dados (Bellazzi & Zupan, 2008). O propósito do teste é acessar a generalização e robustez dos sistemas.

Para separar os conjuntos de aprendizagem e validação é necessário estimar a performance da predição. Modelos de mineração de dados podem ser complexos e em casos extremos podem "relembrar" cada instância dos dados que aprenderam. Tais modelos

executam perfeitamente sobre os dados que foram usados para a aprendizagem, mas não com um novo caso que não se iguale a alguma das instâncias dos dados do conjunto de aprendizagem. É dito que tais modelos não tiveram uma boa generalização devido a um super ajustamento dos dados (*overfitting*). As técnicas de mineração de dados mais recentes incluem mecanismos eficientes para evitar o super ajustamento, como poda em árvores de decisão, limitar a complexidade para as redes neurais, e apenas a seleção de regras mais significantes para modelagem de regras de decisão. Entretanto, apenas a avaliação de um conjunto de dados independente pode garantir que a boa performance não resulte em super ajustamento.

Modelos preditivos podem ser avaliados com base em sua performance preditiva e compreensibilidade. Performance preditiva é mais fácil de quantificar e estatísticas típicas incluem métricas tais como sensibilidade, especificidade e acurácia da classificação. Compreensibilidade é uma medida subjetiva que é calculada participando do domínio dos especialistas. Como isto pode ser muito difícil, os modelos geralmente preferem encontrar respostas para questões do tipo:

- Dado dois modelos, qual é o mais fácil de ser entendido?
- Qual deles explica melhor as decisões?
- Qual deles faz com que os especialistas tenham mais confiança em usar?

Se compreensibilidade e explicação são levados em consideração, os algoritmos de mineração de dados podem ser classificados primeiramente usando a performance estatística preditiva e, então, os especialistas no domínio podem selecionar o modelo final, dentre os melhores classificados, baseado em sua compreensibilidade e habilidade de explicação.

Como mencionado anteriormente, para estimar essas estatísticas que avaliam a performance preditiva, uma abordagem vantajosa é aplicar a estratégia hold-out: um subconjunto para o conjunto de aprendizagem, o conjunto de treinamento, é usado para construir o modelo, enquanto outro subconjunto, o conjunto de teste, é usado para estimar a acurácia do modelo. Entretanto, o procedimento de hold-out é um pouco ineficiente. Uma estratégia típica é treinar com dois terços dos dados do conjunto e então testar com o um terço restante. Tal estratégia pode não ser aplicável a um conjunto com poucos dados pois os algoritmos para aprendizagem do modelo prognóstico pode ter problemas devido ao reduzido conjunto para treinamento, enquanto que o conjunto teste pode ser ainda insuficiente para alcançar os limites do intervalo de confidência desejados. Um método para tratar esses problemas é a k-fold cross-validation. Com a validação cruzada, os dados são divididos em um número k de subconjuntos os quais contêm aproximadamente um número igual às instâncias dos dados e aproximadamente combinam a distribuição da saída do conjunto de treinamento (validação cruzada estratificada). Tipicamente, o conjunto de treinamento é dividido em dez subconjuntos (10-fold cross validation). Então,
os dados dos nove subconjuntos são usados para a modelagem enquanto o subconjunto restante é usado para testar o modelo resultante e avaliar a estatística. O processo de treinamento e teste é repetido dez vezes, sendo que cada vez é usado um subconjunto diferente para o teste. Médias estatísticas são então feitas e caracterizam o método modelado. Além da validação cruzada, outras abordagens de divisão dos dados podem ser usadas tais como *'leave-one-out' cross validation* (onde k = 1 na validação cruzada), random sampling, bootstrap, dentre outros (Han & Kamber, 2006; Bellazzi & Zupan, 2008; Arimura et al., 2009).

Uma atenção especial deve ser dada aos parâmetros de estimação. A maioria dos métodos de mineração de dados dependem de um conjunto de parâmetros que definem o comportamento do algoritmo de aprendizagem e, diretamente ou indiretamente, influenciam na complexidade dos modelos resultantes. Por exemplo, o grau de poda pode ser feito por indução de árvores de decisão, o número de unidades na camada oculta pode ser ajustada para feed-forward nos modelos de redes neurais e o nível requerido da significância estatística pode ser definido por regras de decisão. Enquanto a descoberta do melhor conjunto de parâmetros pode ser caracterizada como uma busca no espaço de parâmetros que aplica algumas otimizações em técnicas do estado da arte, profissionais geralmente definem um conjunto dos valores mais prováveis de parâmetros e, novamente por meio da validação cruzada, avaliam cada conjunto separadamente para encontrar o melhor. A avaliação dos métodos de mineração de dados, então, permite não apenas a classificação das técnicas de mineração de dados, mas também identifica o conjunto apropriado de parâmetros a ser usado. Observe também que a classificação das características, seleção de um subconjunto e construção podem ter seus próprios parâmetros, o que também requer otimização (Bellazzi & Zupan, 2008).

A seguir são descritos alguns métodos de avaliação que visam medir a efetividade dos resultados de classificação em mineração de dados e sistemas CAD.

#### Método taxa do erro

Com esse método, o classificador simplesmente prediz a classe de um dado do conjunto teste. Se for correta, será contado como sucesso, caso contrário, isso será contado como um erro. A taxa do erro sobre o conjunto teste mede a acurácia preditiva do classificador e pode ser calculada de acordo com a Equação 4.8.

$$Taxa-erro = \frac{n \text{úmero de casos de erro}}{n \text{úmero total de casos no conjunto teste}}$$
(4.8)

Existem várias situações em que a taxa de erro de classificação não é apropriada para a avaliação de modelos de classificação (Prati et al., 2008). Uma situação comum se dá quando o número de exemplos em cada uma das classes é muito desbalanceado. Além disso, a taxa de erro assume custos iguais para os erros tanto para a classe majoritária quanto para a classe minoritária. Em medicina, o custo de classificar incorretamente um paciente doente como sadio para uma dada doença grave é muito maior do que classificar um paciente sadio como doente pois, no primeiro caso, a falha no diagnóstico pode levar à morte do paciente.

#### Matriz de confusão

Para classificação binária, que leva em conta duas classes no conjunto de treinamento, um método comum é chamado matriz de confusão. Este método leva em consideração o custo de predições erradas. A matriz de confusão é semelhante aos métodos de precisão e revocação, os quais contêm informação sobre classificações preditas e atuais feitas pelo classificador (Thabtah, 2007).

Tabela 4.9: Matriz de confusão

|        |          | Prec     | lito     |
|--------|----------|----------|----------|
|        |          | Negativo | Positivo |
| Classe | Negativo | TN       | FP       |
|        | Positivo | FN       | TP       |

A Tabela 4.9 representa uma matriz de confusão, a qual contém informações sobre classificações atuais e preditas feitas por um classificador. **TP** corresponde aos *verdadeiros-positivos* que representam o número de casos em que um objeto é positivo. **FN** (falso-negativo) representa o número de predições incorretas que um objeto é negativo. **TN** (verdadeiro-positivo) representa o número de predições corretas que um objeto é negativo. **FP** (falso-positivo) representa o número de predições incorretas que um objeto é negativo. **FP** (falso-positivo) representa o número de predições incorretas que um objeto é positivo. Baseado na matriz de confusão apresentada na Tabela 4.9, a seguir são apresentadas algumas medidas considerando duas classes.

A **acurácia**, ou **precisão total** é a proporção dos casos no conjunto teste que foram corretos.

$$\operatorname{acurácia} = \frac{TN + TP}{TN + FP + FN + TP}$$
(4.9)

A **sensibilidade**, ou precisão da classe positiva, é a proporção dos casos positivos que foram corretamente classificados.

sensibilidade = 
$$\frac{TP}{FN + TP}$$
 (4.10)

A taxa de erro da classe negativa (TFP) é a proporção de casos negativos que foram classificados incorretamente como positivos:

$$TFP = \frac{FP}{TN + FP} \tag{4.11}$$

A **especificidade**, ou precisão da classe negativa, é definida como a proporção de casos negativos que foram classificados corretamente:

$$especificidade = \frac{TN}{TN + FP}$$
(4.12)

A taxa de erro da classe positiva (TFN) é a proporção de casos positivos que foram incorretamente classificados como negativos:

$$TFN = \frac{FN}{FN + TP} \tag{4.13}$$

É importante ressaltar que os valores ótimos devem alcançar 100% de sensibilidade e 100% de especificidade.

# 4.5 Considerações finais

Neste capítulo foram discutidos os principais tópicos de mineração de dados e de imagens explorados neste trabalho. A mineração de imagens é um processo complexo, envolvendo o tratamento das imagens, como processamento e extração de características (abordados no Capítulo 2), e a etapa de descoberta de conhecimento a partir dos dados. O foco deste trabalho é a tarefa de associação, usada tanto para extrair padrões quanto para a etapa de classificação. Esta tarefa foi escolhida devido ao fato de as regras por ele geradas serem intuitivas.

A partir do próximo capítulo são apresentados os métodos desenvolvidos nesta tese, os quais abordam métodos de representação de imagens e classificadores associativos.

Capítulo

# Desenvolvimento de extratores de características de forma usando redes complexas

Como foi visto anteriormente no Capítulo 4, a primeira fase de um sistema de mineração de imagens é a preparação da base de dados de imagens e a extração de suas características para representação de imagens.

A literatura apresenta vários métodos para analisar imagens e objetos usando o contorno da forma, como é descrito no Capítulo 2, e a base da maioria deles considera o contorno como uma cadeia de pontos conectados. Nesta abordagem, a sequência de pontos na fronteira é importante, já que é usada para extrair o descritor de forma que seja capaz de caracterizar a forma da imagem.

Backes et al. (2009), Backes (2010) e Casanova et al. (2013) propõem o uso de redes complexas para caracterizar o contorno da imagem, considerando o contorno como um conjunto de pontos e modela este conjunto como um grafo. Dessa maneira, o método por eles proposto não necessita que os *pixels* sejam adjacentes e sequenciais, já que o modelo de grafo apenas leva em consideração a distância entre os elementos do contorno. As características topológicas são derivadas a partir da dinâmica do crescimento da rede, as quais estão relacionadas aos aspectos físicos da rede. A desvantagem deste método é que a rede precisa ser recriada para cada 'evolução dinâmica', a partir de diferentes limiares incrementais.

Neste capítulo é apresentado um novo extrator de características que usa redes complexas e cria apenas um modelo de grafo para cada imagem. Primeiramente a imagem é segmentada e seu contorno é modelado em uma rede livre de escala. Medidas topológicas da rede criada são usadas para compor o vetor de características. Aqui foram propostos quatro vetores de características, baseados nas medidas de grau do vértice, força do vértice e grau de junção. Cada um dos descritores propostos foram analisados e foi verificado quais medidas topológicas extraídas caracterizaram melhor as imagens.

A inovação deste método está em modelar o contorno da imagem em uma rede complexa livre de escala e a grande vantagem deste novo extrator é que o vetor de característica produzido é pequeno, quando comparado a vetores produzidos usando momentos de Zernike e texturas de Haralick, por exemplo. A dimensão do vetor depende do número de medidas que irão ser extraídas da rede, sendo que os experimentos realizados mostraram que um vetor composto por apenas duas características é suficiente para caracterizar bem as imagens. Além disso, também há um aumento na precisão da classificação de imagens com o uso do novo descritor. Dessa maneira, o novo extrator lida com o problema da maldição da dimensionalidade. Os experimentos também mostram que os descritores gerados tendem a ser invariantes à escala, rotação e a pequenas deformações.

Para a análise do método desenvolvido são apresentados quatro estudos de caso, sendo o primeiro relativo a objetos em diversas rotações; o segundo é constituído de objetos em diversas rotações, escalas e também com pequenas deformações não-rígidas; o terceiro estudo de caso inclui uma base de objetos diversos; e o quarto e último estudo de caso envolve uma base real de regiões de interesse de massas benignas e malignas em mamografias digitais, que serviram para desenvolver um protótipo de um sistema de auxílio ao diagnóstico de câncer de mama, que é o objetivo deste trabalho.

Vale ressaltar que parte dos resultados apresentados neste capítulo, relativos ao estudo de caso de aplicação do extrator de característica desenvolvido em um conjunto de dados médicos, foram publicados na forma de artigo científico no *SPIE Medical Imaging 2012* (Watanabe et al., 2012b).

A seguir, está descrito o extrator de características proposto.

# 5.1 Descrição do método: modelando imagens como uma rede livre de escala

Na literatura, são realizadas propostas de modelagem de imagens usando redes complexas, mas nenhuma delas representa imagens como um modelo livre de escala. O método aqui proposto apresenta a modelagem de uma imagem em uma rede complexa livre de escala, a partir de seu contorno e de seu esqueleto. O Algoritmo 4 apresenta os passos para realizar esta modelagem.

| Algoritmo 4: Modelagem da imagem em uma rede complexa livre de escala   |
|---|
| <b>Dados</b> : Imagem segmentada binarizada                             |
| <b>Resultado</b> : Imagem modelada em uma rede complexa livre de escala |
| 1 Obter o contorno C com a aplicação do filtro laplaciano 3x3;          |
| 2 Obter o esqueleto S;  |
| $\mathbf{s}$ para cada ponto $s_i$ do esqueleto faça                    |
| 4   Encontrar o ponto $c_i$ de C mais próximo de $s_i$ ;                |
| 5 Atualizar a matriz de adjacências ponderada e não direcionada:        |
| $W(i,j) = W(j,i) = d(s_i, c_j);$  |
| 6 fim   |
| 7 retorna A matriz de adjacências W                                     |

A entrada do algoritmo é uma imagem previamente segmentada e binarizada. Os métodos de binarização e segmentação aplicados devem ser adequados a cada tipo de imagem. Depois de segmentada e binarizada a imagem, o primeiro passo na modelagem é obter o esqueleto e o contorno da forma da imagem binarizada (linhas 1 e 2 do algoritmo). O contorno C (linha 1) é obtido pela convolução da imagem binarizada com um filtro laplaciano 3x3, o qual é um operador escalar baseado na derivada de 2<sup>a</sup> ordem usado para detecção de bordas:

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$
 (5.1)

Na linha 2, o esqueleto<sup>1</sup> é obtido por meio da remoção dos *pixels* do contorno dos objetos, mas não permitindo que os objetos se 'quebrem'. Os pixels que restam formam o esqueleto da imagem. Assim, os objetos são reduzidos a linhas, sem mudar a essência estrutural da imagem.

O segundo passo é realizar a modelagem de uma rede complexa livre de escala a partir do contorno e do esqueleto da imagem (linhas 3, 4, 5 e 6 do Algoritmo 4). Considere que o contorno seja representado como  $C = [c_1, c_2, ..., c_n]$ , onde cada  $c_j$  é um vetor na forma  $c_i = [x_i, y_i]$ , cujos componentes são valores numéricos discretos representando as coordenadas do ponto j do contorno. Seja  $S = [s_1, s_2, ..., s_m]$  os pontos do esqueleto, também representando as coordenadas do ponto i do esqueleto.

A rede é criada da seguinte maneira. Primeiro, a rede é iniciada com todos os pontos do esqueleto. Considere que todos os pontos da rede tem grau 0, i.e., inicialmente não há arestas na rede, apenas nós. Então, cada ponto  $s_i$  do esqueleto S é conectado ao ponto  $c_j$  mais próximo de  $s_i$ . Neste trabalho foi usada a distância euclidiana para encontrar o ponto mais próximo, mas outras funções de distância podem ser usadas. Dessa forma,

 $<sup>^1 \</sup>rm Neste trabalho foi usada a implementação de obtenção do esqueleto que está descrita em Gonzalez et al. (2003) e Zhang & Suen (1984).$ 

uma matriz de adjacências ponderada não direcionada é calculada como:

$$W_{i,j} = W_{j,i} = d(s_i, c_j), 1 \le i \le |S|,$$
(5.2)

onde  $d(s_i, c_j)$  é a distância euclideana do ponto  $s_i$  ao seu vizinho mais próximo no contorno, o ponto  $c_j$ , e |S| é o número de pontos do esqueleto.

Observe que esta construção segue as duas regras básicas do modelo de rede de Barabási-Albert: crescimento e ligação preferencial. O crescimento se dá com a adição de novos nós do contorno a cada iteração, e a ligação preferencial é caracterizada pelos nós do contorno mais próximos dos pontos do esqueleto. A rede construída terá no máximo 2 \* |S| nós e terá exatamente 2 \* |S| arestas. Mas observe que, pela regra de ligar um ponto do esqueleto ao ponto mais próximo do contorno, nem todos os pontos do contorno farão parte da rede, e, além disso, alguns pontos do esqueleto serão ligados em um mesmo ponto no contorno. Então, a rede sempre terá um número maior de arestas do que nós, o que leva à criação de *hubs*. A Figura 5.1 mostra um exemplo da rede complexa livre de escala gerada pelo método proposto. É possível observar que alguns nós estão altamente conectados enquanto outros têm poucas conexões. Em particular, os pontos do esqueleto estão conectados a apenas um ponto no contorno, enquanto que os *hubs* que aparecem na rede estão presentes nos pontos do contorno.



Figura 5.1: Rede complexa livre de escala

## 5.1.1 Descritores

Após modelada a rede complexa livre de escala, medidas topológicas da rede podem ser extraídas para comporem os vetores de características. Neste trabalho foram propostos descritores baseados no grau do vértice (nó), na força do vértice e no grau de junção.

#### Descritor: Grau do Vértice - Grau

Este descritor é composto pelas medidas de grau máximo  $k_{\kappa}$  (Eq. 3.5) e grau médio  $k_{\mu}$  (Eq. 3.6). Entretanto, é necessário realizar a normalização do grau dos nós pelo número de nós na rede (N). Esta normalização é realizada para reduzir a influência do tamanho da rede sobre os descritores. Assim, este vetor de características é composto por apenas duas medidas  $[k_{\kappa}, k_{\mu}]$ .

#### Descritor: Força do Vértice - Força

Este vetor de características é composto pelas medidas de força máxima do nó ms (Eq. 3.8) e pela força média do nó as (Eq. 3.9). Logo, este vetor tem apenas duas características.

#### Descritor: Grau de Junção - Junção

Este vetor de características é composto pela entropia H (Eq. 3.10), pela energia E (Eq. 3.11) e pela média do grau de junção P (Eq. 3.12). Essas medidas foram calculadas a partir da rede livre de escala artificialmente modelada, como descrita no início desta seção. Assim, o vetor de características tem três atributos: [H, E, P].

#### Descritor: Grau-Força-Junção - Completo

Este descritor é composto pela junção dos três descritores apresentados anteriormente. Assim, seu respectivo vetor de características possui sete atributos:  $[k_{\mu}, k_{\kappa}, ms, as, H, E, P]$ .

# 5.2 Experimentos

Nesta seção são apresentados 3 estudos de caso para mostrar que o modelo proposto gera redes complexas livres de escala e também para mostrar a eficiência dos descritores propostos. Cada estudo de caso está descrito da seguinte maneira:

- Descrição da base: é apresentada uma breve descrição da base de imagens, contendo o número de imagens por classe, o local onde ela foi adquirida e algumas características visuais.
- Avaliação da rede complexa: nesta etapa são analisadas as redes complexas geradas pelo modelo proposto de algumas imagens da base de imagens em estudo. Nestas redes, busca-se a presença de *hubs* (nós com alto número de arestas). Também é gerado e analisado o gráfico plotado em uma escala log-log do número de arestas pelo número de nós (descrito na Seção 3.2.3). Neste gráfico é analisada a distribuição das arestas, e, para ser uma rede livre de escala, precisa mostrar que

a maioria dos nós têm poucas conexões e alguns têm um grande número de arestas. A aparência deste gráfico deve tender a uma reta para redes com muitos nós.

• Análise dos descritores: para avaliar a eficiência dos descritores propostos (Grau, Junção, Força, Completo), as características das imagens são extraídas e é usada a abordagem 10-fold cross validation para as fases de treinamento e teste para a classificação das imagens. O classificador usado foi o C4.5, o qual constrói uma árvore de decisão na fase de treinamento. As medidas avaliadas foram a taxa de acerto (precisão) da classificação total das imagens e a taxa de acerto de cada classe individualmente. Nos casos onde a classificação era binária (para a base de Mama, cuja classificação é massa maligna ou benigna), foram avaliadas também as medidas de sensibilidade e especificidade. Essas medidas estão descritas na Seção 4.4.

Um descritor topológico simples, o qual foi denominado "Forma" nos experimentos realizados, foi desenvolvido para comparação com os descritores propostos. Após a binarização da imagem e a detecção da região de interesse, foram extraídos da ROI de cada imagem 11 valores para comporem o vetor de características: área, tamanho do eixo maior, tamanho do eixo menor, excentricidade, orientação, área convexa, área preenchida, número de Euler, solidez, amplitude e perímetro, cujas definições estão apresentadas na Seção 2.2.3.

O pipeline do *framework* desenvolvido para a avaliação do método proposto está apresentado na Figura 5.2. Considerando primeiramente a fase de treinamento, inicia-se com um conjunto de imagens. Para cada imagem, são realizados os seguintes passos. Primeiro é realizada a segmentação e a binarização da imagem. Depois, são extraídos o contorno e o esqueleto da imagem. A partir do contorno e do esqueleto, a imagem é modelada em uma rede complexa livre de escala. Após a modelagem, são extraídas medidas topológicas da rede criada, as quais serviram para compor o vetor de características. Estes vetores de características extraídos das imagens do conjunto de treinamento são usados para a construção de uma árvore de decisão. Na fase de teste, uma nova imagem, diferente das que foram usadas na fase de treinamento, é processada da mesma maneira que na fase anterior, passando pela fase de segmentação e binarização, extração do contorno e do esqueleto e modelagem da rede complexa livre de escala. O vetor de característica obtido é então submetido à árvore de decisão construída na fase anterior e a saída do sistema é a classificação da nova imagem ou , no caso de um sistema de auxílio ao diagnóstico, uma sugestão de diagnóstico.

Os descritores desenvolvidos foram comparados com outros descritores da literatura: momentos de Zernike e estatísticas de Haralick, além do descritor de Forma descrito anteriormente. Os vetores de características gerados pelos momentos de Zernike foram obtidos pelos polinômios de Zernike de ordem 30, compondo um vetor de características de 256 atributos. Já os descritores de Haralick foram compostos por 140 elementos obtidos



Figura 5.2: Pipeline do *framework* desenvolvido.

segundo está detalhado em Felipe et al. (2003); Ribeiro (2008). Primeiramente, os níveis de cinza das imagens foram reduzidos para 16 níveis. Uma matriz de co-ocorrência foi gerada para cada imagem, para as direções de  $0^{\circ}$ ,  $45^{\circ}$ ,  $90^{\circ}$  e  $135^{\circ}$ , e para as distâncias 1, 2, 3, 4 e 5. Logo, foram produzidas vinte matrizes de  $16 \times 16$  elementos inteiros por imagem. Para cada matriz, foram calculados sete valores, descritos na Tabela 5.1, gerando assim, os 140 atributos.

| Tabela 5.1: | Descritor   | de   | texturas   | de  | Haralick   | е   | posições    | $\operatorname{dos}$ | valores | no | vetor | de |
|-------------|-------------|------|------------|-----|------------|-----|-------------|----------------------|---------|----|-------|----|
|             | característ | tica | (Felipe et | al. | , 2003; Ri | bei | iro, 2008). |                      |         |    |       |    |

| Nome                 | Equação  | Significado       | Posição   |
|----------------------|--|-------------------|-----------|
| Step                 | $\sum_{i} \sum_{j} P(i,j)$                             | distribuição      | 1-20      |
| Variância            | $\sum_{i} \sum_{j} (i-j)^2 P(i,j)$                     | constraste        | 21 - 40   |
| Entropia             | $\sum_{i} \sum_{j} \check{P}(i,j) \log(P(i,j))$        | suavidade         | 41-60     |
| Energia              | $\sum_{i}\sum_{j}P(i,j)^{2}$                           | uniformidade      | 61-80     |
| Homogeneidade        | $\sum_{i} \sum_{j} \frac{P(i,j)}{1+ i-j }$             | homogeneidade     | 81-100    |
| $3^o$ Momento        | $\sum_{i} \sum_{j} (i-j)^3 P(i,j)$                     | distorção         | 101 - 120 |
| Inversa da Variância | $\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{P(i,j)}{(i-j)^2}$ | contraste inverso | 121-140   |

## 5.2.1 Estudo de caso 1: Base ALOI

Para a validação deste novo extrator foram usadas bases de imagens heterogêneas, no sentido de melhor validá-lo. Essa base de imagens foi coletada da Amsterdam Library of Object Images  $(ALOI)^2$ . Essas imagens já estão binarizadas e cada classe possui um

<sup>&</sup>lt;sup>2</sup>http://staff.science.uva.nl/~aloi/

objeto sob diversos ângulos de visão, obtidos por rotações em um plano. Para maiores detalhes desta base, ver Geusebroek et al. (2005).

A base ALOI preparada para este trabalho consiste em 222 imagens, sendo que cada classe possui 74 imagens. Todas as imagens possuem resolução de  $192 \times 144$  pixels. A Figura 5.3 apresenta um exemplo de imagem de cada classe.



Figura 5.3: Exemplos de imagens da base ALOI para a classe 1, 2 e 3, respectivamente.

#### Avaliação da rede complexa gerada

As Figuras 5.4, 5.5 e 5.6 mostram a rede complexa gerada para as Figuras 5.3(a), 5.3(b) e 5.3(c), respectivamente, e o gráfico na escala log log do número de arestas vs. o número de nós.



**Figura 5.4:** Imagem da classe 1. (a) Modelagem da rede complexa e (b) gráfico na escala log.

O modelo proposto gerou uma rede com 606, 833 e 447 nós, respectivamente para as classes 1, 2 e 3, e os gráficos correspondentes apresentam a distribuição dos nós em relação ao número de arestas. Pelos gráficos apresentados, é possível ver que muitos nós possuem poucas arestas e poucos nós possuem muitas arestas. Além disso, quanto maior o número



**Figura 5.5:** Imagem da classe 2. (a) Modelagem da rede complexa e (b) gráfico na escala log.



**Figura 5.6:** Imagem da classe 3. (a) Modelagem da rede complexa e (b) gráfico na escala log.

de nós na rede, mais ela se parece com um reta. Isso mostra que a rede gerada realmente é uma rede complexa livre de escala.

#### Classificação

Como esta base já está binarizada, foram comparados apenas descritores de forma, já que não é possível considerar nem a cor e nem a textura das imagens. Para a comparação com os descritores de redes complexas, foram usados o descritor de Zernike e o descritor de forma simples, previamente descrito no início da Seção 5.2. A Tabela 5.2 mostra a comparação dos resultados.

| Descritores | Precisão Total | Classe 1 | Classe 2 | Classe 3 |
|-------------|----------------|----------|----------|----------|
| Grau        | 0.9550         | 0.9460   | 0.9190   | 1.0000   |
| Junção      | 0.8780         | 0.8650   | 0.8780   | 0.8920   |
| Força       | 0.9190         | 1.0000   | 0.8510   | 0.9050   |
| Completo    | 0.9860         | 1.0000   | 0.9590   | 1.0000   |
| Forma       | 0.9730         | 0.9860   | 0.9590   | 0.9730   |
| Zernike     | 0.9820         | 0.9860   | 0.9590   | 1.0000   |

Tabela 5.2: Base ALOI.

Analisando os resultados da Tabela 5.2, pode-se observar que o descritor Completo, baseado em redes complexas, foi o que obteve o melhor desempenho na tarefa de classificação das imagens. Este descritor conseguiu classificar com 100% de precisão os objetos da classe 1 e 3. Além disso, considerando o número de características geradas, este descritor possui apenas 7 elementos, enquanto que o descritor de Zernike, o qual gerou resultados bem próximos do descritor completo, possui 256 características.

A seguir está a árvore de decisão gerada na fase de treinamento usando o descritor Completo:

força média 
$$\leq 0.486218$$
  
grau médio  $\leq 0.527943:2$   
grau médio  $> 0.527943:3$   
força média  $> 0.486218:1$ 
(5.3)

Analisando a árvore de decisão criada, pode-se observar que foram selecionadas as características força média e grau médio para a construção da árvore. Assim, com apenas duas características, é possível caracterizar bem as imagens dessa base, e a abordagem de descritor usando redes complexas livres de escala se mostrou então adequada para a caracterizar as imagens na tarefa de classificação.

Como esta base de imagens possui as imagens em diversas rotações, pode-se inferir dos resultados que o descritor baseado em redes complexas que usa as medidas topológicas de força média e grau médio é pouco sensível à rotação, ou seja, tende a ser invariante à rotação.

#### 5.2.2 Estudo de caso 2: Objetos

Esta base de imagens foi coletada do grupo MPEG7 ( $MPEG7 \ CE \ Shape-1 \ Part \ B$ )<sup>3</sup>. Para este estudo, foram escolhidas 8 classes de imagens, sendo que as classes de 1 a 7 possuem

 $<sup>^{3}{\</sup>rm Imagens}$  disponíveis em http://www.imageprocessingplace.com/root\_files\_V3/image\_databases.htm

20 elementos cada, e a classe 8 possui 16 elementos, totalizando 376 imagens. Uma característica interessante desta base é que seus objetos são apresentados com pequenas deformações não-rígidas, e em diferentes escalas e rotações. Desta maneira, é possível avaliar se o método proposto é invariante à rotação, escala e pequenas deformações. A Figura 5.7 apresenta um exemplo de imagem de cada classe e a Figura 5.8 apresenta alguns exemplos de 3 classes de imagens diferentes.



Figura 5.7: Exemplo de objetos da base MPEG7.



Figura 5.8: Exemplos de figuras de algumas classes.

É importante ressaltar também que as imagens não possuem a mesma resolução. Uma imagem da classe 8 (carros), por exemplo, pode ter uma resolução de  $1110 \times 488$  pixels enquanto outra da mesma classe pode ser de  $309 \times 166$  pixels.

#### Avaliação da rede complexa gerada

A seguir são apresentadas as redes complexas e os gráficos na escala log do número de arestas vs. o número de nós de três imagens de classes diferentes. A resolução da imagem e o número de nós da rede gerada para cada uma destas imagens estão apresentados na Tabela 5.3.

Analisando os gráficos das imagens, pode-se observar que todas apresentam muitos nós com poucas ligações (arestas) e poucos nós com muitas ligações, caracterizando, então, a presença de *hubs*. Além disso, comparando o número de nós das redes complexas geradas,

| Figura      | Resolução da imagem | Número de nós da rede |
|-------------|---------------------|-----------------------|
| Figura 5.9  | $256 \times 256$    | 994                   |
| Figura 5.10 | $258 \times 442$    | 6158                  |
| Figura 5.11 | $352 \times 288$    | 964                   |

Tabela 5.3: Dados das imagens processadas



Figura 5.9: (a) Modelagem da rede complexa e (b) gráfico na escala log. A imagem binarizada possui uma resolução de  $256 \times 256$  pixels e a rede gerada possui 994 nós.



Figura 5.10: (a) Modelagem da rede complexa e (b) gráfico na escala log. A imagem binarizada possui uma resolução de 2258×442 pixels e a rede gerada possui 6158 nós.

observa-se que a rede que possui mais nós (6158 nós) é a que mais possui uma aproximação de reta. Estes fatos mostram que as redes geradas são livres de escala.



Figura 5.11: (a) Modelagem da rede complexa e (b) gráfico na escala log. A imagem binarizada possui uma resolução de 352 × 288 pixels e a rede gerada possui 964 nós.

#### Classificação

Para esta base foram comparados com os descritores propostos os descritores de Forma e Zernike. A Tabela 5.4 mostra a comparação dos resultados usando o classificador C4.5 (árvore de decisão).

| Descritores | PT    | C1   | C2   | C3   | C4   | C5    | C6   | C7   | C8   |
|-------------|-------|------|------|------|------|-------|------|------|------|
| Grau        | 0.904 | 0.95 | 0.95 | 0.95 | 0.85 | 0.938 | 1    | 0.85 | 0.75 |
| Junção      | 0.84  | 0.8  | 0.8  | 0.85 | 0.75 | 0.938 | 0.8  | 0.9  | 0.9  |
| Força       | 0.929 | 0.85 | 1    | 0.95 | 0.8  | 0.938 | 0.95 | 0.95 | 1    |
| Completo    | 0.955 | 1    | 1    | 0.9  | 0.85 | 1     | 1    | 0.95 | 0.95 |
| Forma       | 0.955 | 0.95 | 0.85 | 0.95 | 0.95 | 1     | 1    | 0.95 | 1    |
| Zernike     | 0.865 | 0.75 | 0.55 | 1    | 1    | 0.813 | 1    | 0.9  | 0.9  |

**Tabela 5.4:** Base OBJETOS. Legenda: PT (Precisão Total); Ci (Classe i,  $1 \le i \le 8$ ).

Analisando os resultados da Tabela 5.4, pode-ser observar que o descritor Completo, baseado em redes complexas, liderou o melhor desempenho na tarefa de classificação das imagens juntamente com o descritor Forma. Entretanto, o descritor completo conseguiu 100% de acerto na categorização de 4 classes enquanto que o classificador Forma conseguiu este mesmo resultado apenas para 3 classes. Além disso, o classificador Completo foi o que obteve o maior número de classes melhor categorizadas (6 classes). O segundo melhor resultado foi o descritor Força, classificando com 92.9% de acerto as imagens. O que obteve o resultado menos satisfatório foi o descritor Junção, com 84% de acerto, seguido do descritor Zernike, com 86.5%. Considerando o número de características gerados, o descritor Completo possui 7 atributos, enquanto que o descritor Forma possui 11 atributos. Ambos os descritores possuem um pequeno número de atributos.

Pelas características da base, pode-se inferir que o descritor Completo e o descritor Forma caracterizaram bem as imagens, mostrando-se invariáveis à rotação, escala e pequenas deformações não-rígidas.

Aqui não será apresentada a árvore de decisão gerada na fase de treinamento, mas os atributos que foram selecionados pelo algoritmo foram força média, força máxima, energia, grau máximo e grau médio. Observe que o grau médio e a força média continuam fazendo parte dos atributos selecionados pela árvore de decisão.

#### 5.2.3 Estudo de caso 3: Imagens diversas

Nesta seção são apresentadas as redes complexas e os gráficos na escala log log do número de arestas vs. o número de nós para outros tipos de imagens. Esta análise é para verificar se as redes geradas realmente possuem a característica de serem redes livres de escala. A imagem original da Figura 5.12 foi retirada do site da BigStock<sup>4</sup> e a imagem da Figura 5.13 foi retirada do blog Virgem em Câncer<sup>5</sup>.



Figura 5.12: (a) Modelagem da rede complexa e (b) gráfico na escala log. A imagem binarizada possui uma resolução de 450 × 442 pixels e a rede gerada possui 17529 nós.

<sup>&</sup>lt;sup>4</sup>www.bigstock.com, código da imagem 7663814

 $<sup>^{5} \</sup>rm http://virgemcancer.wordpress.com/2010/06/30/mamografia-3-d-traz-nova-dimensao-para-o-exame-de-cancer-de-mama/$ 



Figura 5.13: (a) Modelagem da rede complexa e (b) gráfico na escala log. A imagem original possui uma resolução de  $320 \times 320$  pixels e a rede gerada possui 5184 nós.

As redes apresentadas nas Figuras 5.12 e 5.13 possuem mais de 5 mil nós. Pode-se ver mais claramente no gráfico que estas redes realmente tendem a ser uma reta, apresentado a característica que muitos nós possuem poucas arestas e poucos nós possuem muitas arestas. Com isso, é mostrado que as redes geradas são livres de escala.

Outro fator importante, é que as imagens que geram redes com a distribuição log log de número de arestas vs. o número de nós são as que apresentam seus esqueletos mais complexos, o que possibilita a geração de mais pontos (nós) na rede. Isto ocorre, por exemplo, quando o contorno da imagem original possui ruídos, ou seja, não é uniforme, como é mostrado na Figura 5.13.

A seguir, o modelo proposto é aplicado em uma base real de imagens de mamografia.

# 5.2.4 Estudo de caso 4: desenvolvimento de um sistema de auxílio ao diagnóstico de câncer de mama

O câncer de mama lidera como a segunda causa de morte por câncer em mulheres, sendo o primeiro o câncer de pulmão (Society, 2012). Para reduzir a taxa de mortalidade causada pelo câncer de mama, a detecção precoce da doença é muito importante. A mamografia é atualmente a modalidade de imagem mais eficaz para o rastreamento do câncer de mama. Entretanto, entre 10% e 30% dos casos de câncer de mama não são percebidos nos exames de mamografia (Mohanty et al., 2011) e, em alguns casos, os radiologistas novatos não conseguem detectar os tumores devido a falta de experiência.

Para a classificação das massas, o especialista radiologista inicialmente classifica as imagens baseando-se na forma da lesão (Wei & Li, 2008). Se há infiltrações nos tecidos adjacentes, resultando em um contorno irregular, isso significa que a massa representa um tumor maligno. Por outro lado, se a massa apresenta um contorno regular com uma borda bem delimitada, a massa tende a ser um tumor benigno (Felipe et al., 2006).

Dessa maneira, na aplicação da mamografia para o diagnóstico de câncer de mama, tecnologias de detecção auxiliada por computador desempenham um papel muito importante e muitos sistemas CAD têm sido desenvolvidos. Com a adoção de tecnologias de refinamento de imagens e reconhecimento de padrões, sistemas CAD podem auxiliar radiologistas a encontrar mais facilmente e efetivamente regiões anormais na mama (Tang & Liu, 2011).

Em sistemas de detecção de câncer de mama auxiliados por computador, tecnologias de classificação de massa é uma das principais tecnologias usadas em sistemas CAD. Primeiramente é realizada a segmentação da massa e depois é aplicada a tarefa de classificação para classificar as massas em benignas ou malignas.

#### Base de imagens

Esta base de imagens consiste de 250 regiões de interesse coletadas da base de dados DDSM (*Digital Database for Screening Mammography*)<sup>6</sup>. Esta base é composta por 99 massas benignas e 151 massas malignas.

A Figura 5.14 apresenta exemplos de imagens da classe de massa benigna e maligna.



Figura 5.14: Exemplo de uma mamografia digital de tumor (a) benigno e (b) maligno.

#### Segmentação de imagens mamográficas e detecção do tumor

Na fase de preprocessamento, as imagens foram segmentadas usando uma versão melhorada do algoritmo EM/MPM, proposto por Balan (2007), como descrito na Seção 2.1. Como neste método a segmentação das imagens é realizada diante de um número fixo de diferentes regiões de textura, para as imagens de mama foram utilizadas neste trabalho 5 regiões. Esta escolha foi baseada em diversos testes experimentais empíricos, observando sempre o resultado final visual da segmentação.

Após a segmentação da massa em 5 regiões, foi feita a escolha da região principal. A escolha foi baseada nas características visuais das imagens da base de dados em estudo, na

<sup>&</sup>lt;sup>6</sup>http://marathon.csee.usf.edu/Mammography/Database.html

qual todas são regiões de interesse (ROIs) centradas. Dessa maneira, a escolha da região foi baseada no centroide da imagem. A Figura 5.15 apresenta o resultado da aplicação do método de segmentação em uma imagem de massa de uma mamografia digital.



Figura 5.15: (a) Imagem original; (b) Imagem segmentada em 5 regiões; (c) Máscara da região principal.

#### Avaliando a rede complexa

As Figuras 5.16 e 5.17 mostram a rede complexa e o gráfico na escala log log do número de arestas vs. o número de nós para uma massa benigna e uma maligna, respectivamente.



Figura 5.16: (a) Modelagem da rede complexa e (b) gráfico na escala log. A imagem original possui uma resolução de 72 × 72 pixels e a rede gerada possui 217 nós.

Analisando as redes e os gráficos gerados, é possível observar a presença de *hubs*. E interessante notar que a imagem original da Figura 5.16 possui baixa resolução e sua rede gerada possui apenas 217 nós. E mesmo assim, o gráfico mostra que há um maior número de nós com poucas ligações (arestas) e poucos nós com muitas ligações. Ambos os gráficos podem ser aproximados por uma reta, o que mostra que as redes complexas geradas são livres de escala.



Figura 5.17: (a) Modelagem da rede complexa e (b) gráfico na escala log. A imagem original possui uma resolução de  $255 \times 255$  pixels e a rede gerada possui 1897 nós.

#### Análise dos descritores

Para avaliar os descritores propostos usando medidas topológicas da rede complexa modelada, foi realizada a tarefa de classificação em massas benignas e malignas, cuja saída serve como uma segunda opinião ao especialista médico. Foram calculadas as medidas de acurácia, sensibilidade e especificidade, descritas na Seção 4.4, e foi considerado como sendo o caso positivo a classe dos tumores malignos (tem câncer de mama), e da classe negativa os tumores benignos (não tem câncer de mama).

As medidas calculadas foram comparadas com as medidas geradas pelos descritores de Zernike, Forma e Haralick, cujos resultados estão apresentados na Tabela 5.5

| Descritores | Acurácia | Sensibilidade | Especificidade |
|-------------|----------|---------------|----------------|
| Grau        | 0.74     | 0.821         | 0.616          |
| Junção      | 0.708    | 0.642         | 0.808          |
| Força       | 0.66     | 0.642         | 0.687          |
| Completo    | 0.748    | 0.841         | 0.606          |
| Forma       | 0.7      | 0.808         | 0.535          |
| Zernike     | 0.58     | 0.623         | 0.515          |
| Haralick    | 0.69     | 0.74          | 0.62           |

Tabela 5.5: Base mama 250 imagens.

Analisando os resultados da Tabela 5.5, pode-se observar que o descritor Completo mais uma vez obteve os melhores resultados. Possui o maior valor de acurácia e de sensibilidade, o que mostra que o método é um pouco mais sensível do que específico, o que é desejável no domínio médico. O que apresentou a menor acurácia foi o descritor de Zernike, com apenas 58% de acurácia.

Como o descritor Completo foi o que apresentou o melhor resultado, a seguir está apresentada a árvore de decisão gerada na fase de treinamento. Pode-se observar na árvore de decisão gerada que a característica selecionada foi o grau médio. Ou seja, apenas o grau médio é suficiente para categorizar bem as imagens desta base, em relação aos outros extratores:

grau médio 
$$\leq 0.220323:2$$
  
grau médio  $> 0.220323:1$  (5.4)

# 5.3 Considerações finais

Neste capítulo foi apresentado um novo extrator de características baseado em redes complexas. O método proposto gera redes complexas livre de escala para cada imagem e são extraídas das redes medidas topológicas para comporem o vetor de características da imagem. Pelos resultados obtidos, pode-se observar pelos gráficos na escala log log do número de arestas vs. o número de nós a presença de *hubs* nas redes geradas, e também que um grande número de nós possuíam poucas arestas enquanto que poucos nós possuíam muitas arestas. Com isso, ficou demostrado que as redes geradas pelo método proposto são, de fato, redes livres de escala e, quanto maior o número de nós na rede, mais o gráfico tende a ser uma reta.

Também foi analisada a capacidade dos descritores propostos em categorizar imagens por meio da classificação. Foi observado que as medidas topológicas grau médio e força média da rede foram as que obtiveram maior sucesso na tarefa de classificação. As bases de imagens escolhidas envolveram rotação, escala e pequenas deformações das imagens, o que torna possível inferir que o método proposto é invariante à rotação, escala e pequenas deformações não-rígidas.

Para compreender o aprendizado da base, foi usado o classificador C4.5, o qual constrói uma árvore de decisão na fase de treinamento.

No próximo capítulo é apresentado um novo classificador associativo, o qual gera regras simples, facilitando a compreensão dos resultados obtidos.

Capítulo 6

# Desenvolvimento de classificadores baseados em regras de associação estatísticas

O uso de classificadores associativos foi escolhido por fornecer um modelo de aprendizado de fácil entendimento, o que é importante no domínio médico. Quando a entrada de dados é no domínio de valores contínuos, a maioria dos classificadores associativos primeiro discretizam os dados para depois realizar a etapa de geração de regras. Entretanto, a fase de discretização de dados pode gerar inconsistências na base e perda de informação significativa.

Outra desvantagem da maioria dos classificadores associativos é que os usuários precisam fornecer vários parâmetros de entrada ao algoritmo de mineração das regras, tarefa essa que é trabalhosa e muitas vezes exige que o usuário tenha um conhecimento prévio da base de imagens em questão.

Neste capítulo são apresentados dois novos classificadores associativos: o SACMiner e o MinSAR. Ambos classificadores evitam a fase de discretização de dados, por meio da utilização de regras de associação estatísticas. O MinSAR ainda estima automaticamente os parâmetros do algoritmo de mineração de regras de associação.

Os resultados apresentados neste capítulo foram publicados na forma de capítulo no livro Enterprise Information Systems - Selected Papers - Lecture Notes in Business Information Processing (Watanabe et al., 2011) e também como artigos científicos nas conferências 12th International Conference on Enterprise Information Systems (ICEIS) (Watanabe et al., 2010) e *IEEE 11th International Conference on Machine Learning and Applications (ICMLA)* (Watanabe et al., 2012c).

# 6.1 O classificador SACMiner

Nesta seção é apresentado o método SACMiner (*Statistical Associative Classifier Miner*), que é um novo método de classificação baseado em regras de associação estatísticas para auxílio ao diagnóstico. Este classificador associativo recebe valores contínuos como dados de entrada e não utiliza a discretização de dados, evitando, assim, inconsistências e perda de informação durante a aprendizagem do modelo, ao contrário da maioria do métodos de classificação associativa. Dois novos algoritmos são usados neste método para a construção do classificador. O primeiro, chamado de StARMiner\* (*Statistical Association Rule Miner\**), minera regras de associação estatísticas sobre valores de características contínuos, selecionando as características que melhor representam as imagens. O segundo algoritmo é o V-Classifier (*Voting Classifier*), o qual usa as regras mineradas pelo StARMiner\* para classificar as imagens.

### 6.1.1 Descrição do SACMiner

O método proposto está apresentado na Figura 6.1 e no Algoritmo 5. O método está dividido em duas partes. A primeira, a fase de treinamento, inclui a representação das imagens do conjunto de treinamento e as regras de associação estatísticas são mineradas pelo algoritmo StARMiner<sup>\*</sup>. Na segunda fase, a fase de teste, são extraídas as características das imagens do conjunto de teste e os seus respectivos vetores de características são submetidos ao classificador V-Classifier.

| Algoritmo 5: Passos do método proposto.  |
|--|
| Dados: Imagens do conjunto de treinamento, uma imagem de teste                   |
| <b>Resultado</b> : Sugestão de diagnóstico (Classe da imagem de teste)           |
| 1 Extrair as características das imagens do conjunto de treinamento;             |
| 2 Executar o algoritmo StARMiner <sup>*</sup> para minerar regras de associação; |
| <b>3</b> Extrair as características da imagem teste;                             |
| 4 Executar o classificador V-Classifier;   |
| 5 retorna Laudo sugerido (classe)  |

A seguir estão detalhados os passos do SACMiner.

#### Algoritmo StARMiner\*

O StARMiner<sup>\*</sup> é um modelo de classificação supervisionado cujo objetivo é encontrar regras de associação estatística sobre os vetores de características extraídos das imagens,



Figura 6.1: Pipeline do método SACMiner.

selecionando os atributos que melhor discriminem as imagens em categorias (classes). Ele retorna regras relacionando os intervalos das características selecionadas e as classes da respectiva imagem.

Formalizando, seja  $x_j$  uma classe de uma imagem e  $f_i$  uma característica (atributo). Seja  $V_{min}$  e  $V_{max}$  os valores dos limites inferior e superior, respectivamente, de um intervalo, onde  $V_{min} \leq V_{max}$ . Uma regra minerada pelo algoritmo StARMiner\* tem a forma:

$$f_i[V_{min}, V_{max}] \to x_j. \tag{6.1}$$

Um exemplo de uma regra minerada pelo StARMiner\* é

$$10[-0.47, 0.89] \to \text{massa benigna.} \tag{6.2}$$

Esta regra indica que imagens que possuem o valor da  $10^{a}$  característica no intervalo fechado [0.47, 0.89] tendem a ser imagens da com tumor benigno.

Para executar a mineração de regras de associação, a base de dados de treinamento é percorrida apenas uma vez. É calculada a média e o desvio padrão para cada atributo e o Z-valor (Equação 6.4), usado no teste de hipóteses.

Para que a característica seja selecionada, duas restrições de interesse devem ser satisfeitas:

- **Restrição 1:** A característica  $f_i$  deve apresentar um comportamento uniforme em todas as imagens da classe  $x_i$ .
- **Restrição 2:** A característica  $f_i$  deve ter um comportamento nas imagens da classe  $x_j$  diferente do comportamento das imagens das outras classes.

Para formalizar estas restrições, considere as seguintes notações:

- T: conjunto de vetores de características das imagens;
- $x_j$ : uma classe de imagem;
- $T_{x_j} \in T$ : o subconjunto dos vetores de características de imagens da classe  $x_j$ ;
- $f_i$ : a i-ésima característica do vetor de características;
- $\mu_{f_i}(T_{x_j}) \in \sigma_{f_i}(T_{x_j})$ : respectivamente, média e desvio padrão dos valores da característica  $f_i$  das imagens da classe  $x_j$ ;
- $\mu_{f_i}(T T_{x_j}) \in \sigma_{f_i}(T T_{x_j})$ : respectivamente, média e desvio padrão dos valores da característica  $f_i$  das imagens que não estão na classe  $x_j$ .

Uma regra  $f_i[V_{min}, V_{max}] \rightarrow x_j$  é gerada pelo algoritmo somente se a regra satisfaz os limiares, definidos pelo usuário:

- $\Delta \mu_{min}$ : é a o valor de diferença mínimo permitido entre a média dos valores da característica  $f_i$  das imagens da classe  $x_j$  e o restante das imagens da base de dados;
- $\sigma_{max}$ : é o desvio padrão máximo permitido da característica  $f_i$  da classe  $x_j$ ;
- $\gamma_{min}$ : é a confiança mínima para rejeitar o teste de hipósteses  $H_0$ .

O teste de hipóteses  $H_0$  afirma que a média dos valores de  $f_i$  dentro e fora da classe  $x_j$  são estatisticamente iguais:

$$H_0: \mu_{f_i}(T_{x_j}) = \mu_{f_i}(T - T_{x_j}), \tag{6.3}$$

e o Z-valor é calculado como:

$$Z_{i,j} = \frac{\mu_{f_i}(T_{x_j}) - \mu_{f_i}(T - T_{x_j})}{\frac{\sigma_{f_i}(T_{x_j})}{\sqrt{|T_x|}}}$$
(6.4)

Os valores de  $V_{min}$  e  $V_{max}$  são calculados como:

$$V_{min} = \mu_{f_i} - \sigma_{max} \tag{6.5}$$

$$V_{max} = \mu_{f_i} + \sigma_{max} \tag{6.6}$$

O pseudo-código do algoritmo está apresentado no Algoritmo 6.

Algoritmo 6: Algoritmo StARMiner\*. Dados: Base de dados T: tabela de vetores de características  $\{x_j, f_1, f_2, ..., f_n\}$ , onde  $x_j$  é a classe da imagem e  $f_i$  são as características das imagens;  $\Delta \mu_{min}, \sigma_{max} \in \gamma_{min}$ .

| <b>Resultado:</b> Regras mineradas   |
|--|
| 1 Percorra a base de dados $T$ ;   |
| 2 para cada classe $x_j$ faça  |
| $3$ para cada característica $f_i$ faça  |
| 4 Calcule $\mu_{f_i}(T_{x_i}) \in \mu_{f_i}(T - T_{x_i});$   |
| 5 Calcule $\sigma_{f_i}(T_{x_i}) \in \sigma_{f_i}(T - T_{x_i});$   |
| 6 Calcule Z-valor $Z_{ij}$ ; (Eq. 6.4)   |
| 7 fim  |
| s para cada característica $f_i$ faça  |
| 9   se $(\mu_{f_i}(T_{x_j}) - \mu_{f_i}(T - T_{x_j})) \ge \Delta \mu_{min} AND \sigma_{f_i}(T_{x_j}) \le \sigma_{max} AND$ |
| $(Z_{i_i} < Z_1 \ \boldsymbol{OR} \ Z_{i_i} > Z_2) \ \mathbf{então}$   |
| 10 Escreva $f_i [\mu_{f_i} - \sigma_{max}, \mu_{f_i} + \sigma_{max}] \rightarrow x_j;$                                     |
| 11 fim   |
| 12 fim   |
| 13 se Nenhuma regra for encontrada então   |
| 14 Escolha a característica $f_i$ com o maior Z-valor correspondente;  |
| 15 Escreva $f_i \left[ \mu_{f_i} - \sigma_{max}, \mu_{f_i} + \sigma_{max} \right] \rightarrow x_j;$                        |
| 16 fim   |
| 17 fim   |
|  |

O algoritmo StARMiner<sup>\*</sup> tem a propriedade interessante de que o número máximo de regras mineradas pela classe  $x_j$  é o número de características (k). A complexidade deste algoritmo é  $\Theta(ckN)$ , onde N é o número de instâncias da base de dados, k é o número de características e c é o número de classes. O StARMiner<sup>\*</sup> é baseado na ideia do algoritmo de seleção de características StARMiner, proposto por Ribeiro et al. (2009a) e descrito na Seção 4.2.1. As principais diferenças entre o StARMiner e o StARMiner<sup>\*</sup> é que o segundo algoritmo tem a vantagem de minerar regras semanticamente relevantes, que podem ser usadas em classificadores associativos. Outra diferença é que enquanto que o algoritmo StARMiner apenas seleciona as características, relacionando classes a características que as melhor representem, o StARMiner<sup>\*</sup> encontra regras relacionando classes e intervalos de características, onde um comportamento particular ocorreu.

#### **Classificador V-classifier**

Após geradas as regras de mineração, ou seja, o modelo para realizar a classificação, foi desenvolvido um algoritmo baseado na ideia de 'contar votos'. Para cada classe, é contado

o número de regras que são satisfeitas. O resultado final da contagem é normalizado pelo número de regras de cada classe. A saída do algoritmo é então a classe que obteve a maior porcentagem de votos. A este algoritmo chamamos de *Voting-Classifier* ou, simplesmente, *V-Classifier*, cujo pseudo-código está descrito no Algoritmo 7.

| Algoritmo 7: V-Classifier  |
|--|
| <b>Dados</b> : Regras mineradas na forma $f_i[\mu_{f_i} - \sigma_{max}, \mu_{f_i} + \sigma_{max}] \rightarrow x_j$ , e um vetor de |
| características $g$ de uma nova imagem, onde $g_i$ são as características  |
| <b>Resultado</b> : Sugestão de diagnóstico (classe da nova imagem)   |
| 1 para cada classe $x_i$ faça  |
| $2     voto(x_j) = 0;$   |
| 3 para cada característica $f_i$ faça  |
| 4   se $g_i \in [\mu_{f_i} - \sigma_{max}, \mu_{f_i} + \sigma_{max}]$ então  |
| 5 $voto(x_j) = voto_{x_j} + 1;$  |
| 6 fim  |
| 7 fim  |
| <b>s</b> Divida $voto(x_j)$ pelo número de regras da classe $x_j$ ;  |
| 9 fim  |
| 10 retorna a classe $\max(voto(x_j))$  |
|  |

Analisando o algoritmo, pode-se observar que o algoritmo V-Classifier tem ordem de complexidade linear com relação ao número de regras. Dessa forma, o custo computacional do SACMiner é baixo, considerando também que o algoritmo StARMiner\* é linear com relação ao número de imagens. Com relação ao número de regras, deve ser ressaltado que o StARMiner\* tem a propriedade de que o número máximo de regras mineradas para a classe  $x_j$  é o número de características do vetor de características das imagens vezes o número de classes, e o número mínimo de regras é o número de classes (uma regra por classe). Entretanto, como o algoritmo StARMiner\* faz a seleção de atributos, é possível ajustar os parâmetros de entrada para que o número de características selecionadas seja reduzido e, consequentemente, o número de regras geradas.

#### 6.1.2 Experimentos

Nesta seção são apresentados três estudos de caso realizados para validar o método SACMiner na tarefa de sugerir diagnóstico para imagens médicas. Os experimentos foram realizados empregando as abordagens *holdout*, empregando 25% das imagens para a fase de teste e o restante das imagens para treinamento, e *leave-one-out*.

Para mostrar a eficiência do método, comparamos seus resultados com os resultados gerados pelos classificadores 1-NN, C4.5, naive bayes e 1R. As medidas usadas na comparação foram acurácia (Eq. 4.9), sensibilidade (Eq. 4.10) e especificidade (Eq. 4.12). Para calcular estes valores, considere:

- Verdadeiro positivo: massas malignas corretamente classificadas como malignas;
- Falso positivo: massas benignas classificadas incorretamente como malignas;
- Verdadeiro negativo: massas benignas corretamente classificadas como benignas;
- Falso positivo: massas malignas classificadas incorretamente como benignas.

#### Experimento 1: Base de 250 imagens

Esta base de imagens consiste de 250 regiões de interesse coletadas da base de dados DDSM (*Digital Database for Screening Mammography*)<sup>1</sup>. Esta base é composta por 99 massas benignas e 151 massas malignas.

O extrator de características usado é o mesmo descrito na Seção 5.2, com 11 características. Após a binarização da imagem e a detecção da região de interesse, foram extraídos da ROI de cada imagem 11 valores para comporem o vetor de características: área, tamanho do eixo maior, tamanho do eixo menor, excentricidade, orientação, área convexa, área preenchida, número de Euler, solidez, amplitude e perímetro, cujas definições estão apresentadas na Seção 2.2.3.

Os vetores de características do conjunto das imagens de treinamento foram submetidos ao algoritmo StARMiner<sup>\*</sup> para minerar regras de associação estatística. Este algoritmo gerou as seguintes regras:

$$1[-0.0120, 0.1770] \to Benigna$$
 (6.7)

$$6[-0.0075, 0.1825] \to Benigna$$
 (6.8)

$$7[-0.0133, 0.1767] \to Benigna$$
 (6.9)

$$2[0.2973, 0.4873] \to Maligna$$
 (6.10)

Nestas regras, as características 1,2,6 e 7 são, respectivamente, a área, tamanho do eixo maior, área convexa e área preenchida. A primeira regra, por exemplo, significa que massas cuja área está no intervalo [-0.0120, 0.1770] tendem a ser benignas. Para este experimento, consideramos um nível de confiança de 90% para teste Z e para calcular os intervalos das regras.

Após geradas as quatro regras, os vetores de características das imagens do conjunto de teste foram introduzidas no classificador construído na fase de treinamento. Os resultados usando a abordagem *holdout* e *leave-one-out* estão apresentados nas Tabelas 6.1 e 6.2.

Analisando a Tabela 6.1, pode-se observar que o SACMiner apresenta os maiores valores de acurácia e especificidade na abordagem *holdout*. Quando analisamos os valores de sensibilidade, podemos notar que o classificador naive bayes obteve os melhores

<sup>&</sup>lt;sup>1</sup>http://marathon.csee.usf.edu/Mammography/Database.html

| Classificadores | Acurácia | Sensibilidade | Especificidade |
|-----------------|----------|---------------|----------------|
| SACMiner        | 0.8548   | 0.8461        | 0.8611         |
| 1R              | 0.7258   | 0.8260        | 0.6666         |
| naive bayes     | 0.6290   | 0.9130        | 0.4615         |
| C4.5            | 0.7585   | 0.7391        | 0.7692         |
| 1-NN            | 0.6129   | 0.6521        | 0.5897         |

 

 Tabela 6.1: Comparação entre SACMiner e outros classificadores usando a abordagem holdout.

 Tabela 6.2: Comparação entre SACMiner e outros classificadores usando a abordagem leave-one-out.

| Classificadores | Acurácia | Sensibilidade | Especificidade |
|-----------------|----------|---------------|----------------|
| SACMiner        | 0.7680   | 0.7788        | 0.7603         |
| 1R              | 0.7680   | 0.7885        | 0.7534         |
| naive bayes     | 0.7360   | 0.8750        | 0.6370         |
| C4.5            | 0.7440   | 0.6154        | 0.8356         |
| 1-NN            | 0.6760   | 0.6154        | 0.7192         |

resultados. Entretanto, quando analisamos seu valor de especificidade, o naive bayes não classifica bem imagens da classe benigna.

Na Tabela 6.2, o classificador SACMiner alcança os maiores valores de acurácia juntamente com o classificador 1R. Neste caso, podemos inferir que a abordagem de regras de associação foi a melhor para classificar as massas. Mas é importante ressaltar que uma vantagem do SACMiner sobre o 1R é que o SACMiner não demanda do passo de discretização dos dados, enquanto que o 1R necessita. Além disso, SACMiner gerou apenas quatro regras, enquanto que o 1R produziu oito regras. Todas as regras geradas pelo 1R foram baseadas na característica tamanho do eixo maior, o segundo atributo do vetor de características, as quais estão descritas a seguir:

| SE $L < 0.1840$ ENTAO Benigna (6.1) | 1 | L | ) | ) |  |
|-------------------------------------|---|---|---|---|--|
|-------------------------------------|---|---|---|---|--|

- SENÃO SE L < 0.2181 ENTÃO Maligna(6.12)
- SENÃO SE L < 0.2367 ENTÃO Benigna (6.13)
- SENÃO SE L < 0.2572 ENTÃO Maligna (6.14)
- SENÃO SE L < 0.2716 ENTÃO Benigna (6.15)
- SENÃO SE L < 0.3126 ENTÃO Maligna (6.16)
- SENÃO SE L < 0.3424 ENTÃO Benigna(6.17)
- SENÃO SE  $L \ge 0.3424$  ENTÃO Maligna. (6.18)

#### Experimento 2: A base 569 ROIs

Esta base consiste de 569 vetores de características obtidos do repositório de aprendizado de máquina UCI<sup>2</sup> (Asuncion & Newman, 2007). Os atributos deste vetores descrevem as características de núcleo da célula presente na imagem. As características foram calculadas a partir de massas de mama e estão classificadas em massas benignas e malignas. Para cada uma dos três núcleos celulares, as seguintes dez características foram extraídas: médias das distâncias entre o centro e os pontos sobre o perímetro, desvio padrão dos valores em nível de cinza, perímetro, área, suavidade, compacidade, concavidade, pontos côncavos, simetria e dimensão fractal. Portanto, o vetor de característica tem 30 atributos, 10 para cada núcleo. As classes estão distribuídas em 357 massas benignas e 212 malignas.

O algoritmo StARMiner<sup>\*</sup> minerou 19 regras para cada classe. Os resultados usando as abordagens *holdout* e *leave-one-out* estão apresentados nas Tabelas 6.3 e 6.4, respectivamente.

| Tabela 6.3: | Comparação | entre | SACMiner | e outros | classificadores | usando a | a abordagem |
|-------------|------------|-------|----------|----------|-----------------|----------|-------------|
|             | holdout.   |       |          |          |                 |          |             |

| Classificadores | Acurácia | Sensibilidade | Especificidade |
|-----------------|----------|---------------|----------------|
| SACMiner        | 0.9859   | 0.9888        | 0.9811         |
| 1R              | 0.8943   | 0.9186        | 0.8571         |
| naive bayes     | 0.9155   | 0.9186        | 0.9107         |
| C4.5            | 0.9295   | 0.9419        | 0.9107         |
| 1-NN            | 0.9577   | 0.9767        | 0.9286         |

 Tabela 6.4:
 Comparação entre SACMiner e outros classificadores usando a abordagem leave-one-out.

| Classificadores | Acurácia | Sensibilidade | Especificidade |
|-----------------|----------|---------------|----------------|
| SACMiner        | 0.9525   | 0.9860        | 0.8962         |
| 1R              | 0.9015   | 0.9356        | 0.8443         |
| naive bayes     | 0.9349   | 0.9580        | 0.8962         |
| C4.5            | 0.9384   | 0.9524        | 0.9151         |
| 1-NN            | 0.9525   | 0.9580        | 0.9434         |

Analisando os resultados da Tabela 6.3, podemos observar que SACMiner obteve os melhores resultados de acurácia, sensibilidade e especificidade. Além disso, quando comparamos os resultados usando a abordagem *leave-one-out* (Tabela 6.4), também observamos que o método alcançou o maior valor de acurácia, apresentando o mesmo resultado que o 1-NN, e o maior valor de sensibilidade.

<sup>&</sup>lt;sup>2</sup>http://archive.ics.uci.edu/ml/datasets.html

#### Experimento 3: Base de dados de tecido de mama

Um outro tipo de exame que pode ser usado para distinguir tecido mamário e especialmente para detectar câncer de mama é a espectroscopia de impedância elétrica (*electrical impedance spectroscopy - EIS*) (Silva et al., 2000). Obtivemos uma base de vetores de características baseada nesta abordagem do UCI Machine Learning Repository. Esta base é composta por 106 vetores de características. Cada vetor é composto por 9 características e é classificado em uma das seis classes de tecidos que foram considerados usando medidas de impedância elétrica:

- Carcinoma: 21 casos;
- Fibro-adenoma: 15 casos;
- Mastopatia: 18 casos;
- Glandular: 16 casos;
- Conjuntivo: 14 casos;
- Adiposo: 22 casos.

As três primeiras são classes de tecidos patológicos (doentes) e as últimas três são classes de tecidos normais. Maiores detalhes do procedimento de obtenção dos dados assim como a classificação dos casos e das frequências usadas podem ser encontrados em Silva et al. (2000).

Para esta base de dados, a classificação consistiu em discriminar cada uma das classes. Para avaliar o poder de discriminação do SACMiner entre as classes, foi usada a abordagem *holdout*, na qual 75% dos dados foram randomicamente selecionados para comporem a base de imagens do conjunto de treinamento e o restante foi usado para o conjunto teste. Para cada classificador foi calculada a média da acurácia das classes e a acurácia de cada classe separadamente.

Na fase de treinamento, o algoritmo StARMiner<sup>\*</sup> minerou 34 regras de associação estatística, como se segue:

Carcinoma: 4 regras Fibro-adenoma: 8 regras Mastopatia: 7 regras Glandular: 7 regras Conjuntivo: 5 regras Adiposo: 3 regras

De acordo com a Tabela 6.5, SACMiner obteve o melhor valor de acurácia média, alcançando 88.5% de precisão. Além do mais, o classificador SACMiner também apresentou os melhores resultados de acurácia para identificar cada classe individualmente,

| Acurácia      | SACMiner | 1R    | naive | C4.5  | 1-NN  |
|---------------|----------|-------|-------|-------|-------|
|               |          |       | bayes |       |       |
| Total         | 0.885    | 0.577 | 0.731 | 0.769 | 0.731 |
| Carcinoma     | 1        | 1     | 1     | 1     | 0.833 |
| Fibro-adenoma | 0.75     | 0     | 0.333 | 0     | 0     |
| Mastopatia    | 0.75     | 0.25  | 0.5   | 0.25  | 0.25  |
| Glandular     | 1        | 0     | 0.5   | 1     | 1     |
| Conjuntivo    | 1        | 0.75  | 1     | 1     | 1     |
| Adiposo       | 0.8      | 1     | 0.8   | 1     | 1     |

**Tabela 6.5:** Comparação dos valores de acurácia entre SACM<br/>iner e outro classificadores usando a abordagem holdout

o que significa que SACMiner é melhor que os outros classificadores ou que tem o mesmo desempenho em distinguir tecidos Carcinoma, Fibro-adenoma, Matopatia, Glandular e Conjuntivo. O ganho de precisão alcançado é de até 11.6% maior quando comparado com o classificador C4.5, o qual obteve o segundo melhor resultados (76.9% de acurácia).

### 6.1.3 Considerações sobre o SACMiner

Até aqui foi apresentado um novo classificador associativo, o SACMiner, o qual emprega regras de associação estatística para auxiliar o diagnóstico por computador de câncer de mama. Os resultados obtidos usando bases reais mostraram que o método desenvolvido alcançou os maiores valores de acurácia, quando comparado com os resultados obtidos usando classificadores descritos na literatura (1-R, naive bayes, C4.5 e 1-NN). Além disso, o método apresentou um equilíbrio adequado entre sensibilidade e especificidade, sendo um pouco mais sensível do que específico, o que é aceitável no domínio médico, já que é mais preciso para identificar os verdadeiros positivos (casos patológicos). O modelo de aprendizado gerado pelo classificador, por meio do algoritmo StARMiner\*, é de fácil compreensão, fazendo com que seja melhor compreendido pelo especialista quais as medidas que levaram o classificador a sugerir a classe de saída dada.

Além disso, o uso da tarefa de regra de associação estatística possibilitou, por meio do algoritmo StARMiner<sup>\*</sup>, que o número de regras geradas fosse reduzido e também que o novo classificador não precisasse realizar a fase de discretização dos dados, fase esta que pode gerar mais processamento e também inconsistência na base de dados.

Entretanto, é importante ressaltar que este método ainda necessita que o usuário forneça os limiares de diferença mínima  $\Delta \mu_{min}$  e desvio padrão máximo  $\sigma_{max}$ , e que poucas alterações nestes valores modificam o modelo de aprendizado gerado e, consequentemente, a precisão da classificação fica afetada.

Deste ponto de vista, é interessante então que seja desenvolvido um método que estime automaticamente estes valores, ficando para o usuário, apenas fornecer o valor de confiança desejado. Por isso, na próxima seção é apresentado o classificador MinSAR, o qual é baseado no SACMiner, mas que estima de forma automática estes parâmetros.

# 6.2 O classificador MinSAR

A maioria dos classificadores associativos, incluindo o SACMiner, requerem que o usuário forneça limiares para a produção dos *itemsets*, i.e., limiares são empiricamente determinados e o usuário precisa testar muitos parâmetros até que consiga obter resultados satisfatórios por meio da combinação destes parâmetros. Isso requer que o usuário tenha conhecimento prévio da base de dados e do problema em questão, além de ter que possuir um certo conhecimento do algoritmo que está sendo usado.

Para lidar com os desafios de evitar a fase de discretização dos dados, gerar poucas regras para o modelo de classificação, e evitar que o usuário tenha que fornecer limiares, foi proposto neste trabalho o classificador MinSAR (*Mining Statistical Association Rules*). Este classificador emprega o uso de regras de associação estatística para construir o modelo de classificação. Ele é baseado no algoritmo SACMiner. A principal diferença entre os dois classificadores é que o MinSAR faz a estimação automática dos parâmetros. Dessa forma, o usuário necessita apenas fornecer o parâmetro de confiança.

#### 6.2.1 Descrição do MinSAR

O pipeline do MinSAR (Figura 6.2) é semelhante ao do SACMiner (Figura 6.1). O que diferencia os dois métodos é o algoritmo responsável por gerar o modelo de classificação.

No SACMiner, o algoritmo chama-se StARMiner<sup>\*</sup>. No MinSAR, o algoritmo chama-se algoritmo MinSAR, o qual está apresentado no Algoritmo 8. O algoritmo StARMiner<sup>\*</sup> e MinSAR são semelhantes. O que os diferencia são as linhas 7, 9 e 10 do Algoritmo 8, onde ocorre a estimação dos parâmetros  $\Delta \mu_{min}$  e  $\sigma_{max}$ . Esses limitares são calculados da seguinte maneira. Para cada classe  $x_j$  e para cada características  $f_i$ , tem-se  $m_{f_i}$  tal que

$$m_{f_{i}} = ||\mu_{f_{i}}(T_{x_{j}}) - \mu_{f_{i}}(T - T_{x_{j}})| - |\mu_{f_{i}}(T_{x_{j}}) - \sigma_{f_{i}}(T_{x_{j}})| - |\mu_{f_{i}}(T - T_{x_{j}}) - \sigma_{f_{i}}(T - T_{x_{j}})||; \qquad (6.19)$$

Assim, para cada classe  $x_i$ , os valores de  $\Delta \mu_{min}$  e  $\sigma_{max}$  são estimados como:

$$\Delta \mu_{min} = \min(m_{f_i}) + \frac{\max(m_{f_i}) - \min(m_{f_i})}{8}; \qquad (6.20)$$


Figura 6.2: Pipeline do método MinSAR.



$$\sigma_{max} = \gamma_{min} * \max(\sigma_{f_i}(T_{x_i})). \tag{6.21}$$

Observe que o desvio padrão máximo será ponderado pelo valor de confiança determinado pelo usuário,  $\Delta \mu_{min} > 0 \in \sigma_{max} > 0$ .

#### 6.2.2 Experimentos

Para avaliar a eficiência do classificador MinSAR, foram comparados os resultados de erro, acurácia, sensibilidade e especificidade com os resultados gerados pelos classificadores C4.5, 1-R, naive bayes e Adaboost. A abordagem usada foi a *10-fold cross-validation*. As bases de imagens usadas foram as mesmas descritas na Seção 6.1.2.

#### Experimento 1: Base de 250 imagens

Esta base é composta por 250 imagens de massas de mama, sendo que 151 são massas malignas e 99 são benignas, como descrito na Seção 6.1.2. O vetor de características é composto por 11 características. Os resultados da aplicação dos classificadores usando esta base de imagens estão descritos na Tabela 6.6.

| Tabela 6.6: | Resultados comparando | MinSAR | com ou | tros classific | cadores pa | ara a | base of | le |
|-------------|-----------------------|--------|--------|----------------|------------|-------|---------|----|
|             | 250 imagens.          |        |        |                |            |       |         |    |

| Classificadores | Erro   | Acurácia | Sensibilidade | Especificidade |
|-----------------|--------|----------|---------------|----------------|
| MinSAR          | 0.2080 | 0.7920   | 0.8767        | 0.6731         |
| SACMiner        | 0.2480 | 0.7520   | 0.7671        | 0.7308         |
| C4.5            | 0.2680 | 0.7320   | 0.8150        | 0.6150         |
| 1R              | 0.2760 | 0.7240   | 0.7603        | 0.6731         |
| naive bayes     | 0.2760 | 0.7240   | 0.8650        | 0.6230         |
| Adaboost        | 0.2440 | 0.7560   | 0.8420        | 0.6350         |

Analisando os resultados da Tabela 6.6 pode-se observar que o classificador MinSAR apresentou os maiores valores de acurácia e sensibilidade, e obteve, juntamente com o 1R, o segundo maior valor de especificidade, sendo que o maior valor de especificidade foi do SACMiner. Comparando o MinSAR com os outros classificadores, este obteve uma precisão 4.8% maior que o Adaboost, 8.2% maior que o C4.5 e 9.4% maior que o 1R e o naive bayes.

Além dos valores de acurácia, pode-se analisar também o número de regras produzidas pelos algoritmos e o número de características selecionadas para produzir as regras, para, assim, analisar o custo computacional do algoritmo. O algoritmo MinSAR selecionou apenas 3 características das 11 que compõem o vetor de características, e produziu apenas 5 regras, das quais 3 regras referem-se à classe benigna e 2 regras à classe maligna. As regras geradas estão listadas a seguir:

Tamanho do eixo maior 
$$[-0.0412, 0.3929] \rightarrow$$
 benigna  
Área Convexa  $[-0.1286, 0.3055] \rightarrow$  benigna  
Perímetro  $[-0.0631, 0.3710] \rightarrow$  benigna  
(6.22)  
Tamanho do eixo maior  $[0.1827, 0.6069] \rightarrow$  maligna  
Perímetro  $[0.1545, 0.5787] \rightarrow$  maligna

Já o algoritmo 1R minerou 8 regras, sendo 4 regras para cada classe, e a característica selecionada foi o tamanho do eixo maior. As regras geradas pelo 1R estão descritas a seguir:

$$0 \leq \text{Tamanho eixo maior} < 0.1840 \rightarrow \text{benigna}$$
  
 $0.1840 \leq \text{Tamanho eixo maior} < 0.2181 \rightarrow \text{maligna}$   
 $0.2181 \leq \text{Tamanho eixo maior} < 0.2367 \rightarrow \text{benigna}$   
 $0.2367 \leq \text{Tamanho eixo maior} < 0.2573 \rightarrow \text{maligna}$   
 $0.2573 \leq \text{Tamanho eixo maior} < 0.2716 \rightarrow \text{benigna}$  (6.23)  
 $0.2716 \leq \text{Tamanho eixo maior} < 0.3127 \rightarrow \text{maligna}$   
 $0.3127 \leq \text{Tamanho eixo maior} < 0.3424 \rightarrow \text{benigna}$   
Tamanho eixo maior >  $0.3424 \rightarrow \text{maligna}$ 

Finalmente, as regras produzidas pelo C4.5 são:

Perímetro 
$$\leq 0.1570 \rightarrow$$
 benigna  
Perímetro  $> 0.1570 \rightarrow$  maligna (6.24)

#### Experimento 2: A base de 569 ROIs

A segunda base de imagens é a de 569 imagens coletadas da UCI Machine Learning, sendo que 357 são massas benignas e 212 são massas malignas, como descrito anteriormente na Seção 6.1.2. O vetor de característica é composto por 30 características.

Os resultados de erro de classificação, acurácia, sensibilidade e especificidade dos classificadores estão apresentados na Tabela 6.7.

Analisando os resultados da Tabela 6.7, pode-se observar que o classificador MinSAR obteve os melhores resultados de acurácia e sensibilidade.

Com relação às regras mineradas, o MinSAR, o C4.5 e o 1R produziram, respectivamente, 46, 13 e 4 regras, e selecionaram 24, 10 e 1 característica, respectivamente.

| Classificadores | Erro   | Acurácia | Sensibilidade | Especificidade |
|-----------------|--------|----------|---------------|----------------|
| MinSAR          | 0.0457 | 0.9543   | 0.9720        | 0.9245         |
| SACMiner        | 0.0826 | 0.9174   | 0.9434        | 0.9020         |
| C4.5            | 0.0668 | 0.9332   | 0.9290        | 0.9360         |
| 1R              | 0.1072 | 0.8928   | 0.8490        | 0.9190         |
| naive bayes     | 0.0703 | 0.9297   | 0.8960        | 0.9500         |
| AdaBoost        | 0.0527 | 0.9473   | 0.9150        | 0.9660         |

Tabela 6.7:Resultados comparando MinSAR com outros classificadores para a base de<br/>569 imagens.

Apesar do MinSAR ter produzido o maior número de regras, é importante ressaltar que as regras geradas por ele são simples de serem entendidas, e cada uma tem apenas um intervalo e uma classe relacionada. Já as regras produzidas pelo C4.5 são compostas, sendo necessário a visita às sub-árvores para formar a regra e também checar muitos intervalos até chegar ao nó folha. Se consideramos cada intervalo como uma regra, como no algoritmo MinSAR, o C4.5 tem o equivalente a 25 regras. Já que o MinSAR obteve uma melhor acurácia nos resultados do que o C4.5, pode-se inferir que, apesar do C4.5 ter produzido menos regras, este algoritmo não considerou todas as características relevantes para categorizar as imagens.

#### 6.2.3 Considerações sobre o MinSAR

Considerando, então, os resultados das duas bases de dados, pode-se observar que o classificador MinSAR alcançou melhores resultados de classificação das imagens que os outros classificadores. A partir destes resultados, pode-se inferir que o algoritmo produz uma boa generalização do modelo de classificação. Algumas outras vantagens do MinSAR é que ele é simples de ser implementado e executa a seleção de características sobre o conjunto das características.

Quando comparado com o SACMiner, pode-se observar que a estimação automática dos parâmetros do MinSAR influenciou significativamente na precisão do resultado das classificações e como o SACMiner é sensível aos parâmetros de entrada, assim como a maioria dos métodos descritos na literatura.

A maioria dos classificadores associativos precisam pré-processar os dados contínuos para convertê-los em dados discretos. Assim, outra vantagem importante do classificador MinSAR é que este não demanda o passo de discretização. Outro ponto relevante é que a maioria dos classificadores associativos precisam gerar intervalos candidatos para então fazer a seleção dos intervalos que irão compor o modelo de classificação. Ou seja, ou os algoritmos adotam uma abordagem *top-down*, na qual o algoritmo inicia com um intervalo grande contendo todos os valores e, recursivamente, encontra pontos de corte para intervalos menores até que determinados critérios sejam alcançados. Ou então, os algoritmo adotam a abordagem *bottom-up*, na qual consideram um número de intervalos determinados pelo conjunto de pontos da fronteira e depois, recursivamente, combinam intervalos adjacentes até que certos critérios de parada sejam alcançados. Ao invés de usar essas abordagens, o classificador MinSAR não gera intervalos candidatos; ele apenas produz um intervalo simples se o valor da característica satisfaz as restrições para alguma classe. Além disso, o MinSAR não requer que o usuário forneça limiares na entrada do algoritmo, como os outros métodos de classificação associativa. Os parâmetros de diferença mínima da média entre duas classes e o desvio padrão máximo de uma classe são automaticamente calculados pelo algoritmo.

## 6.3 Considerações finais

Neste capítulo foram apresentados dois classificadores associativos baseados em regras de associação estatística, para auxílio ao diagnóstico de imagens médicas: o SACMiner e o MinSAR. Estes classificadores não realizam a discretização de dados, trabalhando com os dados contínuos diretamente, e não possuem a fase de geração de *itemsets* candidatos, realizando a seleção de características e a geração de regras que compõem o modelo de classificação em um único passo. O classificador MinSAR também estima automaticamente os limiares necessários na fase de mineração das regras, enquanto que a maioria dos métodos da literatura necessitam que o usuário entre com esses valores. Os resultados dos experimentos realizados com bases reais mostram que os classificadores SACMiner e MinSAR alcançaram valores de acurácia, sensibilidade e especificidade altos comparados com algoritmos da literatura como C4.5, naive bayes, 1R.

Capítulo 7

# Detecção automática de epilepsia no lobo temporal

Epilepsia é uma das desordens neurológicas mais comuns, e afeta pessoas de todo o mundo. É a terceira doença neurológica mais comum e impõe altos custos à sociedade (Sujitha et al., 2010). Epilepsia do lobo temporal mesial (ELTM) associada com esclerose hipocampal é a forma de epilepsia mais comum em adultos (Engel Jr., 2008).

O diagnóstico tradicional para a detecção de ELTM é baseado na análise de alterações ocorridas principalmente no hipocampo e em outras estruturas do lobo temporal, sendo que pelo menos 80% dos casos das crises epilépticas iniciam-se no hipocampo (WO 4th, 2012). Por isso, mudanças patológicas no hipocampo e em outras estruturas do lobo temporal estão bem documentadas em pacientes com ELTM (Sujitha et al., 2010). Entretanto, o diagnóstico fica difícil de ser realizado quando o paciente não apresenta mudanças significativas no volume do hipocampo apresentado no exame de ressonância magnética, o que atrapalha o processo de identificação da lateralidade do foco epiléptico e a detecção precoce da doença (Farid et al., 2012). Logo, a classificação automática de imagens de ressonância magnética é uma ferramenta desejável para melhorar a interpretação das imagens, especialmente quando essas mudanças são sutis ou distribuídas igualmente nas estruturas e podem ser despercebidas na inspeção visual (Focke et al., 2012).

Estudos recentes têm detectado anormalidades em regiões neocorticais em ELTM e têm demostrado que a ELTM não está limitada a uma patologia hipocampal, mas antes uma doença que afeta as regiões neocorticais temporais e extratemporais (McDonald et al., 2008; Blümcke et al., 2012). Lin et al. (2007) encontrou mudanças na medida da espessura cortical em pacientes com ELTM. Os autores ainda encontraram 30% de reduções bilaterais em espessura cortical em várias regiões dos lobos frontal, temporal e ocipital em pacientes com ELTM comparados aos controles. Além disso, eles encontraram que a longa duração da doença estava associada à diminuição cortical nas regiões frontal superior, parietal e parahipocampal, sugerindo que a ELTM crônica é caracterizada por mudanças neocorticais que são progressivas e em regiões específicas.

Diversos métodos têm sido desenvolvidos para realizar a detecção de diferenças em estruturas cerebrais de pacientes com ELTM em comparação a controles. Alguns autores desenvolveram *frameworks* que usam extração de características e um classificador do campo de aprendizado de máquina. McDonald et al. (2008) usou medidas de espessura cortical e análise de discriminante linear (LDA) para diferenciar pacientes com ELTM de controles e obteve valores de 74% de acurácia, sensibilidade e especificidade. Keihaninejad et al. (2010) usou seleção de atributos e classificação baseada em volumes estruturais com SVM (*Suport-Vector Machine*) e obteve 96% de precisão na separação entre pacientes com ELTM e controles, usando 14 medidas de volume de estruturas cerebrais. Focke et al. (2012) comparou 38 pacientes com esclerose hipocampal e ELTM unilateral e 22 sujeitos neurologicamente normais, usando características extraídas da segmentação baseada na massa cinzenta e difusidade média para a representação das imagens, obtendo, respectivamente, 88% e 93% de precisão na classificação de pacientes e controles, usando o classificador SVM.

A maioria destes trabalhos se preocupam apenas com a precisão da classificação e não apresentam uma análise das características mais significativas. Neste capítulo é apresentada uma nova metodologia para detecção automática de epilepsia no lobo temporal mesial, que, além da classificação, apresenta a relação das medidas que mais foram significativas no processo da tomada de decisão pelo classificador. Para a categorização das imagens foram usadas medidas de estruturas de espessura cortical para permitir que pacientes com ELTM sejam diagnosticados, mesmo que apresentem mudanças sutis no hipocampo.

Os resultados apresentados neste capítulo foram publicados na forma de artigo científico no XXIII Congresso Brasileiro de Engenharia Biomédica (Watanabe et al., 2012a).

## 7.1 Descrição do método

Nesta seção é apresentada a metodologia proposta para a detecção automática e para o auxílio ao diagnóstico de ELTM desenvolvida neste trabalho. Esta metodologia usa o classificador 1-NN, que possui baixo custo computacional, e o classificador C4.5, o qual gera uma árvore de decisão na fase de aprendizagem. As medidas usadas para a caracterização das imagens são medidas das estruturas corticais, ao invés da utilização de medidas relacionadas ao volume de estruturas subcorticais, dentre elas, o hipocampo, para permitir que, mesmo exames em que o hipocampo possua mudanças sutis ou apresente mudanças distribuídas em ambos os hemisférios, seja realizada a categorização de pacientes e controles com alta precisão de classificação. Além da classificação, nesta metodologia é abordada a análise das estruturas corticais que mais apresentaram uma representação significativa no processo de classificação. Estas medidas foram selecionadas por meio do algoritmo Relief e também pela análise dos atributos selecionados pelo algoritmo C4.5 para a construção do modelo de árvore de decisão. Dessa forma, além da classificação da imagem, a metodologia desenvolvida é capaz de sugerir um diagnóstico ao médico que apresenta a relação das estruturas mais significativas para o processo de classificação, o que não é possível quando são usados classificadores como o LDA e o SVM.

O pipeline do método proposto para a análise das medidas de espessura cortical está apresentado na Figura 7.1.



Figura 7.1: Método proposto

Inicialmente, a fase de treinamento é executada para construir o classificador. Nesta fase, primeiramente são obtidas as imagens de ressonância magnética (RMI) da cabeça dos participantes, que são os pacientes, pessoas que possuem ELTM, e os controles, que são os sujeitos normais, ou seja, sem epilepsia. Depois, na fase de preprocessamento da imagem, as imagens 3D são reconstruídas e cada hemisfério é segmentado em 34 regiões cerebrais usando o *software* FreeSurfer, o qual será detalhado posteriormente. Ainda usando o FreeSurfer, as medidas de espessura cortical são automaticamente extraídas de cada região para compor o vetor de características. As imagens agora são caracterizadas por estes vetores e é aplicado um algoritmo para selecionar as características mais significativas para evitar alterações na fase de classificação. Desta forma, o classificador é construído

baseado nestes vetores de características. Na fase de teste, um novo conjunto de imagens, igualmente processadas, são usadas para avaliar a classificação. A saída nesta fase é a sugestão de um diagnóstico e relações entre algumas regiões cerebrais, o que torna o papel do processo mais compreensível ao especialista. Cada passo do método está descrito a seguir.

#### 7.1.1 Participantes

Os participantes deste trabalho foram 123 pacientes com ELTM e 174 controles saudáveis. A idade deles estava entre 14 e 59 anos. O consentimento para o uso de seus exames nesta pesquisa foi obtido antes dos participantes realizarem o exame de RMI, e este estudo foi aprovado pelo Comitê de Ética do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto, da Universidade de São Paulo (Processo HCRP n<sup>o</sup> 5961/2009).

O critério de inclusão foi realizado da seguinte maneira:

- semiologia de crises consistentes com ELTM, geralmente com auras epigástricas, autônomas, ou físicas seguidas por inibição comportamental, opacificação progressiva da consciência, automatismos oroalimentares e manuais, e fenômenos autonômicos;
- 2. picos interictais unilaterias ou bilaterais, anterior e mesial temporal;
- 3. monitoramento de video-eletroencefalograma com início das crises decorrentes exclusivamente do lobo temporal;
- 4. RMI com atrofia hipocampal evidente (AH) e um sinal hiper-intenso em sequênias T2-ponderadas, sem ordem de lesão identificada;
- 5. ELT medicalmente refratária, definida como falha de resposta de pelo menos dois fármacos anti-epilépticos, após verificações adequadas. Deve-se ressaltar que 56 pacientes tinham AH do lado direito e 66 pacientes tinham AH do lado esquerdo.

#### 7.1.2 Aquisição de imagens e extração de características

#### Aquisição de imagens

Considerando o processo de aquisição das imagens, todos os sujeitos realizaram o exame em um tomógrafo de ressonância magnética nuclear com um campo de 1.5 Tesla (Siemens Magneton Vision, Erhlängen, Germany), usando uma bobina comercial de cabeça com polarização circular. O protocolo incluiu uma sequência Gradiente-eco 3D (MPRAGE) ponderada em T1 (TR = 9,7ms; TE = 4ms; flip angle = 12graus), com *slab* cobrindo todo o encéfalo e proporcionando partições de 1 mm de espessura no plano sagital, matriz de  $256 \times 256$  pixels e campo de visão (*field of view* - FOV) de 256 mm, produzindo assim um pixel isométrico de 1mm<sup>3</sup>.

#### Extração de características

A espessura cortical é uma medida anatômica utilizada para descrever a espessura das camadas do córtex cerebral nos cérebros dos mamíferos. A análise desta medida mede a espessura da matéria cinzenta presente no córtex, e representa um papel importante no grau de capacidades cognitiva que o indivíduo pode ter (Santos, 2010). A espessura cortical também pode ser definida como a distância em milímetros em três dimensões entre a borda da substância branca com a substância cinzenta até a borda da substância cinzenta com o líquido cefalorraquidiano (Picon, 2012).

Considerando o processamento de imagens, medidas da espessura cortical a partir do dado de RMI é uma tarefa não trivial devido à natureza convoluta do córtex. A espessura do córtex é uma propriedade que só pode ser corretamente medida se o local e a orientação das superfícies cinza/branca e pial forem ambas conhecidas, o que torna o processamento dessas imagens um desafio (Fischl & Dale, 2000). Neste estudo, medidas de espessura foram calculadas usando o software FreeSurfer<sup>1</sup> para construção e cálculo da espessura, como descrito em Dale et al. (1999); Fischl et al. (2001). Estimativas da espessura cortical foram calculadas para regiões de interesse (ROIs) baseadas em giros e lobos, e ponto a ponto através do manto cortical. As ROIs baseadas em giros foram derivadas de um sistema de reconhecimento automático que subdivide a RMI do córtex cerebral volumétrico em 34 regiões neuroanatômicas por hemisfério (4 medial temporal, 5 lateral temporal, 11 frontal, 5 parietal, 4 ocipital, e 5 cinguladas) (McDonald et al., 2008). Este método foi validado por (Desikan et al., 2006) e também por (Han et al., 2006).

Após segmentadas as áreas cerebrais, propomos três vetores de características. O primeiro vetor foi composto pelas medidas das espessuras corticais de cada estrutura cerebral do hemisfério esquerdo, contendo, portanto, 34 características. O segundo vetor foi composto pelas medidas das espessuras corticais de cada estrutura cerebral do hemisfério direito, também contendo 34 medidas. E o terceiro vetor foi composto pelas medidas de cada estrutura cerebral de ambos os hemisférios, contendo, portanto, 68 atributos.

As diferenças entre as medidas de espessuras corticais de homens e mulheres são pequenas. No cérebro humano adulto a zona mais fina é a fissura calcarina, com cerca de 2 mm; já a zona mais espessa é o giro pré-central, com cerca de 4mm (Kabani et al., 2001; Santos, 2010). Devido a estas características, não foi realizada a normalização dos valores dos vetores de características, o que possibilita uma análise considerando as medidas reais.

<sup>&</sup>lt;sup>1</sup>Massachusetts General Hospital, Boston, MA; http://surfer.nmr.mgh.harvard.edu

#### 7.1.3 Seleção de características

A seleção das características relevantes e a eliminação das irrelevantes é um problema central em mineração de imagens. A seleção de características é o problema de escolher um subconjunto menor de características que idealmente é suficiente para caracterizar bem as imagens. Aqui, a seleção de características foi usada para automaticamente detectar tecidos anormais que têm alterações mais significantes, desde que tem-se as espessuras corticais de diversas áreas do cérebro.

Neste trabalho, foi usado o algoritmo Relief. Também foram usadas as características selecionadas pelo algoritmo C4.5, o qual usa a entropia para medir quão relevante é um atributo.

#### 7.1.4 Classificação

Para determinar se a alteração na espessura cortical de cada região poderia diferenciar pacientes com ELTM de controles, no nível do sujeito, foram usados os classificadores 1-NN e árvore de decisão. Os dois classificadores são métodos de classificação binária supervisionados.

## 7.2 Experimentos

Para avaliar o método, primeiramente as imagens foram segmentadas e foram extraídas as espessuras corticais de cada estrutura cerebral para cada hemisfério cerebral, usando o FreeSurfer, como já descrito anteriormente. Então, estas medidas compuseram o vetor de características que alimentou um classificador (1-NN ou árvore de decisão C4.5), primeiramente considerando as medidas de cada hemisfério separadamente (primeiro e segundo vetores de características propostos, com 34 atributos cada um), e então todas as medidas de ambos os hemisférios (terceiro vetor de características proposto, com 68 atributos), cuja descrição dos vetores de características está na Seção 7.1.2. Para a análise do poder de caracterização das imagens, cada vetor de característica foi usado na tarefa de classificação, a qual usou os classificadores 1-NN e C4.5. Para analisar quais destas características possuíam maior relevância para categorizar as imagens, foi usado o algoritmo de seleção de características Relief. Então, os novos vetores de características, compostos pelas características selecionadas, foram submetidos ao algoritmo 1-NN. Com relação à análise de quais medidas foram mais relevantes para a categorização das imagens, também foram analisadas as características selecionadas pelo algoritmo C4.5 para a construção da árvore de decisão.

Para avaliar os resultados da fase de classificação, foi executada a abordagem 10-fold cross validation. Para comparar os resultados, foram calculadas as medidas de acurácia,

sensibilidade e especificidade. Uma predição ótima atinge 100% de sensibilidade (i.e., prediz todas as imagens do grupo de pacientes como paciente) e atinge 100% de especificidade (i.e., não prediz nenhuma imagem do grupo controle como paciente).

## 7.3 Resultados

A classificação em controles e pacientes foi executada usando a espessura cortical de estruturas corticais dos hemisférios esquerdo, direito ou ambos, respectivamente.

#### 7.3.1 Espessura cortical de regiões cerebrais do hemisfério esquerdo

A Tabela 7.1 mostra os resultados para os descritores compostos pela espessura cortical das estruturas do hemisfério esquerdo. Foram consideradas as 10 primeiras características selecionadas pelo algoritmo Relief das 34 que compõem o vetor de características. Este algoritmo selecionou as seguintes regiões: córtex insula, pericalcarine, giro lingual, entorrinal, giro temporal médio, córtex cingulado anterior rostral, parahipocampal, giro temporal transverso, pós-central, e temporal superior.

| Classificador | Características | Acurácia | Sensibilidade | Especificidade |
|---------------|-----------------|----------|---------------|----------------|
| 1-NN          | 34              | 100      | 100           | 100            |
| C4.5          | 34              | 99.33    | 98.37         | 100            |
| 1-NN          | 10              | 100      | 100           | 100            |
| 1-NN          | 3               | 99.66    | 99.19         | 100            |

Tabela 7.1: Espessura cortical de estruturas cerebrais do hemisfério esquerdo.

Pode-se obervar que os valores de acurácia, sensibilidade e especificidade alcançaram 100% de acerto, usando o classificador 1-NN e considerando tanto todas as características do vetor (34 características) como também considerando apenas as 10 primeiras selecionadas pelo algoritmo Relief. Ou seja, com menos de 30% do esforço computacional pode-se obter os mesmos resultados que são obtidos com o conjunto completo de características.

Os resultados usando o C4.5 também são significantes, e a árvore de decisão gerada na fase de treinamento mostra a relação entre o córtex pericalcarino e a ínsula, como descrito a seguir:

SE córtex perical carino  $\leq 2.096~{\rm mm}$ 

SE ínsula  $\leq 2.115$  mm ENTÃO sujeito normal;

SE ínsula > 2.115mm ENTÃO sujeito tem ELTM

SE córtex pericalcarino > 2.096 mm ENTÃO sujeito normal

Estas regras mostram que a espessura cortical do córtex pericalcarino e da ínsula estão correlacionados quando a espessura do córtex pericalcarino é menor ou igual a 2.096mm.

(7.1)

#### 7.3.2 Espessura cortical de regiões cerebrais do hemisfério direito

A Tabela 7.2 apresenta os resultados usando as medidas de espessura cortical apenas de regiões cerebrais do hemisfério direito. As 10 primeiras características selecionadas pelo algoritmo Relief das 34 que compõem o vetor de características foram: córtex da ínsula, córtex pericalcarino, giro lingual, entorrinal, parahipocampal, giro temporal médio, córtex cingulado anterior rostral, pós-central, giro temporal transversal, e giro rostral médio frontal.

| Classificador | Características | Acurácia | Sensibilidade | Especificidade |
|---------------|-----------------|----------|---------------|----------------|
| 1-NN          | 34              | 99.33    | 99.19         | 99.43          |
| C4.5          | 34              | 98.65    | 97.56         | 99.43          |
| 1-NN          | 10              | 100      | 100           | 100            |
| 1-NN          | 3               | 99.33    | 99.19         | 99.43          |

Tabela 7.2: Espessura cortical de estruturas cerebrais do hemisfério esquerdo.

Pode-se observar que os maiores valores de acurácia, sensibilidade e especificidade foram obtidos usando o classificador 1-NN e apenas usando as 10 características selecionadas pelo algoritmo Relief das 34 que compõem foi possível obter valores até 100%. Entretanto, todos os resultados mostraram uma alta taxa de acurácia e especificidade, com valores acima de 99%.

O algoritmo C2.5 também selecionou as mesmas características quando executado usando as medidas do hemisfério esquerdo: córtex pericalcarino e ínsua, e a relação entre elas está apresentada nas seguintes regras:

SE córtex perical carino  $\leq 2.078~{\rm mm}$ 

SE ínsula  $\leq 2.264 \text{ mm} \text{ ENTÃO}$  sujeito normal;

(7.2)

SE ínsula > 2.264mm ENTÃO sujeito tem ELTM

SE córtex perical carino  $> 2.078~\mathrm{mm}$  ENTÃO sujeito normal

Neste experimento, a relação entre ínsula e o córtex pericalcarino é a mesma apresentada usando as medidas do hemisfério esquerdo.

#### 7.3.3 Espessura cortical de regiões cerebrais de ambos hemisférios

Neste experimento, foram consideradas as medidas de espessura cortical de ambos os hemisférios. Os resultados estão apresentados na Tabela 7.3. As 10 primeiras características selecionadas pelo algoritmo Relief das 68 que compõem o vetor de características foram: córtex insular esquerdo, córtex pericalcarino esquerdo, córtex insular direito, córtex pericalcarino direito, giro lingual direito, giro lingual esquerdo, giro entorrinal direito, giro entorrinal esquerdo, giro cingulado rostral anterior, e parahipocampal direito.

| Classificador | Características | Acurácia | Sensibilidade | Especificidade |
|---------------|-----------------|----------|---------------|----------------|
| 1-NN          | 68              | 100      | 100           | 100            |
| C4.5          | 68              | 98.33    | 98.37         | 100            |
| 1-NN          | 10              | 100      | 100           | 100            |
| 1-NN          | 3               | 99.66    | 99.19         | 100            |

Tabela 7.3: Espessura cortical de estruturas cerebrais do hemisfério esquerdo.

As características selecionadas pelo C4.5 para construir a árvore de decisão foram giro pericalcarino direito (dir-pericalcarino) e córtex insular esquerdo (esq-ínsula). A relação entre estas medidas está apresentada nas seguintes regras:

SE dir-perical carino  $\leq 2.078~{\rm mm}$ 

> SE esq-ínsula  $\leq 2.115 \text{ mm ENTÃO sujeito normal;}$ ~ (7.3)

SE esq-ínsula > 2.115mm ENTAO sujeito tem ELTM

SE córtex perical carino  $> 2.078~\mathrm{mm}$  ENTAO sujeito normal

## 7.4 Discussão dos resultados

Os resultados alcançados enfatizam que mudanças patológicas específicas estão presentes no neocórtex de pacientes com ELTM, representado pela redução da espessura cortical. Esta é uma abordagem que mostra que é possível detectar ELTM analisando não apenas reduções de volume, como é explorado em outros estudos.

Os resultados mostram que a espessura cortical é uma boa medida para representar imagens e detectar ELTM. Os resultados obtidos pelo método proposto alcançam valores mais altos de acurácia que outros métodos da literatura. Keihaninejad et al. (2010) alcançou 96% de acurácia usando volumes estruturais como características e o classificador SVM (*Support Vector Machine*). McDonald et al. (2008) classificou com precisão de 74% ambos os grupos usando todas as medidas de espessura cortical e o classificador LDA (*Linear Discriminant Analisys*). O método proposto neste trabalho alcançou 100% de acurácia, sensibilidade e especificidade executando o classificador 1-NN, que pode ser facilmente implementado utilizando estruturas de índice métricos em sistemas de gerenciamento de bases de dados, além de ter custo computacional menor do que o SVM e o LDA.

Outra vantagem do 1-NN sobre outros classificadores é que é um método de classificação simples. Além disso, o 1-NN não possui a fase de treinamento (Kordos et al., 2010), e o tempo da fase de teste pode ser significantemente reduzido pela referência a

priori de uma seleção do vetor, o que precisa ser realizado apenas uma vez ou aplicando estratégias avançadas de buscas dos vizinhos mais próximos.

A vantagem do C4.5 sobre o SVM e o LDA é que este gera regras que podem ser interpretadas pelo usuário, o qual torna-o apto a estudar as relações entre as características.

Além disso, analisando os resultados presentes nas tabelas e as características selecionadas, pode-se inferir que alterações na espessura cortical foram detectadas em ambos os hemisférios em pacientes com ELTM quando comparados aos controles, por causa dos três resultados a seguir. Primeiro, o Relief selecionou as mesmas 9 características de um subconjunto de 10 características para ambos os hemisférios. Segundo, os valores de acurácia, sensibilidade e especificidade usando apenas espessura cortical (esquerda ou direita) alcançaram 100%. Terceiro, quando foram usadas espessuras corticais de estruturas de ambos hemisférios, foram alcançados os mesmos valores de 100%.

As 10 características selecionadas pelo Relief mostram que a perda de espessura cortical aconteceu nas regiões posterior temporal, parietal, ocipital e sensorial-motoras. Estas regiões foram também encontradas em outro estudo (Dabbs et al., 2009), no qual foi analisada a espessura cortical para correlacionar os fenótipos cognitivos em ELT. É importante ressaltar que os experimentos propostos e realizados neste trabalho não excluem a perda da espessura cortical em outras regiões, mas foram usadas apenas 10 características que classificaram com sucesso tanto o grupo pacientes quanto o controle.

Para analisar a relação entre as características, pode-se observar que as características que foram selecionadas pelo algoritmo C4.5 para a construção da árvore de decisão, e as regras geradas. Este classificador selecionou apenas duas regiões em ambos hemisférios: o córtex insular e o córtex pericalcarino.

Anatomicamente, a ínsula tem extensivas conexões com o sistema límbico, incluindo o giro cingulado, córtex entorrinal, amídala e hipocampo. Assim, crises que surgem do lobo temporal mesial rapidamente se espalham para a ínsula (Schwartz, 2005), o que explica a alteração expressiva da espessura cortical nesta região.

O córtex pericalcarino também está fisicamente e diretamente conectado com o lobo temporal, na porção média. Também está conectado com o córtex ocipital e sua fronteira lateral está conectada com o sulco inferior-médio e o sulco calcarino. Pode-se concluir, então, que a ELTM causa alterações na espessura cortical nesta região. Em Jenkins et al. (2011) são apresentadas as seguintes associações: a perda do volume do córtex pericalcarino e o subsequente desenvolvimento de esclerose múltipla e ELTM. Isto reforça a relação entre alteração no córtex pericalcarino e a presença de ELTM.

## 7.5 Considerações finais

Baseado nos resultados alcançados, pode-se afirmar que a medida de espessura cortical é significante em pacientes com ELTM. Dessa forma, a espessura cortical pode ser usada para classificar com sucesso pacientes e controles usando um dos classificadores mais simples, o 1-NN. A seleção de apenas 10 características com o algoritmo Relief e o uso do algoritmo C4.5 torna possível o estudo da relação entre as características que apresentam alterações mais significantes. Além disso, também pode-se encontrar que esta alteração é bilateral, pois pode-se usar apenas a espessura cortical de um dos hemisférios (esquerdo ou direito) e ainda assim alcançar uma precisão de 100% de acerto da classificação. Uma importante aplicação desta técnica é que pode ser usada para detectar automaticamente tecidos em um paciente individual para avaliação pré-cirúrgica. Estudos futuros devem usar esta metodologia em pacientes com ELTM que não apresentam mudanças significativas no volume do hipocampo e que visa a identificação da lateralidade do foco epiléptico.

# Capítulo 8 Conclusão

Sistemas de apoio à decisão clínica são desenvolvidos visando diferentes aspectos de cuidado ao paciente, incluindo prevenção, diagnóstico e planejamento de tratamento. É muito importante enfatizar que os sistemas de apoio à decisão são destinados a complementar e não suplantar o médico no processo de tomada de decisão. Em outras palavras, tais sistemas são destinados a auxiliar o especialista em realizar escolhas dentre diversas alternativas. Esses sistemas não são destinados a automatizar o processo de tal modo que a escolha seja imposta ao usuário. Embora alguns sistemas de apoio à decisão sejam projetados para fornecer recomendações específicas para a consideração, o usuário especialista revisa as sugestões e pode vir a rejeitá-las em favor de uma alternativa diferente (Muralidhar et al., 2011). Portanto, é muito importante que a precisão das sugestões de diagnóstico destes sistemas seja alta, que o conhecimento obtido seja de fácil interpretação e compreensão e que forneça respostas em um tempo aceitável ao usuário. Assim, esta tese buscou desenvolver métodos de mineração de imagens para aumentar a precisão de sistemas de apoio ao diagnóstico por imagens e para gerar conhecimento de fácil compreensão ao usuário em um tempo aceitável.

## 8.1 Principais contribuições

O trabalho desenvolvido e apresentado nesta tese trouxe contribuições para a área de sistemas de auxílio ao diagnóstico e mineração de imagens. As principais contribuições envolveram o desenvolvimento de métodos e técnicas de extração de características de

imagens e de classificadores associativos e também o desenvolvimento de um *framework* de sistema de auxílio ao diagnóstico.

A primeira contribuição foi a redução do *qap* semântico existente entre a representação numérica das imagens e sua interpretação humana por meio do desenvolvimento de extratores de características baseados em redes complexas. A inovação deste método está na modelagem da forma da imagem em uma rede complexa livre de escala a partir de seu contorno e esqueleto. Os extratores propostos, baseados em medidas da topologia da rede, geraram vetores de características compactos, os quais possuem de dois a sete atributos, também lidando, portanto, com o problema da maldição da alta dimensionalidade dos dados. Os experimentos realizados mostraram que os novos descritores propostos representam melhor as imagens que descritores tradicionais da literatura, como os momentos de Zernike e texturas de Haralick, gerando um resultado de classificação mais precisa. A vantagem de se usar redes complexas para a representação de imagens em relação à maioria dos descritores de forma é que não é necessário que os pixels do contorno das imagens sejam adjacentes e sequenciais, já que o modelo leva em consideração apenas a distância entre os elementos do esqueleto até algum ponto do contorno. Isso permite que o método proposto gere descritores invariantes à escala, rotação e à pequenas deformações. Por fim, diferente do método de Backes et al. (2009), que usa redes complexas de pequeno mundo para modelar imagens e recria a rede diversas vezes para a extração das características, o método desenvolvido gera a rede apenas uma vez.

A segunda contribuição deste trabalho foi o desenvolvimento de dois classificadores associativos, o SACMiner e o MinSAR. Os algoritmos desenvolvidos buscaram evitar os problemas que a maioria dos classificadores associativos da literatura possuem, que são: o alto custo computacional de algoritmos para descobrir um conjunto de itens frequentes; a discretização dos dados contínuos, o que pode gerar perda de informação e inconsistência nos dados; e a geração de um grande número de regras, dificultando o entendimento do conhecimento obtido e gerando ambiguidade de regras. Tanto o SACMiner quanto o MinSAR sobrepujaram estes desafios por meio do uso de regras de associação estatísticas. O SACMiner é um classificador que evita a fase de discretização de dados e gera regras simples e diretas, que facilitam a compreensão do especialista médico na análise do diagnóstico. Além disso, este classificador possui um algoritmo de geração de regras de ordem linear com relação ao número de atributos no vetor de características e ao número de classes, percorrendo a base de dados apenas uma vez. Dessa forma, o algoritmo evita a realização de muitos cálculos, como é feito na maioria dos algoritmos de classificadores associativos. Isso é possível porque o algoritmo realiza seleção de atributos e mineração de regras de associação estatísticas em um único passo. O processo de seleção de atributos ainda contribui para reduzir o número de características selecionadas para a geração de regras, lidando ao mesmo tempo com o problema da maldição da alta dimensionalidade dos dados e com o problema da geração de grande número de regras. A precisão de classificação do SACMiner se mostrou superior quando comparada aos resultados obtidos por classificadores bem estabelecidos na literatura como o C4.5, naive bayes e 1-NN. Já o classificador MinSAR é uma extensão do classificador SACMiner, e, portanto, possui todas as vantagens do algoritmo original, com o adicional de não requerer do usuário parâmetros de entrada ao algoritmo de geração de regras, gerando, portanto, um conjunto de regras compacto e relevante, evitando a geração de regras espúrias. Além da maior precisão de classificação destes métodos com relação aos classificadores C4.5, naive bayes, 1-NN e AdaBoost, outra vantagem é que o conhecimento obtido é apresentado por meio de regras simples SE-ENTÃO, o que permite uma fácil compreensão pelo especialista, que pode então analisar o que levou o classificador a tomar determinada decisão, fazer análises das relações das medidas das características com a classe a que estão relacionadas, tornando o sistema mais confiável para ser utilizado.

A terceira contribuição deste trabalho foi o desenvolvimento de um *framework* de sistema de auxílio ao diagnóstico de câncer de mama usando os classificadores associativos descritos anteriormente. A vantagem deste *framework* com relação a outros da literatura é que apresenta uma classificação mais precisa em relação a outros e apresenta o conhecimento obtido na fase de aprendizagem, o que torna o sistema mais confiável para ser utilizado pelo especialista. Isto ocorre devido aos classificadores desenvolvidos utilizados e às características de forma usadas no sistema, as quais carregam uma semântica relevante, como área do tumor detectado, tamanho do eixo menor, dentre outras.

A quarta e última contribuição foi o desenvolvimento de uma metodologia para o auxílio ao diagnóstico de epilepsia no lobo temporal mesial (ELTM) por meio de exames de ressonância magnética. Um problema na representação de imagens é a diferença entre a representação numérica das imagens e sua representação visual (gap semântico). Nesta metodologia foram consideradas medidas de espessura das estruturas corticais. Essas medidas podem ser consideradas características de alto nível, pois agregam conteúdo semântico, tendo um significado bem definido para o especialista. Assim, o uso destas medidas permitiu reduzir tanto o *qap* semântico quanto gerar informações compreensíveis e acessíveis ao radiologista. A inovação desta metodologia é que não usa medidas de volume do hipocampo, como tradicionalmente o diagnóstico é realizado, já que mudanças sutis no volume do hipocampo podem atrapalhar a detecção da ELTM. Assim, este método possui a vantagem de classificar pacientes corretamente, mesmo que estes possuam mudanças sutis no hipocampo. A outra vantagem é que a precisão do diagnóstico usando estas características é alta, chegando a 100% de precisão com o uso de classificadores com baixo custo computacional como o classificador que usa o vizinho mais próximo (1-NN). Por fim, esta metodologia realiza uma análise das medidas que foram mais significativas no processo de classificação. Esta análise é realizada por meio da aplicação do algoritmo

de seleção de atributos Relief e também são considerados os atributos selecionados no modelo de construção da árvore de decisão pelo algoritmo C4.5. Dessa forma, além da classificação da imagem, a metodologia desenvolvida é capaz de sugerir um diagnóstico ao médico que apresenta a relação das estruturas mais significativas para o processo de classificação, o que não é possível quando são usados apenas classificadores como o LDA e o SVM, que são as abordagens da literatura.

## 8.2 Publicações

Esta tese também gerou um conjunto de trabalhos publicados em eventos nacionais e internacionais, os quais estão listados a seguir.

#### Capítulos de livros internacionais

- WATANABE, C. Y. V., RIBEIRO, M. X., TRAINA JUNIOR, C., TRAINA, A. J. M. SACMiner: A New Classification Method Based on Statistical Association Rules to Mine Medical Images. Enterprise Information Systems, Lecture Notes in Business Information Processing, v. 73, p. 249-263, 2011.
- TRAINA, A. J. M., TRAINA JUNIOR, C., BALAN, A. G. R., RIBEIRO, M. X., BUGATTI, P. H., WATANABE, C. Y. V., AZEVEDO-MARQUES, P. M. Feature Extraction and Selection for Decision Making. In: Thomas Martin Deserno. (Org.). Biomedical Image Processing. 1sted.Heidelberg: Springer Berlin Heidelberg, 2011, v., p. 197-223.

#### Congressos internacionais - artigos completos

- WATANABE, C. Y. V., RAMOS, J. S., TRAINA, A. J. M., TRAINA JUNIOR, C. A CAD system based on complex networks theory to characterize mass in mammograms. In: SPIE Medical Imaging 2012: Computer-Aided Diagnosis. San Diego, CA, 2012. v. 8315. p. 1-12.
- WATANABE, C. Y. V., RIBEIRO, M. X., TRAINA JUNIOR, C., TRAINA, A. J. M. Statistical Associative Classification of Mammograms: The SACMiner Method. In: 12th International Conference on Enterprise Information Systems
  ICEIS2010, 2010, Funchal. Proceedings of 12th International Conference on Enterprise Information Systems - Artificial Intelligence and Decicion Support Systems. Portugal: SciTePress, 2010. v. 2. p. 121-128.

#### Congresso internacional - short paper

 WATANABE, C. Y. V., RIBEIRO, M. X., TRAINA, A. J. M., TRAINA JUNIOR, C. A Statistical Associative Classifier with Automatic Estimation of Parameters on Computer Aided Diagnosis. In: 2012 Eleventh International Conference on Machine Learning and Applications (ICMLA), 2012, Boca Raton. 2012 11th International Conference on Machine Learning and Applications, 2012. v. 1. p. 564-567.

#### Workshop internacional - artigo completo

 SILVA, C. Y. V. Watanabe, RIBEIRO, M. X., TRAINA, A. J. M., TRAINA JUNIOR, C. Employing Wavelet Transforms to Support Content-Based Retrieval of Medical Images. In: 8th International Workshop on Pattern Recognition in Information Systems (PRIS 2008), 2008, Barcelona. Anais do PRIS 2008. p. 19-28.

#### Congressos nacionais - artigos completos

- WATANABE, C. Y. V., DINIZ, P. R. B., VELASCO, T. R., HILDENBERG, C. L. O., TRAINA, A. J. M. Automatically detecting temporal lobe epilepsy using cortical thickness measures. In: XXIII Congresso Brasileiro de Engenharia Biomédica, 2012, Porto de Galinhas. XXIII Congresso Brasileiro de Engenharia Biomédica, 2012. p. 1255-1259.
- SILVA, C. Y. V. Watanabe, BUGATTI, P. H., RIBEIRO, M. X., TRAINA JUNIOR, C., TRAINA, A. J. M. Improving CBIR Using Feature Extraction Based on Wavelet Transform. In: XIV Simpósio Brasileiro de Sistemas Multimedia e Web (WebMedia 2008), 2008, Vila Velha, ES. Anais do WebMedia 2008. Porto Alegre: Sociedade Brasileira de Computação, 2008. v. 1. p. 51-58.

## 8.3 Proposta de trabalhos futuros

Apesar desta tese ter trazido contribuições relevantes à área de sistemas CAD e mineração de imagens e ter atingido os objetivos propostos inicialmente, os desafios destas áreas ainda não foram totalmente solucionados. Assim, os principais direcionamentos para outros trabalhos que podem ser realizados a partir desta tese estão listados a seguir:

• Extensão da metodologia de detecção automática de ELTM: neste trabalho foram considerados apenas pacientes com epilepsia unilateral (ou seja, com foco epiléptico em um dos lados). É necessário que sejam adquiridos exames de pacientes que possuam epilepsia bilateral e buscar informar o foco epiléptico, além da classificação em pacientes e controles. Isso seria ainda mais relevante ao médico.

- Aplicação dos classificadores SACMiner e MinSAR no auxílio ao diagnóstico de ELTM: foi visto que o uso destes classificadores no auxílio ao diagnóstico de mama foi adequado, aumentando a precisão da classificação e apresentado um modelo de aprendizagem de fácil compreensão ao especialista. Como na metodologia para o diagnóstico de ELTM já foram usadas características significativas, o uso de classificadores associativos tende a gerar regras interessantes.
- Aplicar o SACMiner e o MinSAR a outros tipos de base de imagens médicas: dessa maneira, é possível avaliar o quão robustos são estes classificadores.
- Aplicação dos extratores de características baseados em redes complexas em outras bases de imagens médicas: também podem ser exploradas outras medidas topológicas e verificar quais delas caracterizam melhor cada tipo de imagem.
- Desenvolvimento de métodos de segmentação de imagens: durante o desenvolvimento dos extratores de características de forma foi necessário realizar a segmentação de imagens. Foi observado que o resultado da segmentação e da detecção das regiões de interesse interfere diretamente no processo de extração de características. Como a fase de segmentação não foi o interesse deste trabalho, foram considerados métodos da literatura para a realização da segmentação de massas em mamografias e de objetos diversos (o algoritmo EM/MPM aperfeiçoado e o algoritmo de Otsu). Para avaliar os extratores propostos, ainda foram considerados base de imagens de objetos previamente segmentados. Entretanto, métodos adequados de segmentação podem aumentar a precisão dos resultados gerados.

## 8.4 Considerações finais

Por ter sido um trabalho interdisciplinar entre a computação e a medicina, além do estudo de métodos e técnicas computacionais, este trabalho também demandou o estudo sobre as bases de imagens médicas escolhidas, o que proporcionou a interação da aluna com pesquisadores de ambas as áreas. É importante ressaltar que os primeiros contatos iniciais com pesquisadores da área médica foram difíceis, pois a linguagem usada por cada área é muito diferente. Este foi mais um desafio que teve que ser vencido, além dos propostos na literatura, que é a compreensão do problema em um domínio diferente do domínio computacional e a aquisição e tratamento de imagens.

Talvez seja por esses motivos que o desenvolvimento de sistemas de auxílio ao diagnóstico médico seja uma área tão fascinante e muito pesquisada atualmente, além de ter o objetivo final de melhorar a saúde e o cuidado do ser humano.

## **Referências Bibliográficas**

- Abraham, R., Simba, J. B., e Iyengar, S. S. (2006). A comparative analysis of discretization methods for Medical Datamining with Naive Bayesian classifier. In 9th International Conference on Information Technology (ICIT'06), pages 1–2.
- Agrawal, R., Imielinski, T., e Swami, A. N. (1993). Mining association rules between sets of items in large databases. In Buneman, P. e Jajodia, S., editors, ACM SIGMOD International Conference on Management of Data, volume 1, pages 207–216. ACM Press.
- Agrawal, R. e Shafer, J. C. (1996). Parallel mining of association rules. *IEEE Transactions on Knowledge and Data Engineering*, 8:962–969.
- Agrawal, R. e Srikant, R. (1994). Fast algorithms for mining association rules. In International Conference on Very Large Databases (VLDB), pages 487–499.
- Akgül, C. B., Rubin, D. L., Napel, S., Beaulieu, C. F., Greenspan, H., e Acar, B. (2011). Content-based image retrieval in radiology: current status and future directions. *Journal of Digital Imaging*, 24(2):208–222.
- Akilandeswari, U., Nithya, R., e Santhi, B. (2012). Review on feature extraction methods in pattern classification. *European Journal of Scientific Research*, 71(2):265–272.
- Albert, R. e Barabasi, A. L. (2001). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(47):1–54.
- Allen, E. A., Erhardt, E. B., e Calhoun, V. D. (2012). Data visualization in the neurosciences: Overcoming the curse of dimensionality. *Neuron*, 74:603–608.
- Antonie, M. L., Chodos, D., e Zaïane, O. (2009). Variations on Associative Classifiers and Classification Results Analyses, chapter 9, pages 150–172. ICI Global.
- Arimura, H., Magome, T., Yamashita, Y., e Yamamoto, D. (2009). Computer-aided diagnosis systems for brain diseases in magnetic resonance images. *Algorithms*, 2(3):925–952.
- Asuncion, A. e Newman, D. (2007). UCI Machine Learning Repository. http://archive.ics.uci.edu/ml/datasets.html.

- Aumann, Y. e Lindell, Y. (1999). A statistical theory for quantitative association rules. In Press, A., editor, *The fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 261–270.
- Azavedo, E., Zackrisson, S., Ingegerd, M., e Arnlind, M. H. (2012). Is single reading with computer-aided detection (CAD) as good as double reading in mammography screening? a systematic review. *BMC Medical Imaging*, 12(22):1–12.
- Backes, A. R. (2010). Estudos de métodos de análises de complexidade em imagens. Tese de doutorado.
- Backes, A. R., Casanova, D., e Bruno, O. M. (2007). Método de aproximação poligonal de contornos utilizando redes complexas. *Journal of Computer Science INFOCOMP*, 6(2):71–80.
- Backes, A. R., Casanova, D., e Bruno, O. M. (2009). A complex network-based approach for boundary shape analysis. *Pattern Recognition*, 42(1):54–67.
- Balan, A. G. R. (2007). Métodos adaptativos de segmentação aplicados à recuperação de imagens por conteúdo. Tese de doutorado.
- Barabasi, A. L. e Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.
- Barabasi, A. L. e Bonabeau, E. (2003). Scale-free networks. *Scientific American*, May 2013:50–59.
- Baralis, E. e Garza, P. (2012). I-prune: item selection for associative classification. International Journal of Intelligent Systems, 27(3):279–299.
- Barrat, A., Barthélemy, M., e Vespignani, A. (2004). Modeling the evolution of weighted networks. *Physical Review E*, 70:1–13. arXiv:cond-mat/0406238.
- Barrat, A., Barthélemy, M., e Vespignani, A. (2008). Dynamical Processes in Complex Networks. Cambridge University Press, UK.
- Barthélemy, M. (2011). Spatial networks. *Physics Reports*, 499(2011):1–101.
- Becker, H., Naaman, M., e Gravano, L. (2010). Learning similarity metrics for event identification in social media. In 3rd ACM International Conference on Web Search and Data Mining, pages 291–300.
- Bellazzi, R. e Zupan, B. (2008). Predictive data mining in clinical medicine: Current issues and guidelines. *International Journal of Medical Informatics (IJMI)*, 77:81–97.
- Blümcke, I., Coras, R., Miyata, H., e Özkara, C. (2012). Defining clinico-neuropathological subtypes of mesial temporal lobe epilepsy with hippocampal sclerosis. *Brain Pathology*, 22(3):402–411.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., e Hwang, D. U. (2006). Complex networks: structure and dynamics. *Physics Reports*, 424(4-5):175–308.

- Bugatti, P. H., Traina, A. J. M., Felipe, J. C., e Traina Jr., C. (2008). A new method to efficiently reduce histogram dimensionality. In Giger, M. L. e Karssemeijer, N., editors, SPIE Medical Imaging 2008: Computer-Aided Diagnosis, volume 1, pages 1–9. Proceedings of SPIE.
- Caldarelli, G. (2007). Scale-free networks. Oxford University Press, Oxford.
- Cardie, C. (1993). Using decision trees to improve case-based learning. In 10th International Conference on Machine Learning, pages 25–32.
- Casanova, D., Backes, A. R., e Bruno, O. M. (2013). Pattern recognition tool based on complex network-based approach. *journal of Physics: Conference Series*, 410:1–4.
- Chen, G., Liu, H., Yu, L., Wei, Q., e Zhang, X. (2006). A new approach to classification based on association rule mining. *Decision Support Systems*, 42:674–689.
- Christopher, J. J. (2011). A statistical approach for associative classification. *European Journal of Scientific Research*, 58(2):140–147.
- Ciresan, D., Meier, U., e Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. In *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), pages 3642–3649.
- Comer, M. L. e Delp, E. J. (2000). The EM/MPM Algorithm for Segmentation of Textured Images: Analysis and Further Experimental Results. *IEEE Transactions on Image Processing*, 9(10):1731–1744.
- Costa, A. F. (2012). Mineração de imagens médicas utilizando características de forma. Dissertação de mestrado.
- Costa, L. F. d. (2004). Complex networks, simple vision. ArXiv Condensed Matter e-prints, pages 1–6. cond-mat/0403346.
- Costa, L. F. d. e Cesar Jr, R. M. (2009). Shape Analysis and Classification Theory and Practice. CRC Press, Boca Raton, CA.
- Costa, L. F. d., Oliveira Jr., O. N., Travieso, G., Rodrigues, F. A., Boas, P. R. V., Antiqueira, L., Viana, M. P., e Rocha, L. E. C. (2011). Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Advances in Physics*, 60(3):329–412.
- Costa, L. F. d., Rodrigues, G. T., e Boas, P. R. V. (2007). Characterization of complex networks: A survey of measurements. *Advances in Physics*, 56(1):167–242.
- Dabbs, K., Jones, J., Seidenberg, M., e Hermann, B. (2009). Neuroanantomical correlates of cognitive phenotypes in temporal lobe epilepsy. *Epilepsy Behaviour*, 15(4):445–451.
- Dale, A. M., Fischl, B., e Sereno, M. I. (1999). Cortical surface-based analysis I: segmentation and surface reconstruction. *Neuroimage*, 9:179–194.
- Das, K. e Vyas, O. P. (2010). A suitability study of discretization methods for associative classifiers. *International Journal on Computer Applications*, 5(10):46–51.

- Dash, M. e Liu, H. (1997). Feature selection for classification. International Journal of Intelligent Data Analysis, 1(3):131–156.
- Deselaers, T., Keysers, D., e Ney, H. (2008). Features for image retrieval: an experimental comparison. *Information Retrieval*, 11(2):77–107.
- Deshpande, D. S. (2011). Association rule mining based on image content. *International journal of Information Technology and Knowledge Management*, 4(1):143–146.
- Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., e Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, 31:968–980.
- Devore, J. L. (2011). Probability and Statistics for Engineering and the Sciences. Duxbury Press, 8th edition.
- Dorogovtsev, S. N. e Mendes, J. F. F. (2003). Evolution of Networks: from Biological nets to the Internet and WWW. Oxford University Press, 1 edition.
- Dougherty, G. (2009). Digital Image Processing for Medical Applications. Cambridge University Press.
- Dua, S., Singh, H., e Thompson, H. W. (2009). Associative classication of mammograms using weighted rules. *Expert Systems with Applications*, 36(5):9250–9259.
- Edelstein, H. A. (1999). Introduction to Data Mining and Knowledge Discovery. Two Crows Corporation.
- Elnakib, A., Gimelâfarb, G., Suri, J. S., e El-Baz, A. (2011). *Medical Image Segmentation:* A Brief Survey, chapter 1, pages 1–39. Springer New York.
- Engel Jr., J. (2008). Surgical treatment for epilepsy. Journal of the American Medical Association (JAMA), 300(21):2548–2550.
- Erdös, P. e Rényi, A. (1959). On random graphs. Publicationes Mathematicae, 6:290–297.
- Fan, J. (2012). Feature Learning Based Multi-scale Wavelet Analysis for Textural Image Segmentation, volume 1 of Advances in Intelligent and Softing Computing, pages 461–466. Springer-Verlag Berlin Heidelberg.
- Farid, N., Girard, H. M., Kemmotsu, N., Smith, M. E., Magda, S. W., Lim, W. Y., Lee, R. R., e McDonald, C. R. (2012). Temporal lobe epilepsy: Quantitative MR volumetry in detection of hippocampal atrophy. *Radiology*, 264(2):542–550.
- Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., e Uthurusamy, R. (1996). Advances in Knowledge Discovery and Data Mining. MIT Press, Cambridge, MA.
- Felipe, J. C., Traina, A. J. M., e Traina Jr., C. (2003). Retrieval by content of medical images using texture for tissue identification. In 16th IEEE Symposium on Computer-based Medical Systems, pages 175–180. IEEE Computer Society.

- Felipe, J. C., Traina, A. J. M., Traina Jr., C., Sousa, E. P. M. d., e Ribeiro, M. X. (2006). Effective shape-based retrieval and classification of mammograms. In 21st Annual ACM Symposium on Applied Computing (SAC'2006), pages 1–6.
- Feng, D., Yang, J., e Liu, C. (2012). An efficient indexing method for content-based image retrieval. *Neurocomputing*, pages 1–12.
- Fischl, B. e Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of* the United States of America, 97:11050–11055.
- Fischl, B., Liu, A., e Dale, A. M. (2001). Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Transactions on Medical Imaging*, 20:70–80.
- Florindo, J. a. B., Casanova, D., e Bruno, O. M. (2013). Fractal measures of complex networks applied to texture analysis. *Journal of Physics: Conference Series*, 410:1–5.
- Focke, N. K., Yogarajah, M., Symms, M. R., Gruber, O., Paulus, W., e Duncan, J. S. (2012). Automated MR image classification in temporal lobe epilepsy. *Neuroimage*, 59(1):356–362.
- Freund, Y. e Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139.
- Gabor, D. (1946). Theory of communication. IEE (London), 93, Part III(26):429-457.
- Ganesan, K., Acharya, U. R., Chua, C. K., Min, L. C., Abraham, K. T., e Ng, K.-H. (2011). Computer-aided breast cancer detection using mammograms: a review. *IEEE Reviews in Biomedical Engineering*, pp(99):1–21.
- Gao, M., Huang, J., Huang, X., Zhang, S., e Metaxas, D. N. (2012). Simplified labeling process for medical image segmentation, volume 7511, pages 387–394. Springer Berlin Heidelberg.
- Gerhardinger, L. C. (2006). Segmentação de imagens e validação de classes por abordagem estocástica. Dissertação de mestrado.
- Geusebroek, J.-M., Burghouts, G. J., e Smeulders, A. W. M. (2005). The Amsterdam Library of Object Images. *International Journal of Computer Vision*, 61(1):103–112.
- Glaser, A. N. (2001). High-Yield Biostatistics. Lippincott Williams and Wilkins.
- Gonzalez, R. C. e Woods, R. E. (2008). *Digital Image Processing*. Pearson Prentice Hall, 3rd edition.
- Gonzalez, R. C., Woods, R. E., e Eddins, S. L. (2003). Digital Image Processing Using MATLAB. Pearson Prentice Hall.
- Han, J., Cheng, H., Xin, D., e Yan, X. (2007). Frequent pattern mining: current status and future directions. *Data Mining and Knowledge Discovery*, 15(1):55–86.

- Han, J. e Kamber, M. (2006). *Data mining: concepts and techniques*. Morgan Kaufmann, San Francisco, CA, second edition.
- Han, J. W., Pei, J., Yin, Y. W., e Mao, R. Y. (2004). Mining frequent patterns without candidate generation: a frequent-pattern tree approach. *Data Mining and Knowledge Discovery*, 8(1):53–87.
- Han, X., Jovicich, J., Salat, D., Kouwe, A. v. d., Quinn, B. T., Czanner, S., Busa, E., Pacheco, J., Albert, M. S., Killiany, R. J., Maguire, R. P., Rosas, D., Makris, N., Dale, A. M., Dickerson, B. C., e Fischl, B. (2006). Reliability of MRI-derived measurements of human cerebral cortical thickness: Th effects of field strength, scanner upgrade and manufacturer. *NeuroImage*, 32:180–194.
- Haralick, R. M., Shanmugam, K., e Distein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(6):610–621.
- Hastie, T., Tibshirani, R., e Friedman, J. (2011). *The elements of statistical learning: data mining, inference and prediction.* Springer Series in Statistics. Springer New York, New York.
- Heravi, M. J. e Zaïane, O. (2010). A study on interestingness measures for associative classifiers. In ACM Symposium on Applied Computing (SAC'2010), pages 1039–1046, New York.
- Holt, J. D. e Chung, S. M. (2002). Mining association rules using inverted hashing and pruning. *Information Processing Letters*, 83(4):211–220.
- Holte, R. C. (1993). Very simple classification rules perform well on most commonly used datasets. *Machine Learning*, 11:63–91.
- Hsu, W., Lee, M. L., e Zhang, J. (2002). Image mining: Trends and developments. *Journal* of Intelligent Information Systems, 19(1):7 23. ISSN:0925-9902.
- Huang, M., Yu, W., e Zhu, D. (2012). An improved image segmentation algorithm based on the Otsu method. In 13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, pages 135–139.
- Jenkins, T. M., Ciccarelli, O., Atzori, M., Wheeler-Kingshott, C. A. M., Miller, D. H., Thompson, A. J., e Toosy, A. T. (2011). Early pericalcarine atrophy in acute optic neuritis is associated with conversion to multiple sclerosis. *Journal of Neurology and Neurosurgery. Psychiatry*, pages 1–5.
- Jeong, S., Kim, S.-W., e Choi, B.-U. (2009). Dimensionality reduction for similarity search with the euclidean distance in high-dimensional applications. *Multimedia Tools and Applications*, 42(2):251–271.
- Jia, H., Yap, P.-T., e Shen, D. (2012). Iterative multi-atlas-based multi-image segmentation with tree-based registration. *NeuroImage*, 59:422–430.
- Jiang, Y., Nishikawa, R. M., Schmidt, R. A., Toledano, A. Y., e Doi, K. (2001). Potential of computer-aided diagnosis to reduce variability in radiologists' interpretations of mammograms depicting microcalcifications. *Radiology*, 220(3):787–794.

- Johnson, C. R. (2012). Biomedical visual computing: Case studies and challenges. Computing in Science Engineering, 14(1):12–21.
- Jose, J. S., Sivakami, R., Maheswari, N. U., e Venkatesh, R. (2012). An efficient diagnosis of kidney images using association rules. *International Journal of Computer Technology* and Electronics Engineering (IJCTEE), 2(2):14–20.
- Kabani, N., Goualher, G. L., MacDonald, D., e Evans, A. C. (2001). Measurement of cortical thickness using an automated 3-D algorithm: A validation study. *NeuroImage*, 13:375–380.
- Kantardzic, M. (2011). Data Mining: Concepts, Models, Methods, and Algorithms. Wiley-IEEE Press, second edition.
- Keihaninejad, S., Heckemann, R. A., Gousias, I. S., Aljabar, P., Hajnal, J. V., Rueckerf, D., e Hammers, A. (2010). Automatic volumetry can reveal visually undetected disease features on brain MR imagens in temporal lobe epilepsy. *Proceeding of International Symposium on Biomedical Imaging: From Nano to Macro*, pages 105–108.
- Kira, K. e Rendell, L. A. (1992). The feature selection problem: traditional methods and a new algorithm. In *Tenth National Conference on Artificial Intelligence*, pages 129–134. AAAI Press.
- Kirsch, A., Mitzenmacher, M., Pietracaprina, A., Pucci, G., Upfal, E., e Vandin, F. (2012). An efficient rigorous approach for identifying statistically significant frequent itemsets. *Journal of the ACM*, 59(3):1–22.
- Kononenko, I. (1994). Estimating attributes: analysis and extensions of relief. In *European Conference on Machine Learning*, pages 171–182. Springer Verlag.
- Kordos, M., Blachnik, M., e Strzempa, D. (2010). Do we need whatever more than k-NN? In 10th international conference on Artificial intelligence and soft computing: Part I, pages 414–421. Springer-Verlag.
- Lei, Y. K., Xu, Y. M., Yang, J. A., Ding, Z. G., e Gui, J. (2012). Feature extraction using orthogonal discriminant local tangent space alignment. *Pattern Analysis and Applications*, 15(3):249–259.
- Li, S., Lee, M. C., e Pun, C. M. (2009). Complex zernike moments features for shape-based image retrieval. *IEEE Transactions on Systems, Man and Cybernetics*, 39(1):227–237.
- Li, Y. J., Ning, P., Wang, X. S., e Jajodia, S. (2003). Discovering calendar-based temporall association rules. *Data and Knowledge Engineering*, 44(2):193–218.
- Li, Z., Liu, X., e Cao, X. (2011). A study on improved eclat data mining algorithm. Advanced Materials Research, 328-330:1896–1899.
- Lin, J. J., Salamon, N., Lee, A. D., Dutton, R. A., Geaga, J. A., e Hayashi, K. M. (2007). Reduced neocortical thickness and complexity mapped in mesial temporal lobe epilepsy with hippocampal sclerosis. *Cerebral Cortex*, 17:2007–2018.
- Lu, C. (2012). Simultaneous nonrigid registration, segmentation, and tumor detection in MRI guided cervical cancer radiation therapy. *IEEE Transaction on Medical Imaging*, 31(6):1213–1227.

- Lustgarten, J. L., Visweswaran, S., Gopalakrishnan, V., e Cooper, G. F. (2011). Application of an efficient Bayesian discretization method to biomedical data. *BMC Bioinformatics*, 12(309):1–15.
- Martins, A. R., Carvalho, V. d. O., e Rezende, S. O. (2007). AprioriGen: Uma extensão do algoritmo Apriori para obtenção de regras de associação generalizadas. Technical report, Universidade de São Paulo.
- Mazid, M. M., Ali, A. B. M. S., e Tickle, K. S. (2010). Improved C4.5 algorithm for rule based classification. *Recent Advances in Artificial Intelligence, Knowledge Engineering* and Data Bases, pages 296–301.
- McDonald, C. R., Hagler Jr., D. J., Ahmadi, M. E., Tecoma, E., Iragui, V., Gharapetian, L., Dale, A. M., e Halgren, E. (2008). Regional neocortical thinning in mesial temporal lobe epilepsy. *Epilepsia*, 49(5):794–803.
- Medicina (2007). Conselho Federal de Medicina. Resolução nº 1.821/2007. Disponível em http://www.conarq.arquivonacional.gov.br./cgi/cgilua.exe/sys/start.htm. Acesso em 11 de outubro de 2012.
- Milgram, S. (1967). The small world problem. Psychology Today, 1:61–67.
- Min, H. e Shuangyuan, Y. (2010). Overview of image mining research. In 5th International Conference on Computer Science & Education, pages 1868–1870, Hefei, China.
- Mitchell, T. M. (1997). Machine Learning. McGraw-Hill Book Company.
- Mohanty, A. K., Senapati, M. R., e Lenka, S. K. (2012). A novel image mining technique for classification of mammograms using hybrid feature selection. *Neural Computing and Applications*, pages 1–11.
- Mohanty, A. K., Swain, S. K., Champati, P. K., e Lenka, S. K. (2011). Image mining for mammogram classification by association rule using statistical and GLCMC features. *International Journal of Computer Science Issues*, 8(5):309–318.
- Mori, S., Kobayashi, Y., e Kuno, Y. (2012). Understanding the meaning of shape description for interactive object recognition, volume 6838 of Lecture Notes in Computer Science, pages 350–356. Springer Berlin Heidelberg.
- Muralidhar, G. S., Bovik, A. C., Sampat, M. P., Whitman, G. J., Haygood, T. M., Stephens, T. W., e Markey, M. K. (2011). Computer-aided diagnosis in breast magnetic resonance imaging. *Mount Sinai Journal of Medicine*, 78:280–290.
- Naganathan, E. R., Narayanan, S., e Kumar, K. R. (2011). FP-growth based new normalization technique for subgraph ranking. *International Journal of Database Management Systems (IJDMS)*, 3(1):81–91.
- Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45(2):167–256.
- Newman, M. E. J. (2010). Networks: An Introduction. Oxford University Press.

- Nishikawa, R. M., Schmidt, R. A., Linver, M. N., Edwards, A. V., Papaioannou, J., e Stull, M. A. (2012). Clinically missed cancer: how effectively can radiologists use computer-aided detection? *American Journal of Roentgenology*, 198(3):708–716.
- Nixon, M. S. e Aguado, A. S. (2008). *Feature Extraction and Image Processing*. Elsevier, second edition.
- Otsu, N. (1979). A thresholding selection method from gray-level histogram. *IEEE Transaction on Systems, Man and Cybernetics*, 9:62–66.
- Pan, H. (2012). Incorporating Domain Knowledge into Medical Image Mining, chapter 8, pages 179–204.
- Pan, H., Tan, X., Han, Q., Feng, X., e Yin, G. (2012). GMA: An approach for association rules mining on medical images, volume 7390 of Lecture Notes in Computer Science, pages 425–432. Springer Berlin Heidelberg.
- Park, J. S., Chen, M. S., e Yu, P. S. (1997). Using a hash-based method with transaction trimming for mining association rules. *IEEE Transactions on Knowledge and Data Engineering*, 9(5):813–825.
- Pasquier, D., Peyrodie, L., Denis, F., Pointreau, Y., Béra, G., e Lartigau, . (2010). Automatic image segmentation for treatment planning in radiotherapy. *Cancer Radiothérapie*, 14(1):S6–S13.
- Pattanaik, S. e Bhalke, D. G. (2012). Efficient content based image retrieval system using mpeg-7 features. *International Journal of Computer Applications*, 53(5):19–24.
- Phyu, T. N. (2009). Survey of classification techniques in data mining. In International MultiConference of Engineers and Computer Scientists, volume 1, pages 727–731, Hong Kong.
- Picon, F. A. (2012). Gêmeos monozigóticos discordantes para transtorno de identidade de gênero: Um estudo da espessura cortical e de morfometria baseada em voxels através de imagens de ressonância magnética estrutural. DissertaÃÃo de mestrado.
- Pires, G. M., Poel, J. v. d., Batista, L. V., e Marques, J. R. T. (2006). Arquitetura para um sistema de diagnóstico auxiliado por computador. In X Congresso Brasileiro de Informática em Saúde - CBIS'2006.
- Prastawa, M., Bullitt, E., e Gerig, G. (2009). Simulation of brain tumors in MR images for evaluation of segmentation efficacy. *Medical Image Analysis*, 13(2):297–311.
- Prati, R. C., Batista, G. E. d. A. P. A., e Monard, M. C. (2008). Curvas ROC para avaliação de classificadores. *IEEE Latin America Transactions*, 6(2):215–222.
- Pu, J., Paik, D. S., Meng, X., Roos, J. E., e Rubin, G. D. (2011). Shape "break-and-repair" strategy and its application to automated medical image segmentation. *IEEE Transac*tion on Visualization and Computer Graphics, 17(1):115–124.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, pages 81–106.
- Quinlan, J. R. (1993). C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, CA.

- Rajendran, P. e Madheswaran, M. (2012). An improved brain image classification technique with mining and shape prior segmentation procedure. *Journal of Medical Systems*, 36(2):747–764.
- Rapoport, A. (1951). Nets with distance bias. Bulletin of Mathematical Biophysics, 13:85–91.
- Rapoport, A. (1953). Spread of information through a population with sociostructural bias: I. assumption of transitivity. *Bulletin of Mathematical Biophysics*, 15:523–533.
- Rebelo, M. S., Furuie, S. S., Gutierrez, M. A., Costa, E. T., e Moura, L. A. (2007). Multiscale representation for automatic identification of structures in medical images. *Computers in Biology and Medicine*, 37:1183–1193.
- Ribeiro, M. X. (2008). Suporte a sistemas de auxílio ao diagnóstico e de recuperação de imagens por conteúdo usando mineração de regras de associação. Tese de doutorado.
- Ribeiro, M. X., Balan, A. G. R., Felipe, J. C., Traina, A. J. M., e Traina Jr., C. (2009a). Mining statistical association rules to select the most relevant medical image features, volume 165 of Studies in Computational Intelligence, pages 113–131. Spring Berlin Heidelberg.
- Ribeiro, M. X., Bugatti, P. H., Traina Jr., C., Marques, P. M. d. A., Rosa, N. A., e Traina, A. J. M. (2009b). Supporting content-based image retrieval and computer-aided diagnosis systems with association rule-based techniques. *Data and Knowledge Engineering*, 68:1370–1382.
- Ribeiro, M. X., Traina, A. J. M., e Traina Jr., C. (2008a). A new algorithm for data discretization and feature selection. In ACM Symposium on Applied Computing (SAC'2008), pages 953–954.
- Ribeiro, M. X., Traina, A. J. M., Traina Jr., C., Rosa, N. A., e Marques, P. M. d. A. (2008b). How to improve medical image diagnosis through association rules: the IDEA method. In 21st IEEE International Symposium on Computer-Based Medical Systems (CBMS'2008), pages 266–271.
- Rish, I. (2001). An empirical study of the naive bayes classifier. In *IJCAI Workshop on Empirical Methods in Artificial Intelligence*, pages 1–6.
- Rodda, S. e Mogalla, S. (2011). A generator based associative classifier for imbalanced datasets. *International Journal of Engineering Science and Technology (IJEST)*, 3(4):2685–2692.
- Rosa, N. A., Felipe, J. C., Traina, A. J. M., Traina Jr., C., Rangayyan, R. M., e Marques, P. M. d. A. (2008). Using relevance feedback to reduce the semantic gap in content-based image retrieval of mammographic masses. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 406–409.
- Roy, S., Saha, A., e Bandyopadhyay, S. K. (2011). Brain tumor segmentation and quantification from MRI of brain. Journal of Global Research in Computer Science (JGRCS), 2(4):155–160.

- Santos, O. M. F. S. (2010). Análise da espessura cortical no cortex visual de pacientes com a doenÃa neuro-degenerativa de Leber. DissertaÃÃo de mestrado.
- Saremi, S., Sejnowski, T. J., e Sharpee, T. (2013). Double Gabor filters are independent components of small translation-invariant image patches. *Neural Computation*, pages 1–18.
- Schwartz, T. H. (2005). Insular seizures: have we been missing the boat? *Epilepsy* Currents, 5(4):147–148.
- Shah, S. e Parikh, S. (2012). Issues in medical diagnosis using computational techniques. In *IEEE Fourth International Conference on Computational Intelligence and Communication Networks (CICN)*, pages 348–354.
- Shahbahrami, A., Pham, T. A., e Bertels, K. (2012). Parallel implementation of gray level co-occurrence matrices and Haralick texture features on cell architecture. *Journal* of Supercomputing, 59:1455–1477.
- Shiraishi, J., Li, Q., Appelbaum, D., e Doi, K. (2011). Computer-aided diagnosis and artificial intelligence in clinical imaging. *Seminars in Nuclear Medicine*, 41(6):449–462.
- Shooshtari, M. A., Maghooli, K., e Badie, K. (2012). Associative classification of mammograms based on parallel mining of image blocks. *Biomedical Engineering: Applications, Basis and Communications*, 24(6):513–524.
- Silva, C. W., Bugatti, P. H., Ribeiro, M. X., Traina Junior, C., e Traina, A. J. M. (2008). Improving cbir using feature extraction based on wavelet transform. In XIV Simpósio Brasileiro de Sistemas Multimedia e Web (WebMedia 2008), volume 1, pages 51–58.
- Silva, J. E. d., Sá, J. P. M., e Jossinet, J. (2000). Classification of breast tissue by electrical impedance spectroscopy. *Medical and Biological Engineering and Computing*, 38:26–30.
- Singh, S. M. e Hemachandran, K. (2012). Content-based image retrieval using color moment and gabor texture feature. *International Journal of Computer Science Issues*, 9(5):299–309.
- Society, A. C. (2012). Cancer Facts & Figures 2012. American Cancer Society, Atlanta.
- Sousa, E. P. M. d. (2006). *Identificação de correlações usando a teoria dos fractais*. Tese de doutorado.
- Souza, A. F. d. (2002). Esqueletos 8-isotrópicos. Dissertação de mestrado.
- Stollnitz, E. J., DeRose, T. D., e Salesin, D. H. (1996). Wavelets for Computer Graphics - Theory and Applications. Morgan Kaufmann Publishers, Inc, San Francisco, CA.
- Sujitha, V., Sivagami, P., e Vijaya, M. S. (2010). Support vector machine based epilepsy prediction using textural features of MRI. *Proceedia Computer Science*, 2:283–290.
- Tan, F. (2007). *Improving feature selection techniques for machine learning*. Tese de doutorado.
- Tan, P., Steinbach, M., e Kumar, V. (2006). Introduction to Data Mining. Addison Wesley.

- Tang, J. e Liu, X. (2011). Classification of breast mass in mammography with an improved level set segmentation by combining morphological features and texture features, volume II, chapter 5, pages 119–136. Springer.
- Thabtah, F. (2007). A review of associative classification mining. *The Knowledge* Engineering Review, 22(1):37–65.
- Tillander, A. (2012). Effect of data discretization on the classification accuracy in a high-dimensional framework. *International Journal of Intelligent Systems*, 27(4):355–374.
- Tiwari, A., Gupta, R. K., e Agrawal, D. P. (2010). A survey on frequent pattern mining: current status and challenging issues. *Information Technology Journal*, 9(7):1278–1293.
- Traina, A. J. M., Traina Jr., C., Balan, A. G. R., Ribeiro, M. X., Bugatti, P. H., Watanabe, C. Y. V., e Azevedo-Marques, P. M. (2011). *Feature extraction and selection for decision* making, chapter 8, pages 197–223. Springer-Verlag, Berlin.
- Traina, A. J. M., Traina Jr., C., Bueno, J. M., Chino, F. J. T., e Marques, P. M. d. A. (2003). Efficient content-based image retrieval through metric histograms. World Wide Web Journal (WWWJ), 6(2):157–185.
- Watanabe, C. Y. V., Diniz, P. R. B., Velasco, T. R., Hildenberg, C. L. O., e Traina, A. J. M. (2012a). Automatically detecting temporal lobe epilepsy using cortical thickness measures. In XXIII Congresso Brasileiro de Engenharia BiomÃdica, pages 1255–1259.
- Watanabe, C. Y. V., Ramos, J. d. S., Traina, A. J. M., e Traina Jr., C. (2012b). A CAD system based on complex networks theory to characterize mass in mammograms. In SPIE Medical Imaging, volume 8315, pages 1–12.
- Watanabe, C. Y. V., Ribeiro, M. X., Traina, A. J. M., e Traina Jr., C. (2012c). A statistical associative classifier with automatic estimation of parameters on computer aided diagnosis. In *IEEE 11th International Conference on Machine Learning and Applications (ICMLA)*, volume 1, pages 564–567, Boca Raton, FL.
- Watanabe, C. Y. V., Ribeiro, M. X., Traina Jr., C., e Traina, A. J. M. (2010). Statistical Associative Classification of Mammograms - The SACMiner Method. Proceedings of the 12th International Conference on Enterprise Information Systems, 2:121–128.
- Watanabe, C. Y. V., Ribeiro, M. X., Traina Jr., C., e Traina, A. J. M. (2011). SACMiner: a new classification method based on statistical association rules to mine medical images, volume 73 of Lecture Notes in Business Information Processing, pages 249–263. Springer.
- Watts, D. J. (1999). Small worlds: the dynamics of networks between order and randomness. Princeton University Press.
- Watts, D. J. e Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393:440–442.
- Webb, S. (2012). Webb's Physics of Medical Imaging. Series in Medical Physics and Biomedical Engineering. Taylor & Francis, second edition.
- Wei, C.-H. e Li, C. T. (2008). Content-Based Retrieval for Mammograms. Proc. IEEE Image Retrieval, Beijing, China.
- Weinland, D., Ronfard, R., e Boyer, E. (2011). A survey of vision-based methods for action representation, segmentation and recognition. *Computer Vision and Image* Understanding, 115:224–241.
- WO 4th, T. (2012). Mesial temporal lobe epilepsy. *Journal of Clinical Neurophysiology*, 29(5):356–365.
- Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z.-H., Steinbach, M., Hand, D. J., e Steinberg, D. (2007). Top 10 algorithms in data mining. *Knowledge and Information Systems (KAIS)*, 14(1).
- Wyatt, J. C. e Altman, D. G. (1995). Prognostic models: clinically useful or quickly forgotten? *BMJ*, 311(1539).
- Xu, J., Faruque, J., Beaulieu, C. F., Rubin, D., e Napel, S. (2012). A comprehensive descriptor of shape: method and application to content-based retrieval of similar appearing lesions in medical images. *Journal of Digital Imagin*, 25(1):121–128.
- Xu, Y. (2012). Multiple clustered instance learning for hitopathology cancer image classification, segmentation and clustering. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 964–971.
- Yamamoto, C. H., Oliveira, M. C. F., Rezende, S. O., e Nomelini, J. (2008). Including the user in the knowledge discovery loop: interactive intemset-driven rule extraction. In 23rd ACM Symposium on Applied Computing (SAC) - Multimedia and Visualization Track, volume 2, pages 1212–1217.
- Yang, Z. L. e Guo, B. L. (2008). Image registration using feature points extraction and pseudo-Zernike moments. In International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IIHMSP '08, pages 752–755.
- Yao, Y. (2012). A novel data mining algorithm based on rough set. Advances in Intelligent and Soft Computing, 114:1115–1121.
- Ye, X., Lin, X., Dehmeshki, J., Slabaugh, G., e Beddoe, G. (2009). Shape based computer-aided detection of lung nodules in thoracic CT images. *IEEE Transaction on Biomedical Engineering*, 56(7):1810–1820.
- Zahn, C. e Roskies, R. Z. (1972). Fourier descriptors for plane closed curves. *IEEE Transactions on Computers*, 21:269–281.
- Zaki, M. J. e Hsiao, C. J. (2000). Scalable algorithms for association mining. *IEEE Transaction on Knowledge and Data Engineering*, 12:372–390.
- Zhang, J., Hsu, W., e Lee, M. L. (2001). Image mininig: issues, frameworks and techniques. In Second International Workshop on Multimedia Data Mining (MDM/KDDâ2001),, pages 13–20.

- Zhang, S. J. e Zhou, Q. (2012). A novel efficient classification algorithm based on class association rules. *Applied Mechanics and Materials*, 135-136:106–110.
- Zhang, T. Y. e Suen, C. Y. (1984). A fast parallel algorithm for thinning digital patterns. Communications of the ACM, 27(3):236–239.
- Zhang, Y. (2012). Background extraction algorithm based on k-means clustering algorithm and histogram analysis.
- Zhu, X. e Deng, H. (2011). A brief review on frequent pattern mining. In *IEEE 3rd* International Workshop on Intelligent Systems and Applications (ISA), pages 1–4.