

4

Avaliação experimental

O presente capítulo apresenta os experimentos realizados com o propósito de avaliar o método de detecção de nódulos pulmonares proposto neste trabalho, principalmente, a segmentação multicritério. Para tanto, inicialmente, são apresentadas as características particulares dos dados utilizados. Em seguida, são apresentados e discutidos os resultados obtidos nos experimentos. Os resultados estão divididos em duas partes, uma sem e a outra com seleção automática do conjunto de critérios para segmentação.

4.1.

Base de dados e programa de implementação

Para avaliar a metodologia para detecção automática de nódulos pulmonares em imagens de tomografia computadorizada proposta, um protótipo foi desenvolvido em linguagem Matlab, o qual foi testado em duas bases de dados apresentadas a seguir. Além do pacote padrão de funções Matlab, foi utilizado o pacote de funções PRTools (Duin et al., 2007) com o qual foi efetuada na fase de seleção de atributos.

4.1.1.

Base de dados HUPE

As imagens desta base de dados foram fornecidas pelo Hospital Universitário Pedro Ernesto (HUPE), no Rio de Janeiro. As imagens foram adquiridas através de um tomógrafo Helical GE Pro Speed sob as seguintes condições: 120 KVp de voltagem e 100 mA de corrente do tubo.

Os exames contêm entre 11 e 98 fatias de 512 x 512 pixels. A dimensão média dos voxels é de 0,67 x 0,67 x 2 mm e as imagens foram quantizadas em 12 bits e armazenadas em formato DICOM.

É importante mencionar que os exames foram feitos sem a injeção de contraste, o que facilitaria a detecção, porém poderia incidir em doença e, ocasionalmente, mortalidade por efeitos alérgicos.

Além das imagens, são utilizadas como referência, as respectivas segmentações dos nódulos, produzidas por um especialista com auxílio de um sistema de análise de nódulos pulmonares chamado Bebúi (Silva et al., 2008). Este programa usa uma técnica de segmentação semi-automática. Para cada nódulo, o especialista precisa colocar uma semente sobre o nódulo, cercar a área do nódulo com um círculo, determinar os limiares inferior e superior de densidade e indicar o grupo de fatias que contém o nódulo. Ao final, se necessário, o especialista ainda edita o resultado da segmentação.

No total, esta base é composta por 33 exames contendo 38 nódulos (benignos e malignos).

4.1.2. Base de dados LIDC

Existem pesquisas e publicações em detecção e diagnóstico auxiliado por computador de nódulos pulmonares há mais de uma década. A maioria dos autores cria sua própria base de dados e validação o que torna difícil a comparação de performance entre diferentes sistemas e metodologias, já que o desempenho de um sistema depende dos dados nos quais este é testado.

Recentemente, uma base de dados pública, conhecida como LIDC (*Lung Image Database Consortium*), vem sendo criada por cinco instituições acadêmicas dos Estados Unidos da América para auxiliar e estimular pesquisadores nesta área (Armato et al. 2004). Atualmente, o LIDC é composto por 85 exames de tomografia computadorizada, armazenadas em formato DICOM, e suas respectivas anotações em arquivos com formato XML.

Cada exame foi analisado por quatro radiologistas em duas fases. Na primeira fase, cada radiologista analisa os exames independentemente dos demais, ou seja, sem troca de informações, e registra suas anotações indicando os nódulos encontrados. Em seguida, na segunda fase, as anotações da primeira fase são distribuídas entre todos os radiologistas, os quais analisam novamente os exames.

Nesta fase, cada radiologista refez livremente suas anotações, decidindo se incluía ou ignorava os nódulos indicados pelos demais radiologistas (podendo até mesmo excluir nódulos que ele próprio indicou como nódulo na primeira fase).

Os radiologistas não são forçados a concordarem com os demais, nem mesmo a entrarem em um consenso. Sendo assim, cada exame contém quatro análises diferentes e independentes, os quais podem ser divergentes entre si. Neste caso, uma mesma região pode ser classificada, assim como delimitada, diferentemente pelos quatro radiologistas.

Para nódulos maiores que três milímetros, os radiologistas delineararam seus contornos e indicaram algumas características como esfericidade, textura, calcificação, diagnóstico, etc (a estas características são atribuídas um valor entre 1 e 5). Enquanto que, para nódulos menores que três milímetros, os radiologistas simplesmente marcaram um ponto sobre o nódulo (aproximadamente, o centróide).

Com a finalidade de reduzir o impacto da subjetividade sobre a análise, dentre os exames disponíveis, foram utilizados na avaliação experimental apenas aqueles exames que continham algum objeto indicado unanimemente como nódulo. Além disso, objetos com diferentes classificações (nódulo/não-nódulo) foram ignorados pelos experimentos do presente trabalho na fase de classificação. Desta forma, foram utilizados 43 exames contendo 77 nódulos (unanimemente assim classificados). Nestes exames, 305 estruturas foram desprezadas, já que a respectiva classificação divergia entre os especialistas. Os exames possuem entre 101 e 525 fatias de 512 x 512 pixels. O tamanho médio dos voxels é de 0,65 x 0,65 x 2 mm e as imagens estão quantizadas em 16 bits.

4.2.

Divisão da base de dados para treinamento e teste

Existem três fases nas quais é necessária parte do banco de dados para treinamento: seleção de variáveis, ajuste de parâmetros de segmentação e treinamento do classificador. Por este motivo, o banco de dados foi dividido como se segue.

4.2.1. Base de dados HUPE

Desta base foi selecionado aleatoriamente um conjunto de cinco exames para o ajuste de parâmetros de segmentação.

Na fase de detecção de nódulos pulmonares, foi utilizada a técnica “deixe um de fora” (*leave-one-out*) para o treinamento do classificador. Portanto, para cada exame a ser testado, os demais foram usados para treinar o SVM. Neste caso, os exames selecionados para ajuste dos parâmetros de segmentação também foram usados no conjunto de treinamento do classificador, porém foram excluídos do conjunto de teste.

Sendo assim, do total de 33 exames, 5 foram utilizados para o ajuste de parâmetros de segmentação e 28 exames foram utilizados para testar o método na fase de detecção através de um classificador treinado por 32 exames (mudados a cada exame de teste; ou seja, para cada um dos 28 exames do conjunto de teste, os outros 27 foram utilizados para treinar o classificador juntamente com os 5 exames utilizados para o ajuste de parâmetros).

4.2.2. Base de dados LIDC

Na fase de ajuste dos parâmetros de segmentação, é indispensável o uso de objetos previamente delimitados por especialistas os quais indicam a segmentação final desejada. Devido a esta necessidade, não é possível utilizar qualquer um dos exames LIDC para tal procedimento, visto que muitos dos nódulos indicados pelos radiologistas foram simplesmente marcados com um ponto (não delimitados). Levando isto em consideração, apenas 23 exames (do total de 43) poderiam ser incluídos neste procedimento.

A fim de não reduzir drasticamente o número de exames para a fase de detecção de nódulos, apenas 10 exames foram separados para o ajuste de parâmetros. Estes exames formaram, então, três grupos:

- Grupo 1: composto por 5 exames;
- Grupo 2: composto pelos 5 exames não incluídos no Grupo 1;
- Grupo 3: composto por todos os 10 exames.

Desta forma, a busca de parâmetros para a segmentação foi realizada em com base nos três grupos para melhor avaliar a robustez e eficiência do método.

A seleção de atributos tanto para segmentação quanto para classificação foi feita a partir do terceiro grupo, ou seja, a partir dos 10 exames separados para o ajuste de parâmetros de segmentação.

Assim como com o banco de dados HUPE, para avaliar a detecção de nódulos, os testes foram realizados em 33 exames (do total de 43) utilizando a técnica de deixar um de fora. Neste caso, os 10 exames utilizados no ajuste de parâmetros, apesar de não participarem do conjunto de teste, foram também utilizados para treinar o classificador.

4.3.

Experimentos com o primeiro grupo de critérios para a segmentação

Nesta seção, é apresentada uma implementação particular da segmentação multicritério.

Neste caso, os critérios utilizados para a segmentação multicritério foram adaptados dos critérios utilizados na segmentação multiresolução para imagens bidimensionais proposta por Baatz et al. 2000a, ou seja, espectral, compacidade e suavidade. O critério espectral é dado pelo desvio padrão das respostas espectrais de cada banda da imagem; a compacidade é dada pela razão entre o perímetro do objeto e a raiz quadrada de sua área; enquanto a suavidade é a razão entre o perímetro do objeto e a área da menor caixa (*bounding box*) que contém o objeto. A seguir, são descritos os critérios a serem utilizados neste grupo de experimentos.

4.3.1.

Critérios selecionados para segmentação de nódulos

É conhecido que, em geral, nódulos pulmonares assumem formas similares a esferas (Way et al, 2006; Ge et al., 2005). Desta forma, além de um critério a partir da densidade dos voxels, são utilizados nos experimentos desta seção para compor a função de heterogeneidade dois atributos de forma: *compacidade* e

desproporção esférica.

A heterogeneidade associada à densidade é medida pelo desvio padrão dos valores dos voxels (σ) pertencentes ao objeto cujo valor é encontrado por:

$$\sigma = \sqrt{\sigma^2} \quad (12)$$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (d_i - \mu)^2 \quad (13)$$

onde σ^2 é a variância dos valores de densidade, d_i é o valor de densidade do i -ésimo voxel, n é o número de voxels pertencentes ao objeto e μ é a média dos valores de densidade.

A compacidade (C) é definida pela eq. (14):

$$C = \frac{\sqrt{a}}{\sqrt[3]{v}} \quad (14)$$

onde a e v são, respectivamente, área superficial e o volume do objeto.

Nota-se que a compacidade mede a ocupação espacial do objeto e é mínima para objetos esféricos.

O segundo critério de forma utilizado, a desproporção esférica (D), foi introduzido e utilizado por Sousa et al. (2007a) na fase de identificação de nódulos pulmonares, tendo-se mostrado promissor nesse sentido. No presente trabalho, a desproporção esférica é utilizada desde a fase de segmentação. Este atributo é calculado pela equação abaixo:

$$D = \frac{a}{4 \cdot \pi \cdot R^2} \quad (15)$$

onde R é o raio da esfera com o mesmo volume do objeto original, isto é,

$$R = \sqrt[3]{\frac{3 \cdot v}{4 \cdot \pi}} \quad (16)$$

A desproporção esférica assume valores pequenos para objetos aproximadamente esféricos e com superfície regular e valores altos para objetos irregulares e alongados; portanto, mede a suavidade da superfície de um objeto.

Ambas as medidas de forma, desproporção esférica e compacidade, são invariantes com relação à rotação, à translação e à escala.

Aplicando estes atributos à equação de heterogeneidade (1), tem-se:

$$h_1 = n_{Obj12} \cdot \sigma_{Obj12} - (n_{Obj1} \cdot \sigma_{Obj1} + n_{Obj2} \cdot \sigma_{Obj2}) \quad (17)$$

$$h_2 = n_{Obj12} \cdot C_{Obj12} - (n_{Obj1} \cdot C_{Obj1} + n_{Obj2} \cdot C_{Obj2}) \quad (18)$$

$$h_3 = n_{Obj12} \cdot D_{Obj12} - (n_{Obj1} \cdot D_{Obj1} + n_{Obj2} \cdot D_{Obj2}) \quad (19)$$

onde h_1 , h_2 e h_3 são heterogeneidades para desvio padrão (das densidades) σ , compacidade C e desproporção esférica D , respectivamente.

Desta forma, a equação de aumento de heterogeneidade global (3) pode ser reformulada da seguinte maneira:

$$H = w_1 h_1 + w_2 h_2 + (1 - w_1 - w_2) h_3 \quad (20)$$

Por conseguinte, o algoritmo de segmentação multicritério, neste caso, depende de três parâmetros: escala (s) e dois pesos (w_1 e w_2).

4.3.2.

Ajuste automático dos valores dos parâmetros de segmentação

Geralmente, nódulos pulmonares são semelhantes entre si, mas ainda apresentam uma variação considerável em densidade e forma de paciente para paciente. Para capturar esta variação, um determinado número de exames com seus respectivos nódulos delimitados manualmente por especialistas foi usado para ajustar os parâmetros de segmentação seguindo o procedimento descrito na Seção 3.3.3, pois representam o resultado esperado da segmentação. Como descrito na Seção 4.2, os parâmetros de segmentação foram ajustados a partir de 4 grupos de imagens, sendo 1 do HUPE e 3 do LIDC.

Nestes experimentos, o GA foi executado com uma população de 50

indivíduos, 100 gerações, sendo que a cada mudança de geração, 70% dos indivíduos foram trocados.

Devido aos algoritmos genéticos serem estocásticos, os valores encontrados em execuções diferentes podem diferir entre si. Portanto, para cada experimento, o método foi executado mais de uma vez para avaliar sua robustez, neste caso, cinco vezes.

A seguir, analisam-se a convergência e a estabilidade do algoritmo genético para a busca de parâmetros da segmentação multicritério apresentada.

4.3.2.1. Convergência do algoritmo genético

Em cada execução de algoritmos genéticos, a população que constitui a primeira geração é usualmente gerada aleatoriamente, pois, na maioria dos casos, não se conhece o comportamento dos valores a serem encontrados. Por isto, a população inicial é composta por indivíduos bastante diversificados. Nas gerações seguintes, a população é gerada, basicamente, a partir da reprodução dos indivíduos da população anterior privilegiando os julgados mais aptos. Sendo assim, a população tende a evoluir para uma solução singular. Quando a população consiste basicamente de indivíduos similares, é dito que a solução convergiu.

A performance de algoritmos genéticos é frequentemente analisada a partir de um gráfico de desempenho. Este gráfico é gerado da seguinte forma:

- O algoritmo genético é executado $N_{\text{experimentos}}$ vezes;
- Em cada geração é armazenado o valor de avaliação do melhor indivíduo da população;
- Calcula-se a média aritmética dos valores de avaliação de todas as execuções por geração e, então, o gráfico é produzido com base nestas médias.

A Figura 8 contém o gráfico de desempenho médio através de cinco execuções dos experimentos para ajuste dos valores dos parâmetros do modelo utilizando o banco de dados HUPE. O eixo horizontal do gráfico refere-se às gerações e o eixo vertical indica a avaliação média em cada execução do melhor indivíduo da população da respectiva geração. Observa-se, neste gráfico, que a

solução converge em torno de 62 gerações. Assim, neste caso, 100 gerações foram de fato suficientes para a procura da solução desejada.

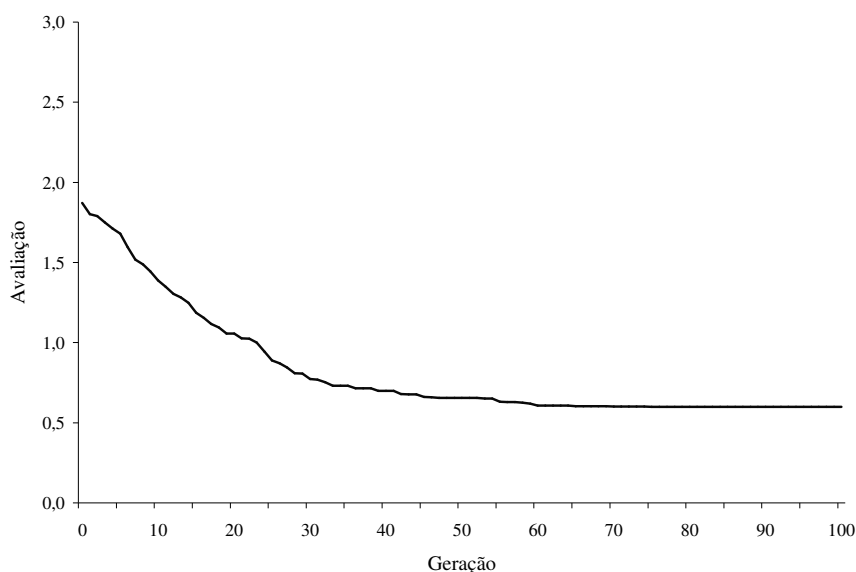


Figura 8 – Convergência de algoritmos genéticos para ajuste de valores dos parâmetros da segmentação com o primeiro conjunto de critérios no banco de dados HUPE.

Este mesmo experimento foi realizado a partir dos três grupos de imagens mencionados na Seção 4.2.2. A Figura 9, a Figura 10 e a Figura 11 apresentam os gráficos de desempenho médio em cinco execuções do GA para ajuste dos valores dos parâmetros do modelo utilizando, respectivamente, o grupo 1, o grupo 2 e o grupo 3 do banco de dados LIDC.

A Figura 12 mostra o gráfico de desempenho médio de todas as execuções feitas com o LIDC, ou seja, uma média entre os experimentos feitos nos grupos 1, 2 e 3 do banco de dados LIDC. Observa-se, neste gráfico, que a solução converge em torno de 72 gerações. Assim, neste caso, cem gerações foram realmente suficientes na busca da solução desejada.

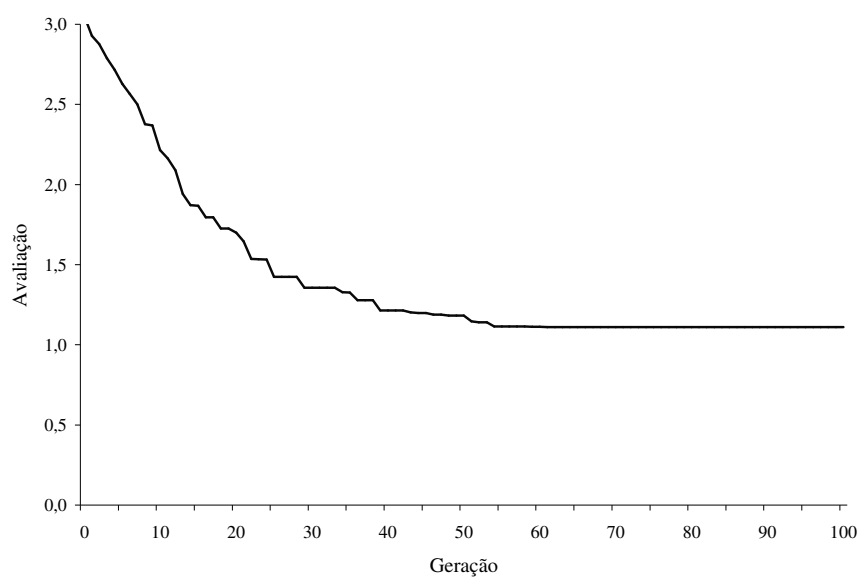


Figura 9 – Convergência de algoritmos genéticos para ajuste de valores dos parâmetros da segmentação com o primeiro conjunto de critérios a partir do primeiro grupo de imagens do banco de dados LIDC.

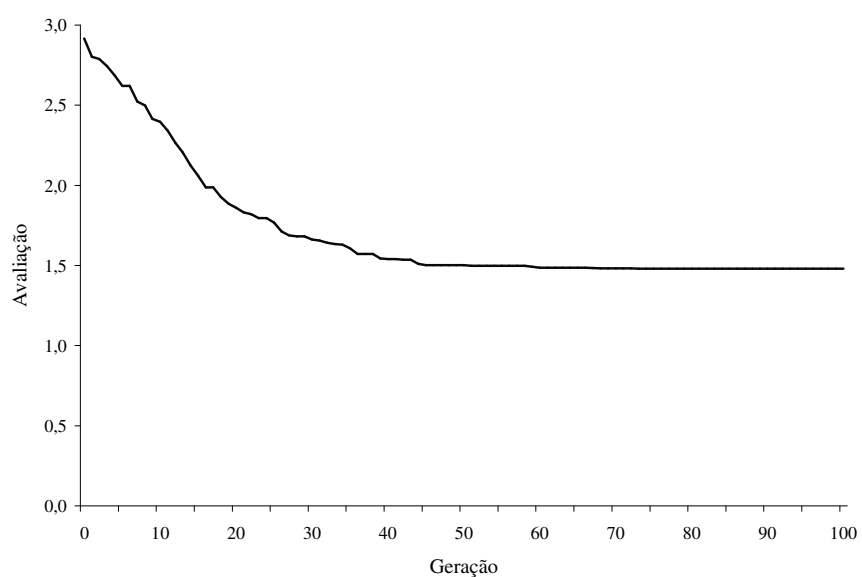


Figura 10 – Convergência de algoritmos genéticos para ajuste de valores dos parâmetros da segmentação com o primeiro conjunto de critérios a partir do segundo grupo de imagens do banco de dados LIDC.

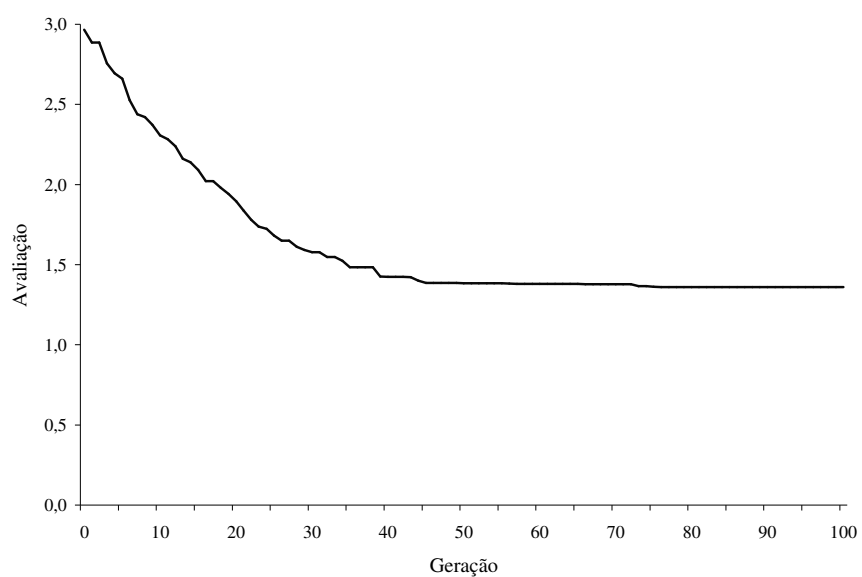


Figura 11 – Convergência de algoritmos genéticos para ajuste de valores dos parâmetros da segmentação com o primeiro conjunto de critérios a partir do terceiro grupo de imagens do banco de dados LIDC.

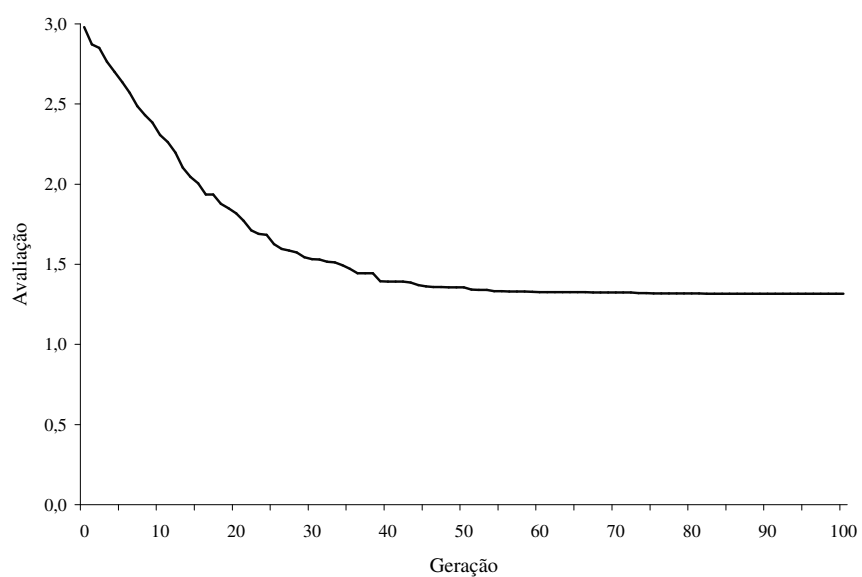


Figura 12 – Convergência de algoritmos genéticos para ajuste de valores dos parâmetros da segmentação com o primeiro conjunto de critérios no banco de dados LIDC.

4.3.2.2.

Análise de estabilidade dos valores dos parâmetros do modelo

Já que o método baseado em algoritmo genético pode apresentar um conjunto de valores de parâmetros diferentes em execuções diferentes, mesmo utilizando os mesmos dados de entrada e mesmos valores para os parâmetros do GA, esta seção tem a finalidade de analisar a estabilidade destes conjuntos.

Os conjuntos de valores dos parâmetros encontrados pelo GA e seus respectivos valores de avaliação em cada execução são apresentados nas próximas tabelas. As colunas correspondem respectivamente ao número da execução, os três parâmetros de ponderação w_1 , w_2 e w_3 (veja a eq. (20)), a escala s e o valor de avaliação do conjunto de parâmetros correspondente.

A Tabela 2 apresenta os resultados obtidos a partir do banco de dados HUPE, enquanto que a Tabela 3, a Tabela 4 e a Tabela 5 contêm os resultados encontrados nos experimentos com os grupos 1, 2 e 3 do banco de dados LIDC.

Note que os valores dos parâmetros encontrados em cada experimento convergiram para uma solução semelhante. Estes resultados sugerem que o procedimento de adaptação proposto é robusto.

Execução	Parâmetros				Avaliação
	w_1	w_2	w_3	s	
1	0,476	0,336	0,188	79	0,59
2	0,474	0,266	0,260	75	0,59
3	0,439	0,311	0,249	76	0,62
4	0,466	0,272	0,262	80	0,63
5	0,417	0,303	0,280	76	0,57
Média	0,454	0,298	0,248	77,2	0,60
Desvio padrão	0,025	0,029	0,035	2,2	0,02

P
M

Tabela 2 – Resultados do procedimento de parametrização da segmentação com o primeiro conjunto de critérios com base no banco de dados HUPE.

Execução	Parâmetros				Avaliação
	w_1	w_2	w_3	s	
1	0,521	0,237	0,242	79	1,08
2	0,556	0,227	0,217	79	1,04
3	0,479	0,174	0,347	80	1,24
4	0,522	0,260	0,218	76	1,07
5	0,510	0,228	0,262	80	1,14
Média	0,518	0,225	0,257	78,8	1,11
Desvio padrão	0,028	0,032	0,054	1,6	0,08

M

P

Tabela 3 – Resultados do procedimento de parametrização da segmentação com o primeiro conjunto de critérios com base no primeiro grupo de imagens do LIDC.

Execução	Parâmetros				Avaliação
	w_1	w_2	w_3	s	
1	0,547	0,197	0,256	75	1,55
2	0,560	0,301	0,139	79	1,45
3	0,545	0,203	0,252	76	1,42
4	0,466	0,224	0,310	79	1,58
5	0,548	0,218	0,234	79	1,41
Média	0,533	0,229	0,238	77,6	1,48
Desvio padrão	0,038	0,042	0,062	1,9	0,08

P

M

Tabela 4 – Resultados do procedimento de parametrização da segmentação com o primeiro conjunto de critérios com base no segundo grupo de imagens do LIDC.

Execução	Parâmetros				Avaliação
	w_1	w_2	w_3	s	
1	0,573	0,211	0,216	77	1,33
2	0,537	0,252	0,211	80	1,19
3	0,441	0,162	0,397	80	1,39
4	0,478	0,220	0,302	76	1,43
5	0,497	0,189	0,314	80	1,44
Média	0,505	0,207	0,288	78,6	1,36
Desvio padrão	0,051	0,034	0,077	1,9	0,10

Tabela 5 – Resultados do procedimento de parametrização da segmentação com o primeiro conjunto de critérios com base no terceiro grupo de imagens do LIDC.

De cada experimento (referente à cada tabela), foram escolhidos dois conjuntos de valores de parâmetros para segmentar as demais imagens da base de dados cujos objetos são classificados como mencionado na Seção 3.5. As soluções escolhidas são aquelas cujas linhas encontram-se sombreadas e indicadas por “P” e “M” que correspondem, respectivamente, à maior e menor avaliação e representam, portanto, o pior (P) e o melhor (M) caso encontrado.

4.3.3. Detecção de nódulos

Após a segmentação das imagens de TC utilizando os parâmetros encontrados conforme apresentado na seção anterior, os objetos são classificados em nódulo ou não-nódulo.

Nestes experimentos, os exames selecionados para ajuste dos parâmetros de segmentação também foram usados no conjunto de treinamento do classificador SVM, porém foram excluídos do conjunto de teste.

Devido ao número reduzido de padrões disponíveis, foi utilizada a técnica “deixe um de fora” (*leave-one-out*) na avaliação. Assim, para cada exame testado, os demais foram usados para treinar o SVM.

A Tabela 6 apresenta resultados informados em outros trabalhos de detecção

de nódulos pulmonares com banco de dados diferentes. Estes resultados são os mesmos relatados nas publicações correspondentes. Nesta tabela, a segunda e terceira colunas apresentam, respectivamente, a sensibilidade (S) e a acurácia (A) apresentadas nos trabalhos indicados na primeira coluna, enquanto que a quarta e a quinta colunas mostram o número de falsos positivos.

Trabalho	S (%)	A (%)	$FP/fatia$	$FP/exame$
Gurcan et al. (2002)	84	–	1,74	–
Lee et al. (2001)	72	–	1,10	–
Suzuki et al. (2003)	80,3	–	0,18	–
Ye et al. (2007)	92,3	–	0,07	12,7
Zhang et al. (2004)	–	83,9	–	3,5

Tabela 6 – Resultados da detecção de nódulos pulmonares em trabalhos encontrados na literatura.

A Tabela 7 transcreve os resultados de trabalhos publicados por outros pesquisadores em detecção de nódulos pulmonares cuja base de dados utilizada é do LIDC. Nota-se pela segunda coluna que apesar de a fonte ser a mesma, o número de exames é distinto entre eles, ou seja, a base de dados utilizada nestes experimentos não é exatamente a mesma.

Trabalho	Banco de dados LIDC	S (%)	$FP/fatia$	$FP/exame$
Opfer & Wiemker 2007	91 exames	60,0	–	2
Özekes 2007	12 pacientes 276 fatias 153 nódulos	93,4	0,594	–
Lee et al. 2008	32 exames 5721 fatias	100,0	1,4	–
Ozekes et al. 2008	16 casos/nódulos 425 fatias	100,0	–	13,4 $FP/caso$
Pu et al. 2008	52 exames	81,5	–	6,5
Dolejsi al. 2009	38 exames 223,6 fatias/exame	89,6	12,03	–

Tabela 7 – Desempenho de trabalhos sobre detecção de nódulos utilizando o LIDC.

Não é possível fazer uma comparação definitiva do desempenho obtido nos experimentos deste trabalho com os apresentados na Tabela 6 e na Tabela 7, pois não foi empregada exatamente a mesma base de dados e, portanto, as condições foram diferentes das aqui indicadas. Porém, estes indicadores ajudam a relacionar o método proposto com a literatura.

A Tabela 8 apresenta o desempenho encontrado nos experimentos realizados na base de dados HUPE o qual também foi utilizado por Sousa (2007b). Entretanto, na etapa de reconhecimento, Sousa incluiu todos os 33 exames, em vez de 28 como neste trabalho, já que os cinco exames usados para ajustar os parâmetros de segmentação não fizeram parte do conjunto de teste. Na segunda coluna, é apresentado o conjunto de exames utilizados para o ajuste de parâmetros da segmentação e a terceira coluna indica pelos símbolos P e M a utilização dos conjuntos de parâmetros no pior e melhor caso, segundo a avaliação, como apresentado na Seção 4.3.2.2. A Figura 13 e a Figura 14 apresentam gráficos comparativos do desempenho dos métodos utilizando esta base de dados.

Segmentação	Ajuste de parâmetros	Caso	S (%)	A (%)	$FP/exame$
Sousa (2007b)	–	–	84,8	95,2	0,42
desvio padrão, compacidade e desproporção esférica	5 exames do HUPE	P	85,5	97,2	0,41
desvio padrão, compacidade e desproporção esférica	5 exames do HUPE	M	86,4	97,8	0,40

Tabela 8 – Resultados da detecção de nódulos pulmonares no banco de dados HUPE.

Mesmo trabalhando com o pior conjunto de valores de parâmetros encontrado na etapa de ajuste, o método proposto obteve nos experimentos com o banco de dados HUPE 85,5% de sensibilidade e 0,41 falsos positivos por exame. Isto indica que o método é capaz de fornecer uma boa sugestão para o especialista, vindo ao encontro do objetivo de um CAD, isto é, fornecer uma segunda opinião sem substituir o especialista. Os resultados utilizando o conjunto de parâmetros do melhor caso não são, de fato, muito diferentes do pior caso o que indica robustez do procedimento de adaptação.

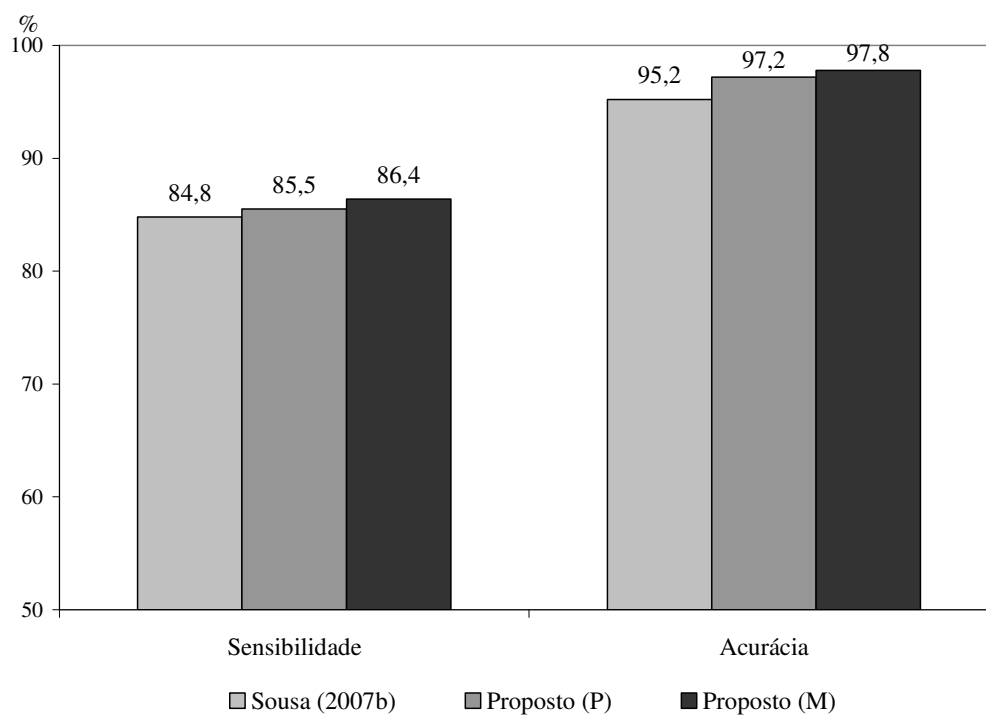


Figura 13 – Desempenho da detecção de nódulos pulmonares na base de dados HUPE.

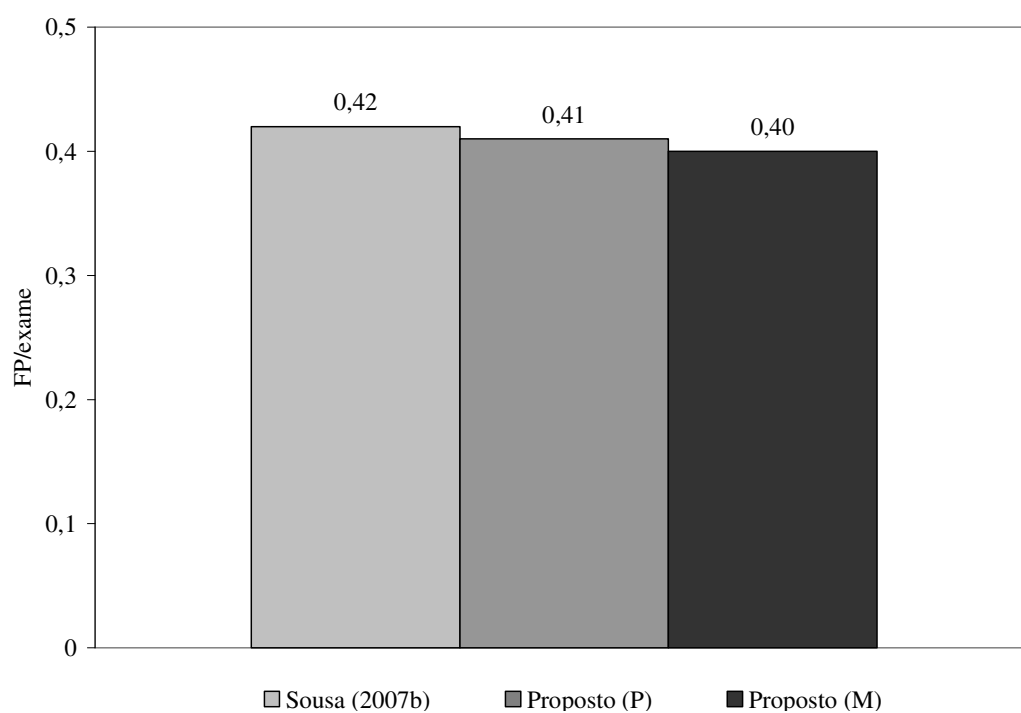


Figura 14 – Falsos positivos por exame da detecção de nódulos pulmonares na base de dados HUPE.

A Tabela 9 apresenta os resultados obtidos pelos experimentos na base de dados LIDC, onde a primeira coluna indica o número de identificação do respectivo experimento. Na segunda coluna, são apresentados os critérios utilizados para a segmentação, sendo que no experimento 1 as imagens foram segmentadas por uma simples limiarização de Otsu sobre os valores dos voxels e, nos demais experimentos, foi empregada a segmentação aqui proposta utilizando os critérios citados. Nota-se que nos experimentos 2, 3 e 4 foram utilizados apenas um critério na segmentação, nos experimentos de 5 a 8 foram empregados dois critérios na segmentação, enquanto que nos experimentos de 9 a 14 foram empregados os três critérios selecionados para este primeiro conjunto de atributos para segmentação. A terceira coluna apresenta qual o grupo de exames do LIDC foi utilizado na fase de ajuste de parâmetros. Na quarta coluna, é indicado qual caso (conjunto de valores de parâmetros) foi utilizado na segmentação, podendo ser pior caso (P), melhor caso (M) ou caso único (U); este último refere-se aos experimentos nos quais o ajuste de parâmetros foi executado uma única vez. As três últimas colunas, finalmente, apresentam os resultados obtidos em tais experimentos. A Figura 15 e a Figura 16 mostram graficamente uma comparação destes resultados, os números dos experimentos são os mesmos indicados na primeira coluna da Tabela 9.

É notável nestes resultados a importância da segmentação dentro do processo de reconhecimento. Entre os experimentos com segmentação monocritério (apenas um critério), aqueles cujo critério refere-se à forma do objeto (compacidade e desproporção esférica) tiveram resultados insatisfatórios. Isto mostra que, para segmentação, os critérios de forma, isoladamente, não são adequados.

Exp	Segmentação	Ajuste de parâmetros	Caso	<i>S (%)</i>	<i>FP/fatia</i>	<i>FP/exame</i>
1	Otsu	–	–	68,4	0,41	62,3
2	desvio padrão	Grupo 3	U	68,2	0,41	62,2
3	compacidade	Grupo 3	U	52,8	0,59	90,6
4	desproporção esférica	Grupo 3	U	53,6	0,62	94,6
5	desvio padrão e compacidade	Grupo 3	P	72,5	0,41	62,1
6	desvio padrão e compacidade	Grupo 3	M	72,7	0,36	55,0
7	desvio padrão e desproporção esférica	Grupo 3	P	74,1	0,35	53,3
8	desvio padrão e desproporção esférica	Grupo 3	M	74,5	0,30	46,2
9	desvio padrão, compacidade e desproporção esférica	Grupo 1	P	76,5	0,30	45,1
10	desvio padrão, compacidade e desproporção esférica	Grupo 1	M	76,5	0,29	44,7
11	desvio padrão, compacidade e desproporção esférica	Grupo 2	P	76,2	0,29	43,4
12	desvio padrão, compacidade e desproporção esférica	Grupo 2	M	76,2	0,28	42,9
13	desvio padrão, compacidade e desproporção esférica	Grupo 3	P	79,6	0,28	42,5
14	desvio padrão, compacidade e desproporção esférica	Grupo 3	M	79,8	0,28	42,2

Tabela 9 – Resultados da detecção de nódulos pulmonares no banco de dados LIDC.

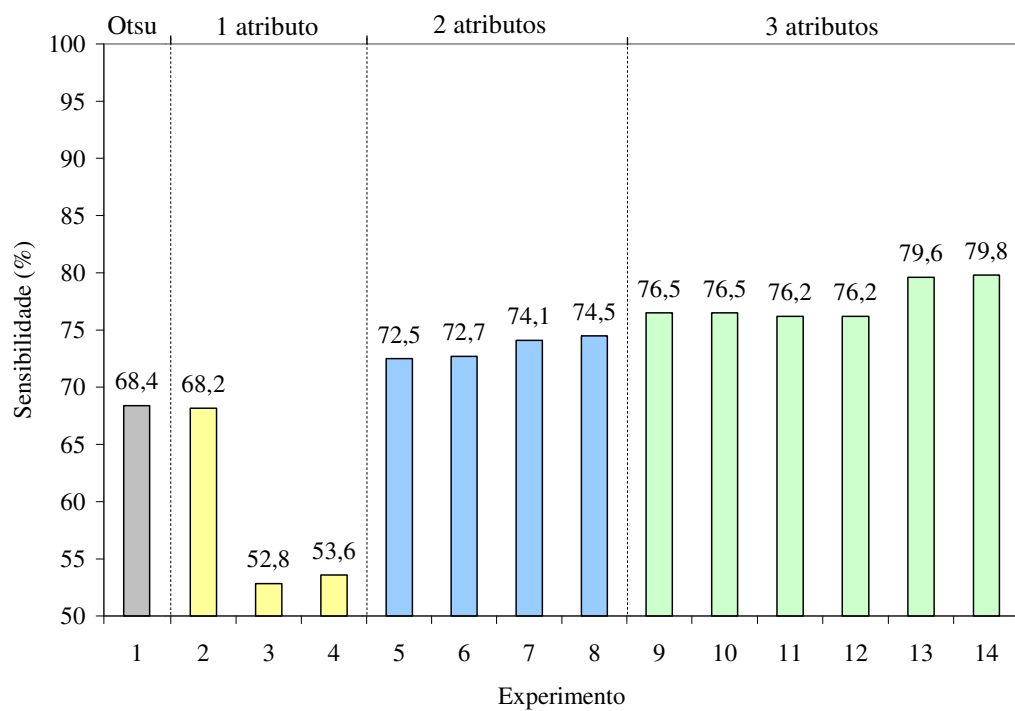


Figura 15 – Desempenho da detecção de nódulos pulmonares na base de dados LIDC com o primeiro conjunto de critérios de segmentação.

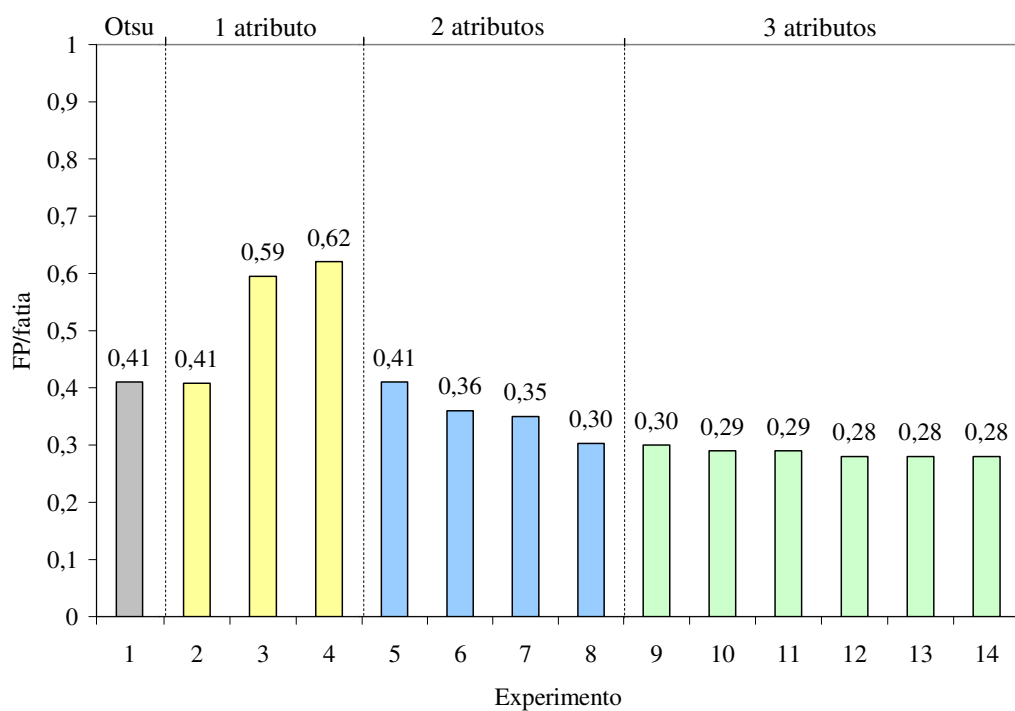


Figura 16 – Falsos positivos por fatia da detecção de nódulos pulmonares na base de dados LIDC com o primeiro conjunto de critérios de segmentação.

É importante ressaltar que os experimentos utilizando mais de um critério na segmentação obtiveram melhores resultados que os utilizando apenas um critério e, os com três critérios foram melhores que os com dois critérios. Portanto, nos experimentos realizados, o acréscimo de atributos na segmentação trouxe consistentemente um ganho de desempenho. Isto mostra que a proposta de utilizar multicritérios na segmentação é promissora.

Os critérios de forma, apesar de não apresentarem bons resultados isoladamente, são bons condutores para a segmentação se utilizados em conjunto com outros critérios normalmente já utilizados (neste caso, baseado na densidade do voxel).

Novamente, nestes experimentos, verificou-se que os resultados de pior e melhor caso, sob condições iguais nas demais configurações (critérios utilizados e conjunto de exames para ajuste de parâmetros de segmentação), foram bastante semelhantes entre si. Isto indica que o método para encontrar os valores dos parâmetros de segmentação utilizando segmentos de referência e GA mostrou-se robusto.

Finalmente, quanto ao grupo de exames utilizados para ajustar os parâmetros de segmentação, os resultados mostraram, como esperado, que quanto maior o grupo de exames utilizado, maior a diversidade de padrões e, portanto, melhor será a adaptação dos parâmetros de segmentação. Ainda assim, os resultados com grupos de exames menores (grupo 1 e 2), obtiveram resultados satisfatórios em relação aos demais. Os resultados sugerem, portanto, que o potencial da técnica de segmentação multicritério não foi totalmente explorado nos experimentos realizados, e os ganhos de desempenho na detecção de nódulos pulmonares podem ser ainda maiores do que mostram as tabelas e gráficos anteriores.

4.4.

Experimentos com o segundo conjunto de critérios para a segmentação

Nesta seção, é apresentada uma segunda implementação da segmentação multicritério. Neste caso, os testes foram realizados sobre a base de dados LIDC. Além disso, outro conjunto de atributos foi introduzido no algoritmo de

segmentação multicritério como se segue.

4.4.1.

Critérios selecionados para segmentação de nódulos

Para os experimentos relatados deste ponto em diante utilizou-se um método automático para escolha dos critérios da segmentação. Portanto, foi utilizada uma técnica de seleção de atributos, como se faz habitualmente quando se deseja reduzir a dimensionalidade em problemas de reconhecimento de padrões.

Para a seleção dos atributos, fez-se uso do pacote de funções para Matlab PRTools versão 4.1 (Duin et al., 2007), sob o método de busca para frente (*forward feature selection*). Neste processo, foram incluídos os descritores geométricos (compacidade, desproporção esférica, densidade esférica, distância radial ponderada, esfericidade e alongação), os descritores de densidade (média, variância, desvio padrão, obliquidade, curtose, energia e entropia) e os descritores de textura (contraste, energia, entropia, homogeneidade e correlação) – todos descritos no apêndice III. Além disso, os exames utilizados são os mesmos selecionados para o ajuste de parâmetros (referente ao grupo 3 do LIDC). Os objetos para a seleção dos atributos a serem utilizados na segmentação multicritério foram gerados da seguinte forma:

- Inicialmente, foi encontrada a região do pulmão como descrito na Seção 3.3.1.
- Em seguida, foi aplicada a limiarização de Otsu incluindo apenas os voxels pertencentes à região do pulmão, encontrando diversas estruturas (nódulos e não-nódulos).
- Então, foram identificados os nódulos delimitados pelos especialistas, sendo que cada delimitação diferente foi dada como um nódulo diferente, ou seja, se um mesmo nódulo foi delimitado diferentemente duas vezes, então dois objetos diferentes são gerados para o nódulo.
- Subtraíram-se as áreas que compõem os nódulos do conjunto de estruturas encontradas pela limiarização e, então, cada componente conectado foi tido como um objeto classificado como não-nódulo.

Como resultado, o método indicou um conjunto de três atributos para classificação, os quais são aqui usados na segmentação multicritério. Os atributos

indicados são: variância, desproporção esférica e alongação.

Aplicando estes atributos à equação de heterogeneidade (1), tem-se:

$$h_1 = n_{Obj12} \cdot \sigma_{Obj12}^2 - (n_{Obj1} \cdot \sigma_{Obj1}^2 + n_{Obj2} \cdot \sigma_{Obj2}^2) \quad (21)$$

$$h_2 = n_{Obj12} \cdot D_{Obj12} - (n_{Obj1} \cdot D_{Obj1} + n_{Obj2} \cdot D_{Obj2}) \quad (22)$$

$$h_3 = n_{Obj12} \cdot E_{Obj12} - (n_{Obj1} \cdot E_{Obj1} + n_{Obj2} \cdot E_{Obj2}) \quad (23)$$

onde h_1 , h_2 e h_3 são heterogeneidades para variância da densidade σ^2 , desproporção esférica D e alongação E , respectivamente (equações para tais atributos encontram-se no apêndice III).

Sendo assim, a equação de heterogeneidade global (3) é definida mais uma vez pela eq. (20) e, portanto, o algoritmo de segmentação multicritério depende novamente de três parâmetros: escala (s) e dois pesos (w_1 e w_2). Como o conjunto de critérios utilizados neste experimento é diferente do anterior, novamente, torna-se necessária a busca dos parâmetros de segmentação como apresentado a seguir.

4.4.2.

Ajuste automático dos valores dos parâmetros de segmentação

Nesta fase, a busca de parâmetros para a segmentação foi realizada em cima dos três grupos do LIDC apresentados anteriormente para melhor avaliar a robustez e eficiência do método.

O GA foi executado 5 vezes para cada grupo de exames sob a seguinte configuração: população com 50 indivíduos, 100 gerações e 70% dos indivíduos deveria ser trocada de uma geração para a seguinte.

A seguir, analisa-se a convergência e a estabilidade do algoritmo genético para a busca de parâmetros da segmentação multicritério dentro destas condições.

4.4.2.1.

Convergência do algoritmo genético

A Figura 17, a Figura 18 e a Figura 19 contêm os gráficos de desempenho

médio através de cinco execuções dos experimentos para ajuste dos valores dos parâmetros do modelo com base nos exames do grupo 1, grupo 2 e grupo 3 do LIDC, respectivamente. A Figura 20 apresenta o gráfico referente à média dos três experimentos citados.

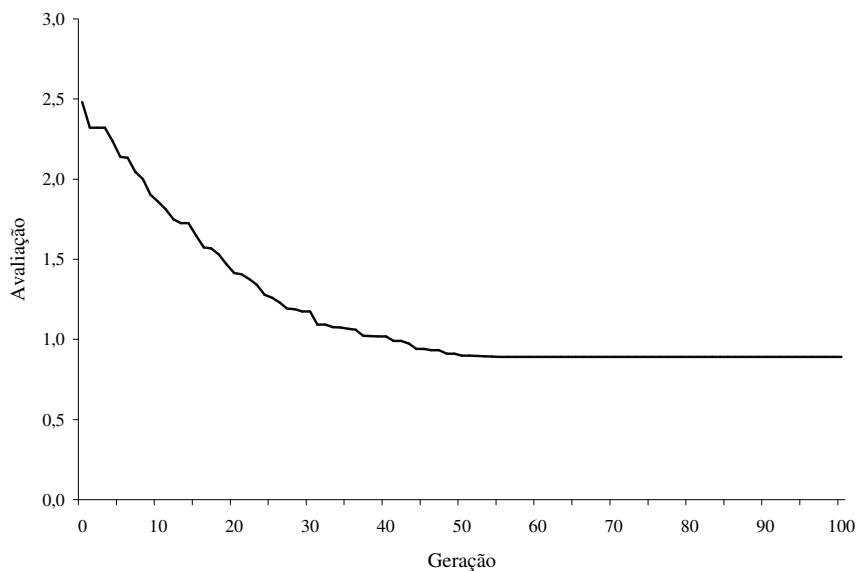


Figura 17 – Convergência de algoritmos genéticos para ajuste de valores dos parâmetros da segmentação com o segundo conjunto de critérios a partir do primeiro grupo de imagens do banco de dados LIDC.

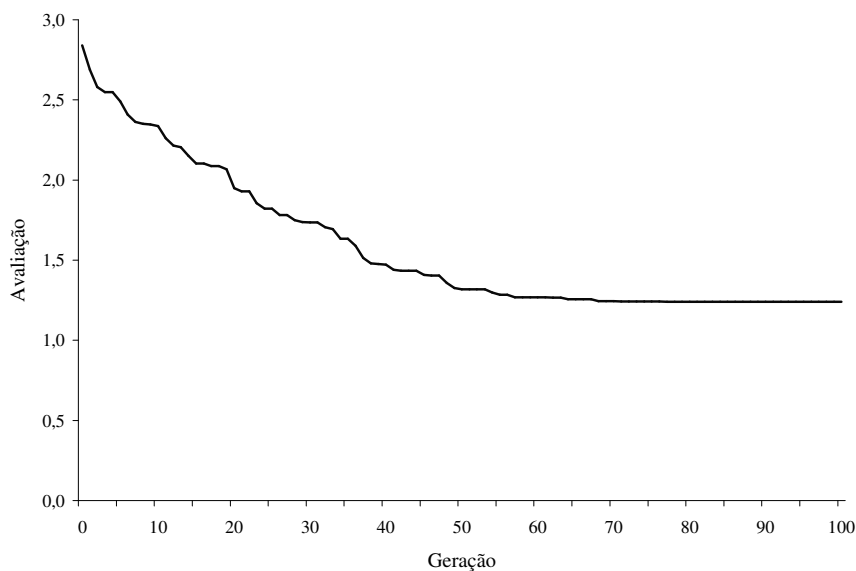


Figura 18 – Convergência de algoritmos genéticos para ajuste de valores dos parâmetros da segmentação com o segundo conjunto de critérios a partir do segundo grupo de imagens do banco de dados LIDC.

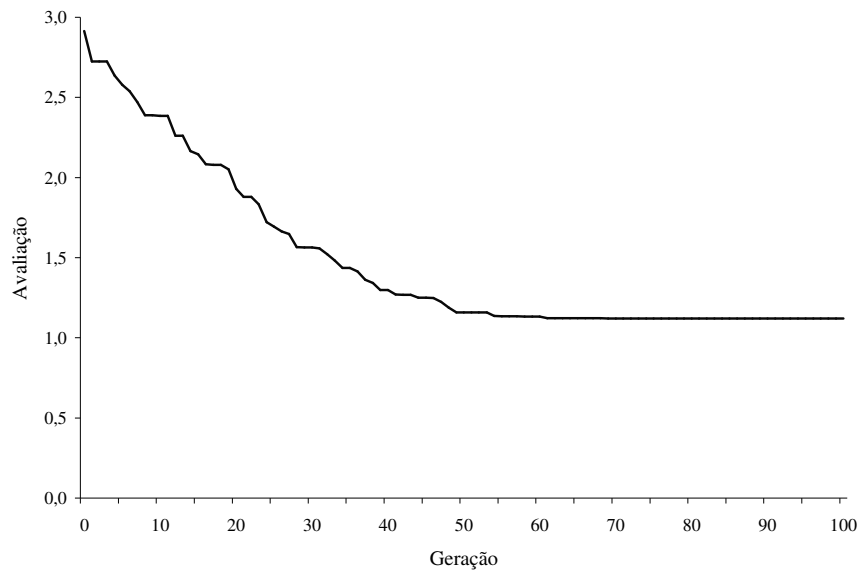


Figura 19 – Convergência de algoritmos genéticos para ajuste de valores dos parâmetros da segmentação com o segundo conjunto de critérios a partir do terceiro grupo de imagens do banco de dados LIDC.

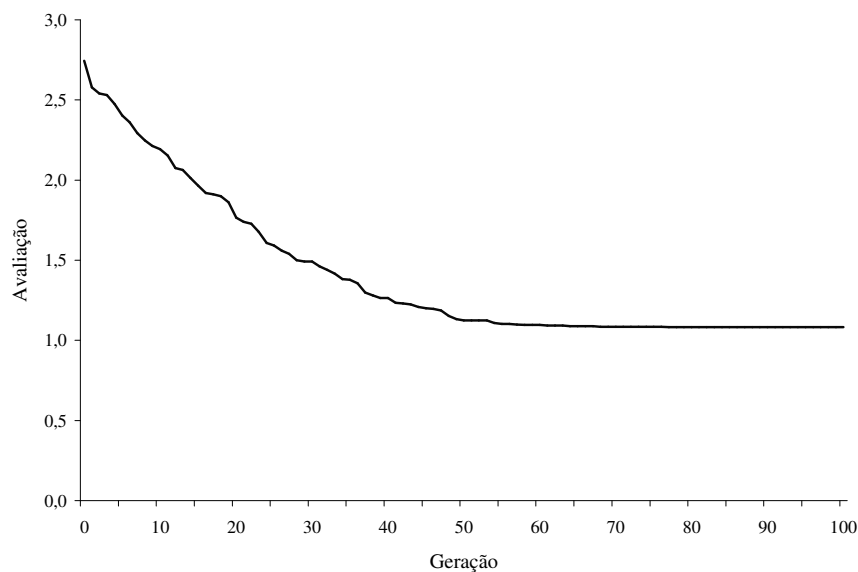


Figura 20 – Convergência de algoritmos genéticos para ajuste de valores dos parâmetros da segmentação com o segundo conjunto de critérios na base de dados LIDC.

Observando estes gráficos, nota-se que a solução converge em aproximadamente 53 (grupo 1), 68 (grupo 2) e 61 (grupo 3) gerações, o que sugere que cem gerações são suficientes para a procura da solução desejada.

4.4.2.2.

Análise de estabilidade dos valores dos parâmetros do modelo

Os conjuntos de valores dos parâmetros encontrados pelo GA e seus respectivos valores de avaliação em cada execução com os grupos 1, 2 e 3 são apresentados na Tabela 10, Tabela 11 e Tabela 12, respectivamente. As colunas correspondem respectivamente ao número da execução, os três parâmetros de ponderação w_1 , w_2 e w_3 , a escala s e o valor de avaliação do conjunto de parâmetros correspondente.

Execução	Parâmetros				Avaliação
	w_1	w_2	w_3	s	
1	0,359	0,125	0,516	75	0,88
2	0,381	0,122	0,497	73	1,01
3	0,358	0,151	0,491	78	0,89
4	0,439	0,198	0,363	75	0,77
5	0,322	0,157	0,521	73	0,91
Média	0,372	0,151	0,478	74,8	0,89
Desvio padrão	0,043	0,031	0,065	2,0	0,09

P

M

Tabela 10 – Resultados do procedimento de parametrização da segmentação com o segundo conjunto de critérios com base no primeiro grupo de imagens LIDC.

Execução	Parâmetros				Avaliação
	w_1	w_2	w_3	s	
1	0,445	0,181	0,374	73	1,22
2	0,325	0,106	0,569	73	1,23
3	0,334	0,113	0,553	74	1,24
4	0,344	0,120	0,536	77	1,27
5	0,411	0,139	0,450	73	1,23
Média	0,372	0,132	0,496	74,0	1,24
Desvio padrão	0,053	0,030	0,082	1,7	0,02

M

P

Tabela 11 – Resultados do procedimento de parametrização da segmentação com o segundo conjunto de critérios com base no segundo grupo de imagens LIDC.

Execução	Parâmetros				Avaliação
	w_1	w_2	w_3	s	
1	0,483	0,213	0,304	75	1,03
2	0,333	0,141	0,526	78	1,18
3	0,352	0,133	0,515	72	1,17
4	0,380	0,189	0,431	74	0,97
5	0,331	0,124	0,545	73	1,23
Média	0,376	0,160	0,464	74,4	1,12
Desvio padrão	0,063	0,039	0,100	2,3	0,11

M

P

Tabela 12 – Resultados do procedimento de parametrização da segmentação com o segundo conjunto de critérios com base no terceiro grupo de imagens.

Observa-se que os valores dos parâmetros encontrados em cada experimento com mesmo grupo de exames convergiram para uma solução semelhante. Os desvios padrão em todos os casos são relativamente pequenos, o que indica novamente que o procedimento de adaptação proposto é robusto.

Para segmentar as demais imagens da base de dados cujos objetos são classificados como apresentado mais adiante, foram escolhidos dois conjuntos de valores de parâmetros de cada conjunto de experimentos com os três grupos de exames diferentes. Os conjuntos escolhidos são os correspondentes às linhas destacadas e que possuem a menor e a maior avaliação, pois referem-se ao melhor (M) e pior (P) caso encontrado.

4.4.3. Segmentação multicritério

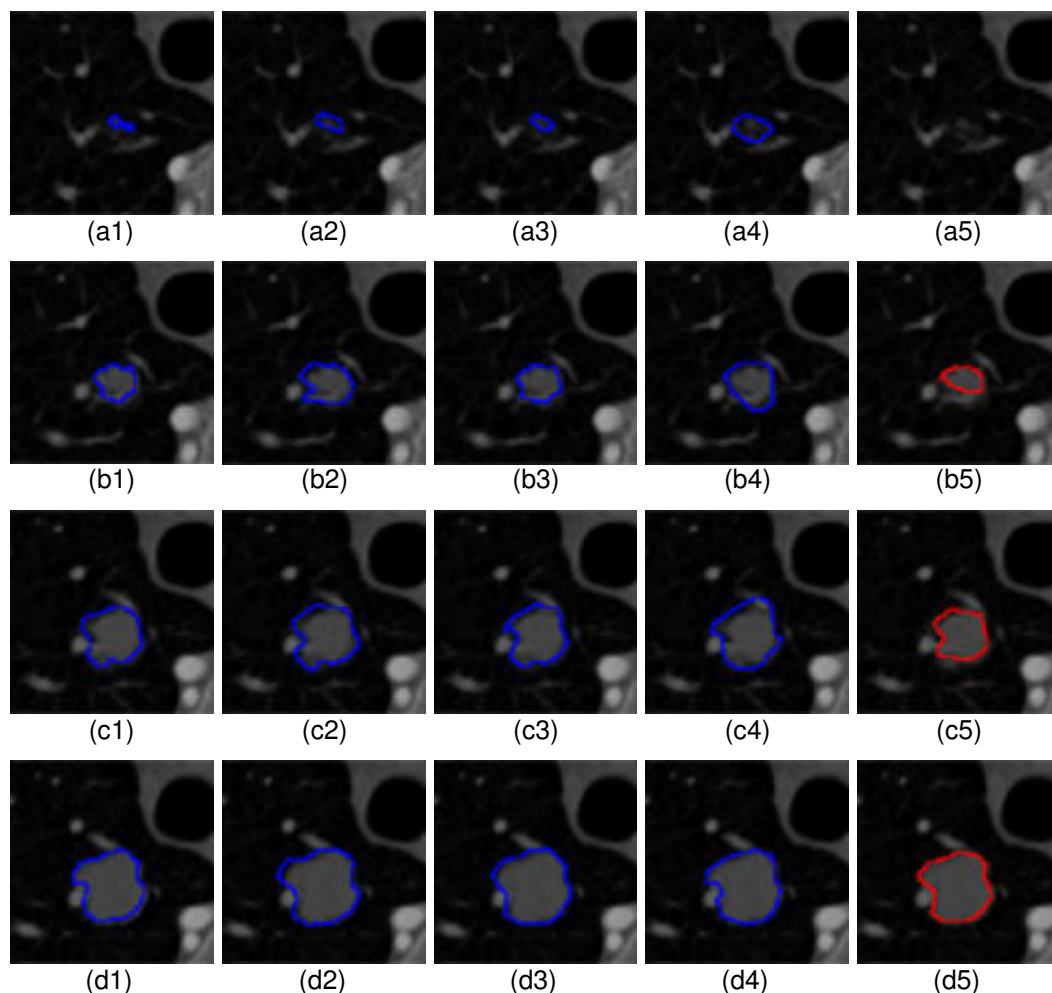
As figuras desta seção ilustram alguns exemplos de resultado da segmentação (Figura 21, Figura 22, Figura 23, Figura 24, Figura 25 e Figura 26).

A Figura 21, a Figura 23 e a Figura 25 apresentam recortes de imagens de TC, fatia a fatia, contendo um nódulo pulmonar com suas respectivas delimitações. Em tais figuras, as letras identificam as fatias e os números identificam a procedência da delimitação do nódulo, sendo que os números de 1 a

4 correspondem a um especialista do LIDC e 5, à segmentação multicritério. Desta forma, as identificações a1, a2, a3, a4 e a5 indicam que as imagens correspondem a uma mesma fatia “a”, cujo nódulo foi delimitado por 4 especialistas (1, 2, 3 e 4) e pelo método de segmentação multicritério (5).

A Figura 22, a Figura 24, e a Figura 26 apresentam, respectivamente, visualizações tridimensionais dos nódulos ilustrados pela Figura 21, Figura 23 e Figura 25 sob dois ângulos diferentes.

Embora o objetivo principal deste trabalho seja a detecção de nódulos e não a exatidão da segmentação, vale a pena notar que os nódulos automaticamente segmentados são, em geral, visualmente semelhantes aos resultados produzidos por especialistas humanos e usados como referência.



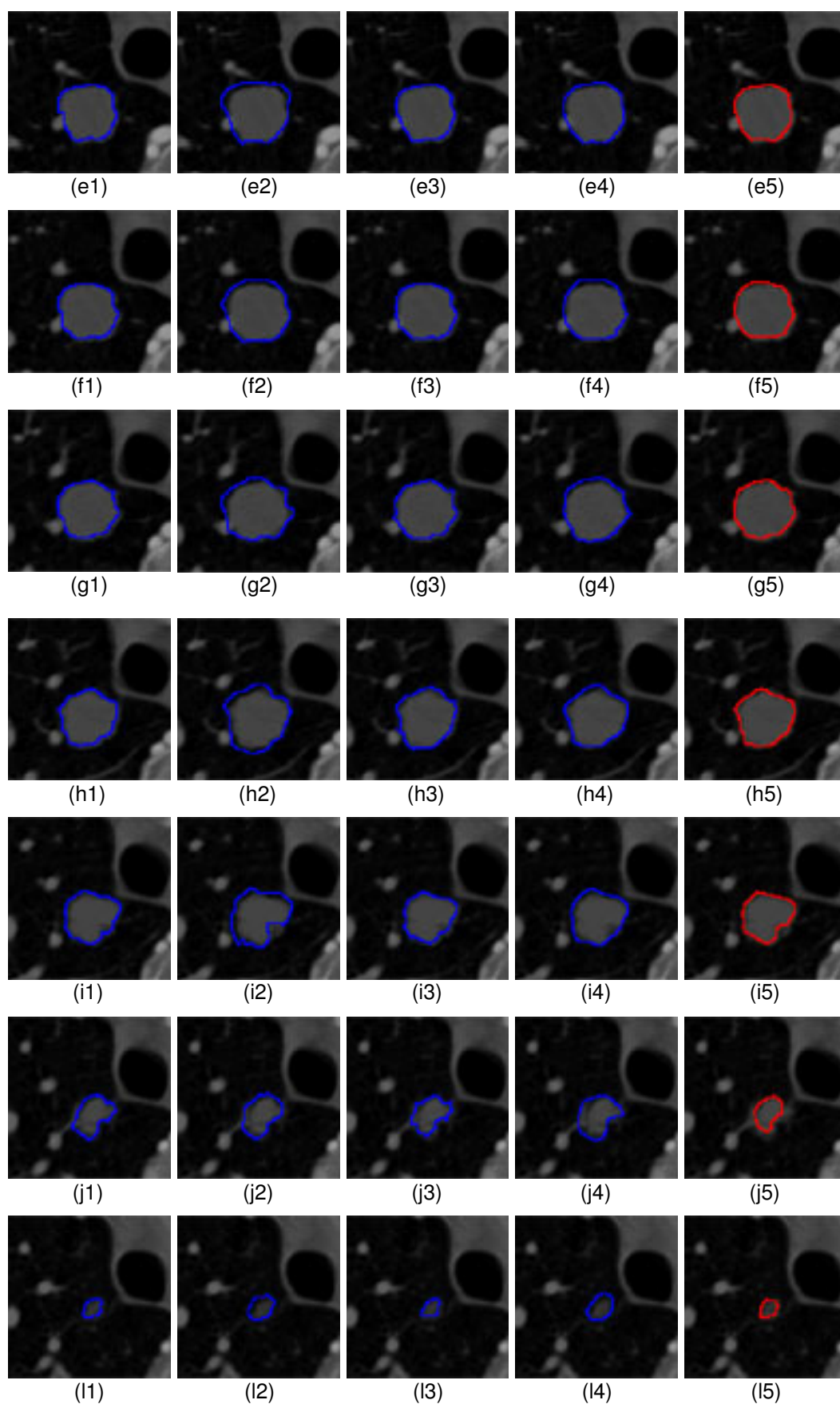


Figura 21 – Nódulo pulmonar delimitado por quatro especialistas e pela segmentação multicritério.

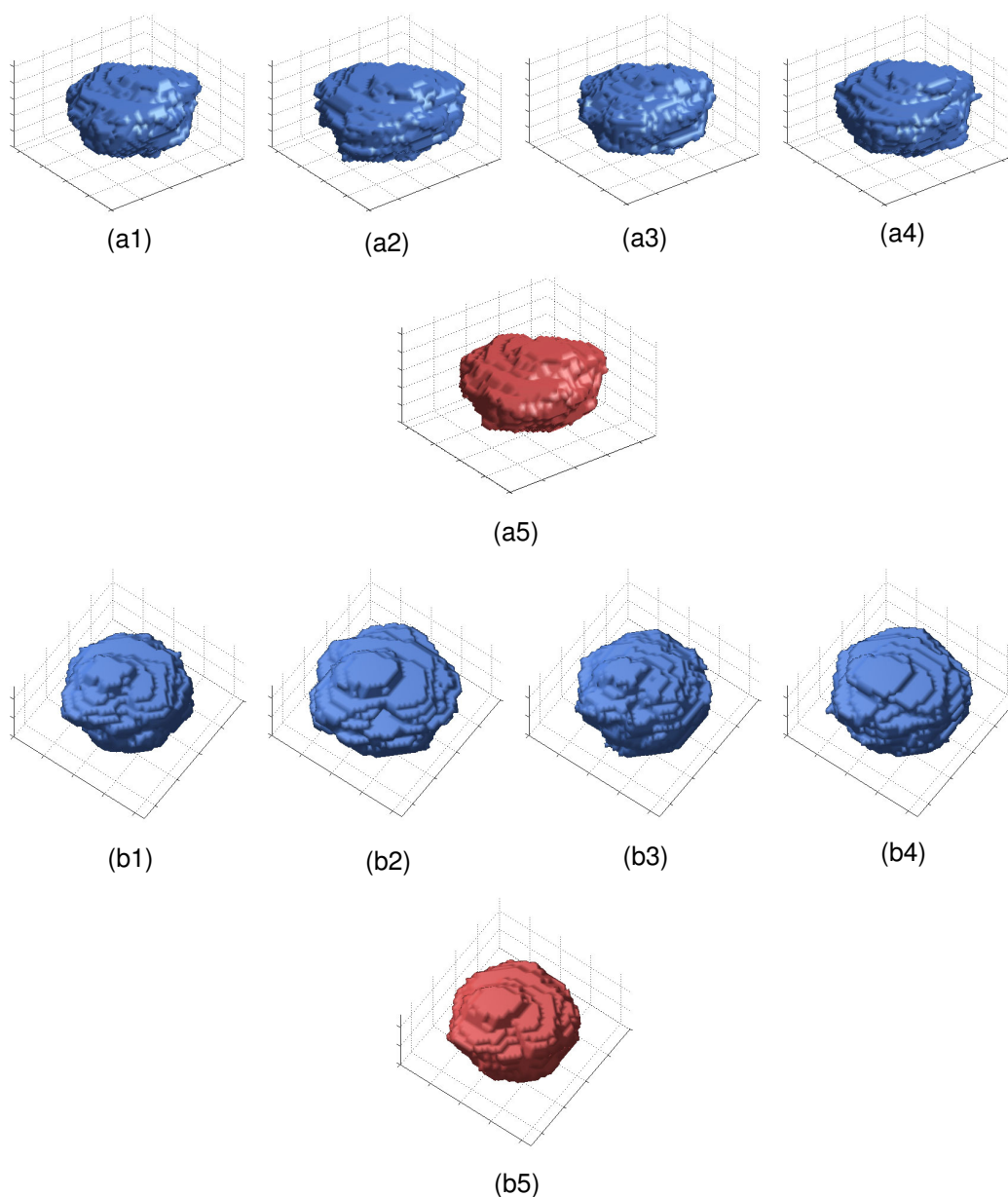


Figura 22 – Duas visualizações de um mesmo nódulo sob diferentes ângulos: (a1) e (b1) delimitados pelo especialista 1; (a2) e (b2) delimitados pelo especialista 2; (a3) e (b3) delimitados pelo especialista 3; (a4) e (b4) delimitados pelo especialista 4; (a5) e (b5) segmentados automaticamente pelo algoritmo proposto.

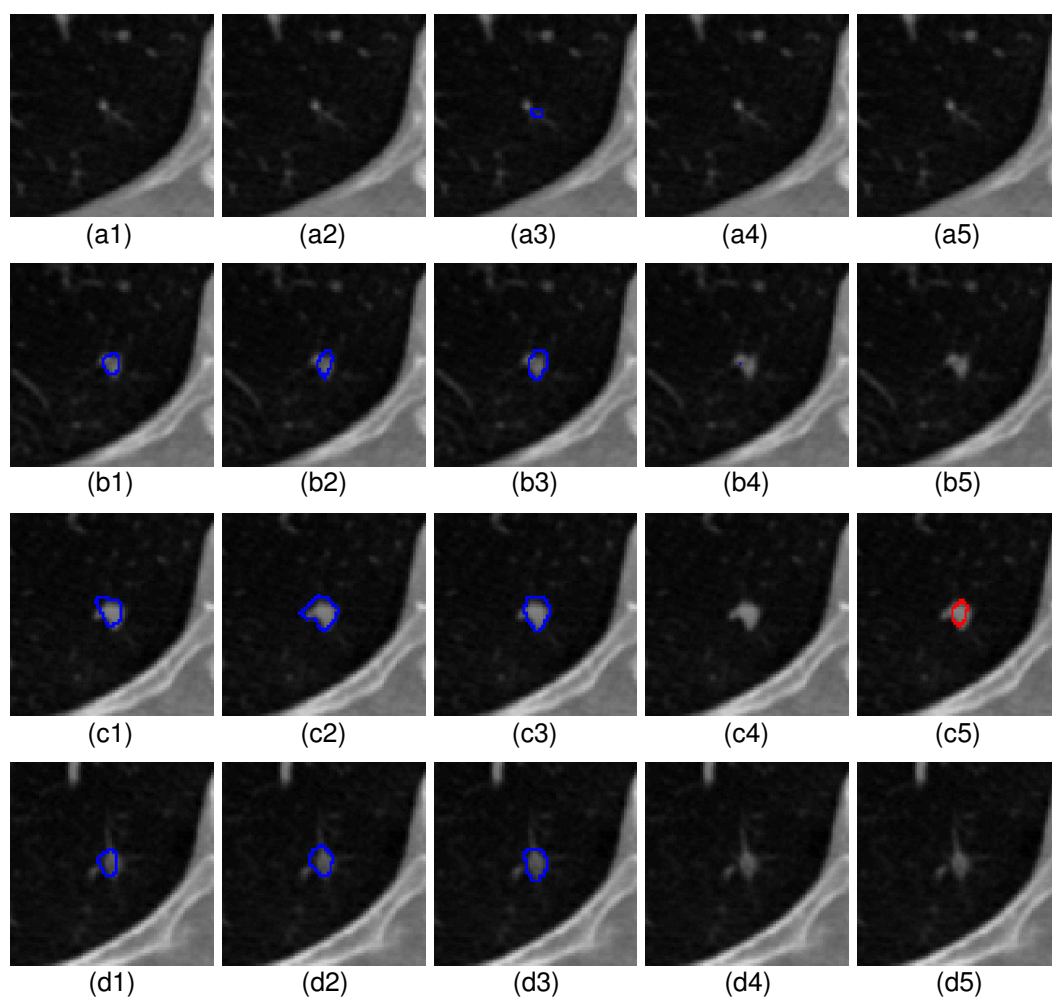


Figura 23 – Nódulo pulmonar delimitado por quatro especialistas e pela segmentação multicritério.

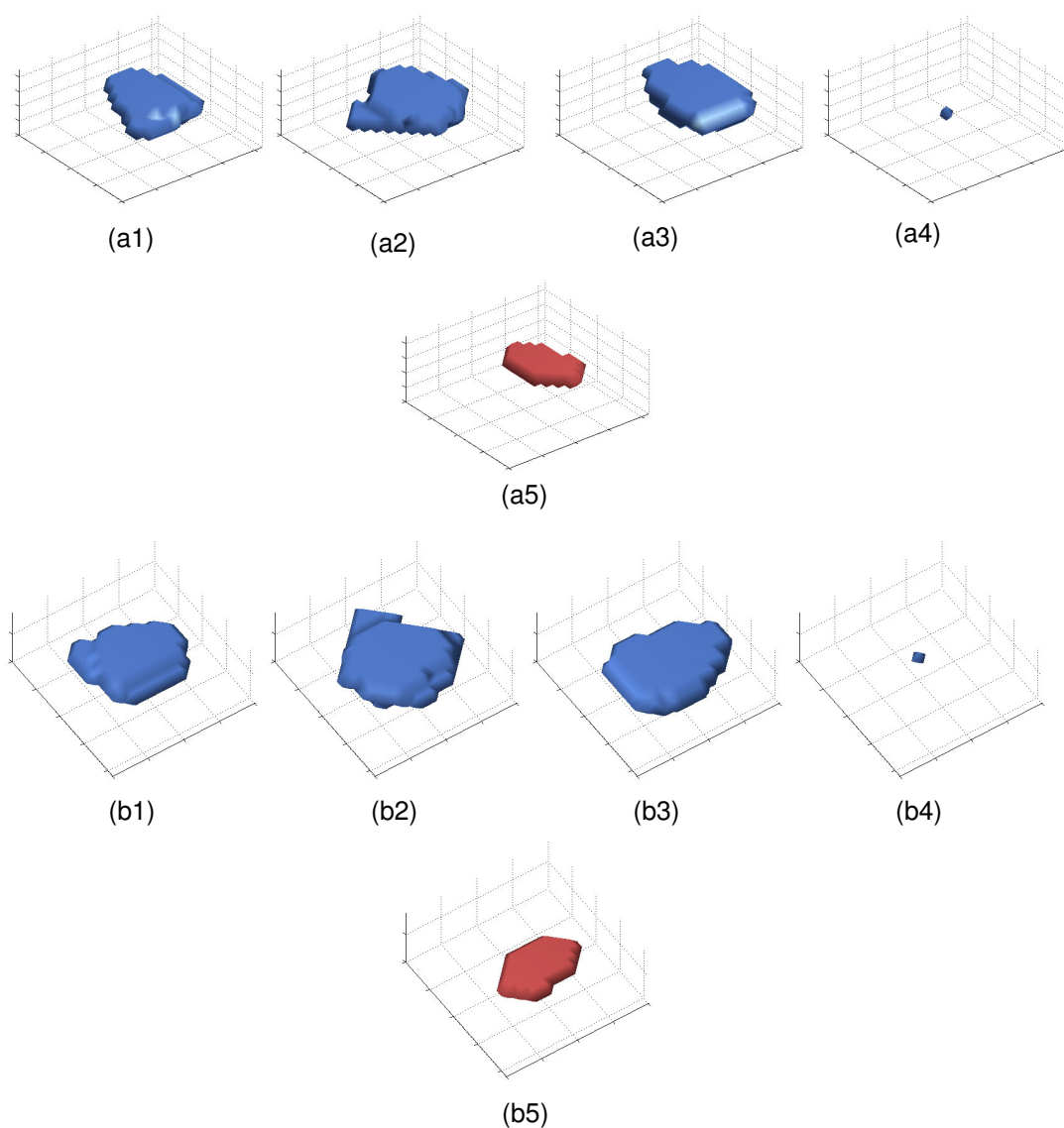
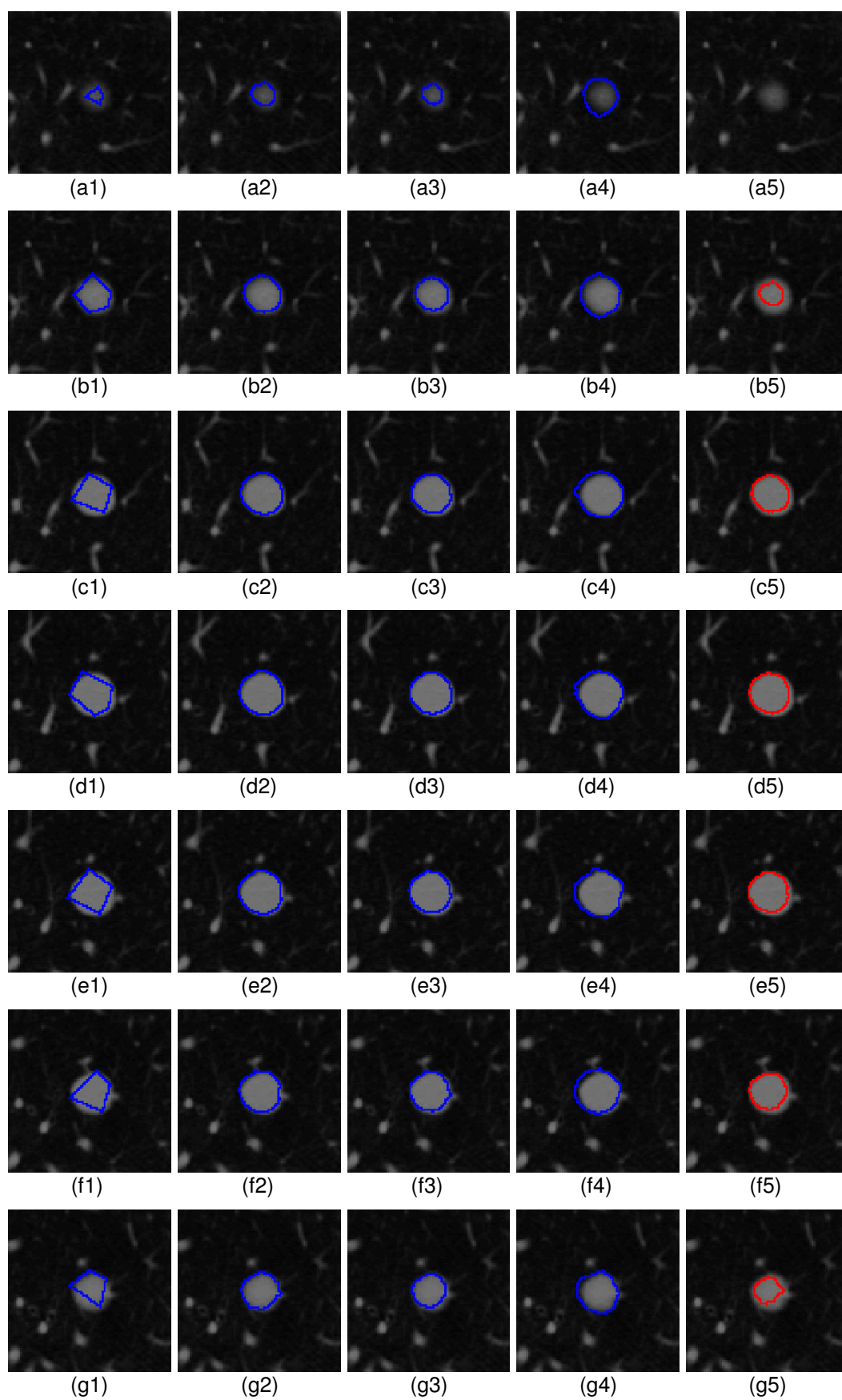


Figura 24 – Duas visualizações de um mesmo nódulo sob diferentes ângulos: (a1) e (b1) delimitados pelo especialista 1; (a2) e (b2) delimitados pelo especialista 2; (a3) e (b3) delimitados pelo especialista 3; (a4) e (b4) delimitados pelo especialista 4; (a5) e (b5) segmentados automaticamente pelo algoritmo proposto.



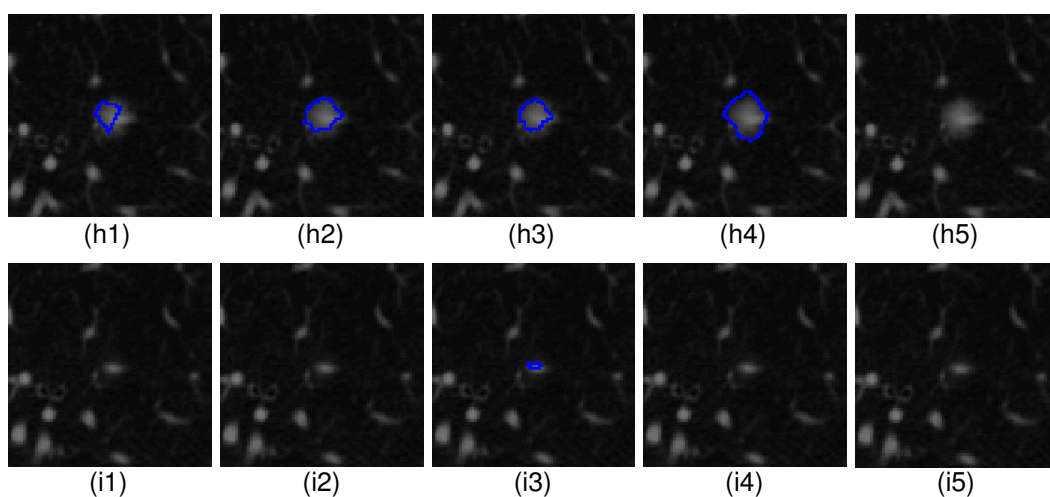


Figura 25 – Nódulo pulmonar delimitado por quatro especialistas e pela segmentação multicritério.

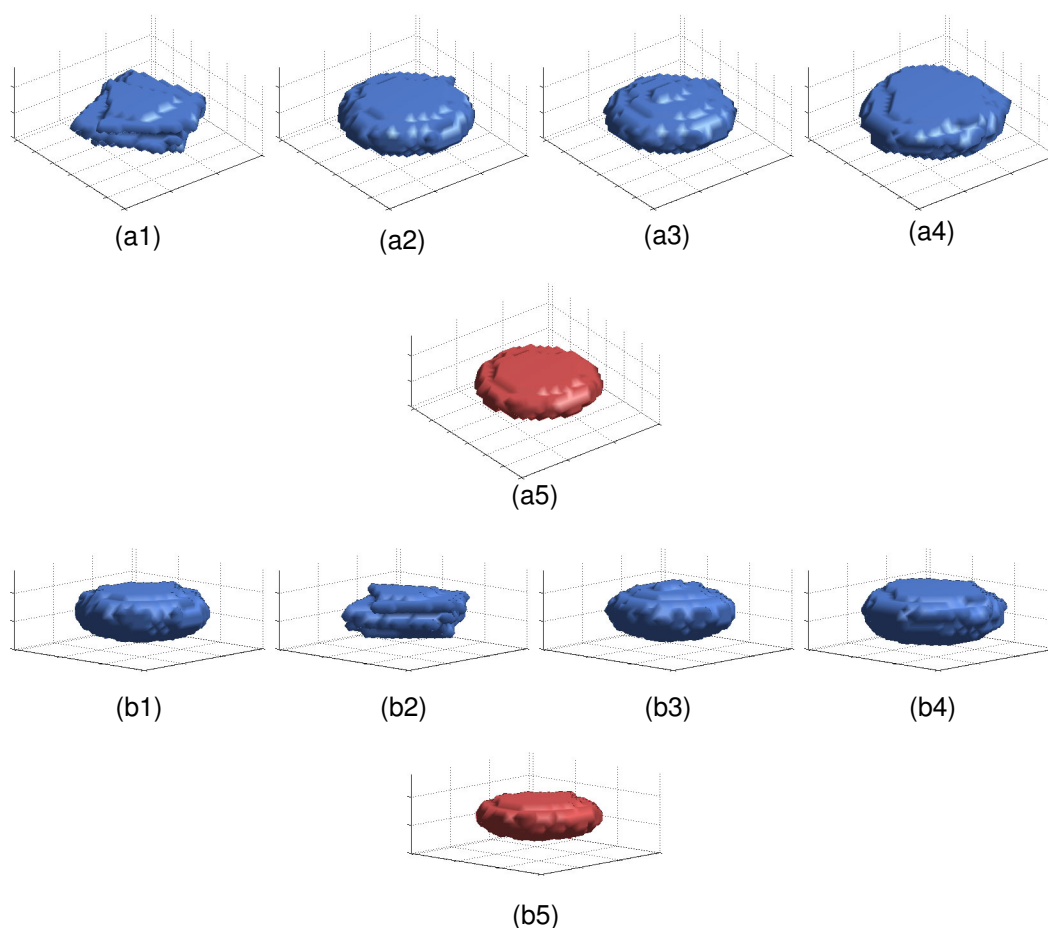


Figura 26 – Duas visualizações de um mesmo nódulo sob diferentes ângulos: (a1) e (b1) delimitados pelo especialista 1; (a2) e (b2) delimitados pelo especialista 2; (a3) e (b3) delimitados pelo especialista 3; (a4) e (b4) delimitados pelo especialista 4; (a5) e (b5) segmentados automaticamente pelo algoritmo proposto.

A Tabela 13 apresenta a avaliação da segmentação dos 33 exames de teste do LIDC, após a busca de parâmetros conforme indicado anteriormente. A avaliação indicada refere-se à função $Fa(X,P)$ calculada pela eq. (6).

O exame 1 da Tabela 13 possui o menor valor de avaliação e, por isso, representa a melhor segmentação realizada segundo este indicador. Este exame corresponde ao mesmo ilustrado na Figura 21. Nota-se que, visualmente, a segmentação foi bastante parecida com as delimitações fornecidas pelos especialistas.

Por outro lado, a Figura 23 refere-se ao nódulo do exame 26 da Tabela 13, o qual teve uma avaliação de 1,288 (uma das piores avaliações, sendo melhor apenas que os exames contendo nódulos que foram indicados pelos especialistas por apenas um voxel). Neste caso, as delimitações fornecidas pelos especialistas também divergirem entre si.

Os exames de 30 a 33 na Tabela 13 referem-se aos exames cujos nódulos são muito pequenos, os quais foram eliminados pela pré-segmentação. Estes nódulos foram excluídos pela limiarização que tinha como objetivo separar a área do pulmão de estruturas internas (vasos sanguíneos e nódulos). Nota-se que o valor de avaliação é 1, pois a soma dos pesos dos voxels não incluídos ao objeto pelo método é igual à soma dos pesos dos voxels que foram incluídos por pelo menos um especialista no segmento desejado (veja a eq. (6)). É importante notar que estes nódulos não foram excluídos pela segmentação multicritério que é a proposta deste trabalho. Entretanto, os resultados da detecção, apresentados na próxima seção, foram prejudicados pelo erro oriundo da pré-segmentação. Isto indica a importância da segmentação no processo de detecção, já que influencia diretamente nos demais procedimentos de processamento de imagens.

Os nódulos dos exames 27, 28 e 29, apesar de pequenos, não foram excluídos na fase de pré-segmentação. Porém, os especialistas apenas apontaram a localização de tais nódulos, não fornecendo, portanto, suas delimitações. Desta forma, os valores da avaliação destes exames são maiores que os demais, já que a referência foi baseada em apenas um voxel.

O exame identificado por 20, o qual obteve uma avaliação de 1,187, é o mesmo ilustrado anteriormente pela Figura 25. Nota-se que a segmentação produzida pelo método neste exame proposto é visualmente boa. Na verdade, foram visualmente boas as segmentações obtidas em todos os exames que

obtiveram avaliação igual ou inferior a este valor. Trata-se da maioria dos exames analisados neste grupo de experimentos.

Exame	Avaliação	Exame	Avaliação	Exame	Avaliação
1	0,558	12	1,004	23	1,241
2	0,631	13	1,054	24	1,249
3	0,654	14	1,060	25	1,283
4	0,691	15	1,102	26	1,288
5	0,701	16	1,110	27	3,300
6	0,794	17	1,132	28	5,300
7	0,841	18	1,149	29	8,300
8	0,865	19	1,165	30	1,000
9	0,927	20	1,187	31	1,000
10	0,988	21	1,208	32	1,000
11	1,001	22	1,210	33	1,000

Tabela 13 – Avaliação da segmentação multicritério com o segundo conjunto de critérios na base de dados LIDC.

4.4.4. Detecção de nódulos

Nesta seção, são apresentados os resultados da classificação dos objetos obtidos pela segmentação das imagens de TC utilizando os parâmetros encontrados.

Novamente, nestes experimentos, os dez exames selecionados para ajuste dos parâmetros de segmentação foram excluídos do conjunto de teste, no entanto foram usados no conjunto de treinamento do classificador SVM. Além disso, foi empregada a técnica “deixe um de fora”. Assim, para cada exame do conjunto de teste, todos os demais foram usados para treinar o SVM.

Para avaliar a influência do método de segmentação proposto sobre o resultado final da classificação, foram realizados alguns experimentos com diferentes segmentações. A Tabela 14 apresenta os resultados obtidos por estes

experimentos na base de dados LIDC. O primeiro experimento corresponde à segmentação pela limiarização de Otsu já apresentada anteriormente (juntamente com os experimentos do primeiro conjunto de critérios). Os demais experimentos utilizaram a metodologia proposta baseada em crescimento de regiões, sendo que nos experimentos 2, 3 e 4 foram utilizados apenas um critério na segmentação, nos experimentos de 5 a 8 foram empregados dois critérios na segmentação, enquanto que nos experimentos de 9 a 14 foram empregados os três critérios selecionados para este primeiro conjunto de atributos para segmentação. A Tabela 14 também mostra qual grupo de exames do LIDC foi utilizado na fase de ajuste de parâmetros e qual caso (pior caso P, melhor caso M ou caso único U) foi utilizado na segmentação.

A Figura 27 e a Figura 28 mostram graficamente uma comparação dos principais indicadores de desempenho quando se trata de detecção de nódulos pulmonares (sensibilidade e falsos positivos por fatia). Nestes gráficos, os números dos experimentos são os mesmos indicados na primeira coluna da Tabela 14.

Mais uma vez, nota-se claramente o importante papel da segmentação no processo de reconhecimento de imagens e, neste caso, para a detecção de nódulos pulmonares. Isto se deve ao fato de que objetos diferentes resultam em medidas descritivas diferentes o que, conseqüentemente, pode implicar em uma classificação errada.

O experimento com segmentação baseada na limiarização de Otsu foi um dos que obteve o pior resultado com sensibilidade de 68,4 % e 0,41 *FP*/fatia. Isto mostra que a simples utilização dos valores de densidade da imagem não é suficiente para descrever objetos.

Exp	Segmentação	Ajuste de parâmetros	Caso	<i>S (%)</i>	<i>FP/fatia</i>	<i>FP/exame</i>
1	Otsu	–	–	68,4	0,41	62,3
2	variância	Grupo 3	U	67,8	0,40	61,7
3	elongação	Grupo 3	U	50,6	0,59	90,1
4	desproporção esférica	Grupo 3	U	53,6	0,62	94,6
5	variância e elongação	Grupo 3	P	73,8	0,32	48,2
6	variância e elongação	Grupo 3	M	74,3	0,32	48,2
7	variância e desproporção esférica	Grupo 3	P	74,4	0,30	46,2
8	variância e desproporção esférica	Grupo 3	M	74,8	0,30	45,9
9	variância, elongação e desproporção esférica	Grupo 1	P	76,8	0,26	40,1
10	variância, elongação e desproporção esférica	Grupo 1	M	78,4	0,26	40,0
11	variância, elongação e desproporção esférica	Grupo 2	P	77,8	0,25	37,5
12	variância, elongação e desproporção esférica	Grupo 2	M	78,0	0,24	37,4
13	variância, elongação e desproporção esférica	Grupo 3	P	80,9	0,24	36,8
14	variância, elongação e desproporção esférica	Grupo 3	M	80,9	0,23	34,8

Tabela 14 – Desempenho de diferentes segmentações avaliadas sobre o banco de dados LIDC.

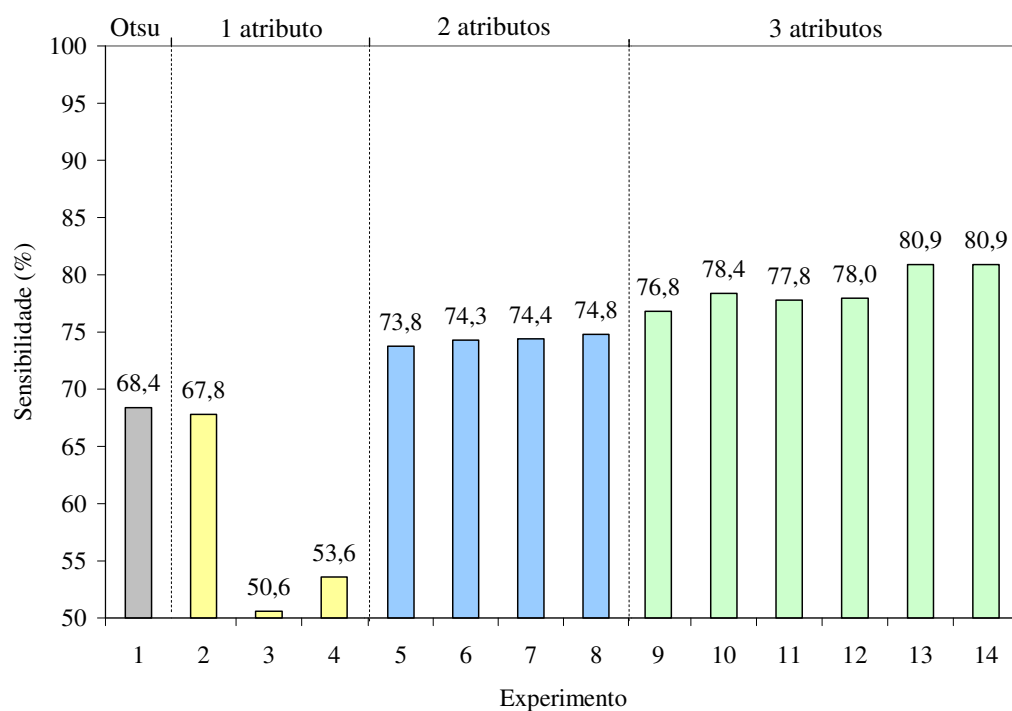


Figura 27 – Sensibilidade da detecção de nódulos pulmonares em diferentes segmentações da mesma base de dados LIDC.

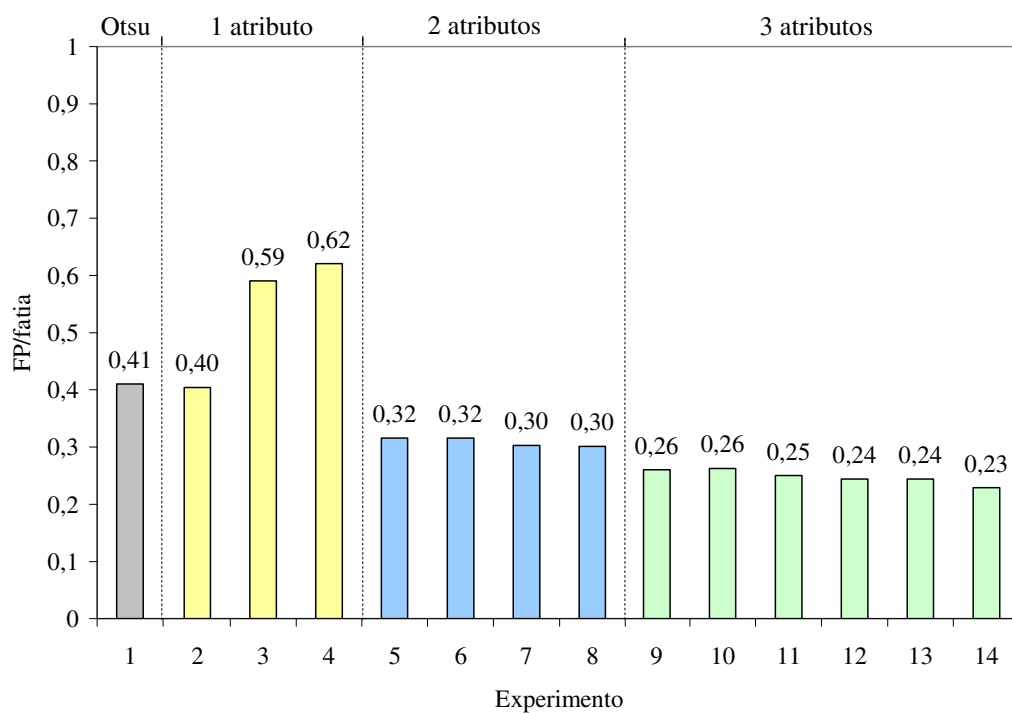


Figura 28 – Falsos positivos encontrados por fatia na detecção de nódulos pulmonares em diferentes segmentações da mesma base de dados.

Entre os experimentos com segmentação utilizando apenas um critério, aqueles cujo critério refere-se à forma do objeto (elongação e desproporção esférica) não produziram bons resultados chegando a, no máximo, 53,6% de sensibilidade. Este resultado reforça o pensamento que, para segmentação, os critérios de forma, isoladamente, não são geram uma boa segmentação.

Os experimentos utilizando mais de um critério na segmentação alcançaram melhores resultados que os utilizando apenas um critério e, os com três critérios foram ainda melhores que os com dois critérios. Isto indica, novamente, que a proposta de utilizar mais de um critério na segmentação é promissora. Os resultados sugerem ainda que a inclusão de mais critérios pode trazer ganhos adicionais de desempenho. Esta possibilidade não pôde ser investigada apropriadamente nestes experimentos por limitação quanto à base de dados – quanto maior o número de critérios, maior o número de parâmetros a estimar e, portanto, maior deve ser o número de exames para treinamento.

Os critérios de forma, apesar de não apresentarem bons resultados isoladamente, são bons condutores para segmentação se utilizados em conjunto com outros critérios normalmente já utilizados (neste caso, baseado na densidade do voxel). Em outras palavras, os critérios de forma não substituem critérios de densidade para segmentação, porém auxiliam neste processo.

O método para encontrar os valores dos parâmetros de segmentação mostrou-se novamente robusto, já que os resultados de pior e melhor caso, sob condições iguais nas demais configurações (critérios utilizados e conjunto de exames para ajuste de parâmetros de segmentação), são bem próximos.

Por fim, os resultados confirmam que o grupo de exames utilizados para ajustar os parâmetros de segmentação pode influenciar no resultado final, já que tal grupo representa o resultado desejado pela segmentação. Quanto maior a diversidade de padrões, expresso aqui pelo número de exames utilizado, melhor será a adaptação do conjunto de valores dos parâmetros de segmentação. Ainda assim, os resultados com grupos de exames mais restritos (grupo 1 e 2) foram satisfatórios em relação aos demais, tendo sensibilidade entre 76,8% e 78,4% com falsos positivos entre 0,24 e 0,26 por fatia. Os dois últimos experimentos, que correspondem à utilização de mais exames na fase de ajuste de parâmetros, tiveram melhor desempenho, o que se justifica pelo fato de ter um conjunto de referência maior e, portanto, mais representativo, chegando a 80,9% de

sensibilidade e 0,24 e 0,23 falsos positivos por fatia no pior e melhor caso, respectivamente.

Os experimentos realizados demonstraram, portanto, as vantagens de se utilizarem múltiplos critérios, isto é, múltiplos atributos para comporem o critério de segmentação. Os resultados sugerem que os ganhos de desempenho serão ainda maiores, havendo maior número de exames para a fase de treinamento, o que permitirá uma melhor estimativa dos parâmetros do algoritmo de segmentação e que mais atributos sejam utilizados para compor o critério de homogeneidade.

Por fim, os resultados apresentados indicam que o método proposto é capaz de fornecer uma boa sugestão para o especialista, vindo ao encontro do objetivo de um CAD, isto é, fornecer uma segunda opinião sem substituir o especialista.

4.5. Custo computacional

O maior problema encontrado durante os experimentos está relacionado ao tempo de processamento do algoritmo de segmentação.

O tempo gasto para a segmentação de um exame inteiro varia de acordo com o número de fatias do mesmo, mas foi de aproximadamente de 1h30min em média.

Para o procedimento de ajuste de parâmetros, a segmentação foi realizada várias vezes, pois para cada conjunto de parâmetros avaliado, uma nova segmentação deveria ser executada. Neste caso, os exames não foram segmentados por completo, já que a avaliação do conjunto de parâmetros baseia-se apenas na segmentação do nódulo. Assim, a avaliação de cada indivíduo durou 10min em média (o tempo de segmentação também varia de acordo com o conjunto de parâmetros) e cada execução do GA durou aproximadamente 11 dias.

Como os experimentos foram realizados em MATLAB, não há dúvida de que a implementação do método em uma linguagem mais eficiente computacionalmente deve diminuir significativamente o custo computacional do algoritmo.