

Aristófanês Corrêa Silva

**Algoritmos para Diagnóstico
Assistido de Nódulos
Pulmonares Solitários em
Imagens de Tomografia
Computadorizada**

TESE DE DOUTORADO

DEPARTAMENTO DE INFORMÁTICA
Programa de Pós-graduação em
Informática

Rio de Janeiro
Fevereiro de 2004



Aristófanês Corrêa Silva

**Algoritmos para Diagnóstico Assistido de
Nódulos Pulmonares Solitários em Imagens
de Tomografia Computadorizada**

Tese de Doutorado

Tese apresentada ao Programa de Pós-graduação em
Informática do Departamento de Informática da PUC-Rio
como parte dos requisitos parciais para obtenção do título
de Doutor em Informática.

Orientador: Prof. Marcelo Gattass

Co-Orientador: Prof. Paulo Cezar Pinto Carvalho

Rio de Janeiro
Fevereiro de 2004



Aristófanês Corrêa Silva

**Algoritmos para Diagnóstico Assistido de
Nódulos Pulmonares Solitários em Imagens
de Tomografia Computadorizada**

Tese apresentada ao Programa de Pós-graduação em
Informática do Departamento de Informática do Centro
Técnico Científico da PUC-Rio como parte dos requisitos
parciais para obtenção do título de Doutor em Informática.
Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Marcelo Gattass

Orientador

Departamento de Informática — PUC-Rio

Prof. Paulo Cezar Pinto Carvalho

Co-Orientador

Departamento de Informática — PUC-Rio

Prof. Sidnei Paciornik

PUC-Rio

Prof. Marcelo Dreux

PUC-Rio

Prof. Rodolfo Acatauassú Nunes

UERJ

Prof. Aura Conci

UFF

Prof. Luiz Henrique de Figueiredo

IMPA

Prof. Waldemar Celles

PUC-Rio

Prof. José Eugenio Leal

Coordenador Setorial do Centro Técnico Científico —

PUC-Rio

Rio de Janeiro, 9 de Fevereiro de 2004

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Aristófanês Corrêa Silva

Graduou-se em Ciência da Computação na Universidade Federal do Maranhão – UFMA. Fez mestrado na Universidade Federal do Maranhão – UFMA em Computação Gráfica.

Ficha Catalográfica

Silva, Aristófanês C.

Algoritmos para Diagnóstico Assistido de Nódulos Pulmonares Solitários em Imagens de Tomografia Computadorizada/ Aristófanês Corrêa Silva; orientador: Marcelo Gattass; co-orientador: Paulo Cezar Pinto Carvalho. — Rio de Janeiro : PUC-Rio, Departamento de Informática, 2004.

v., 140 f: il. ; 30 cm

1. Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática.

Inclui referências bibliográficas.

1. Informática – Teses. 2. Diagnóstico de Nódulo Pulmonar Solitário. 3. Textura. 4. Geometria. 5. Análise Discriminante Linear de Fisher. 6. Redes Neurais Perceptron de Múltiplas Camadas. I. Gattass, Marcelo. II. Carvalho, Paulo Cezar Pinto. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. IV. Título.

CDD: 004

Aos meus pais, Acyr e Maria Nilde.

Agradecimentos

Neste longo caminho, vários são os agradecimentos acumulados. Espero contemplar a todos e, se porventura, esquecer de alguém, com certeza se deve a um lapso temporário de memória.

À DEUS, por tudo.

À minha esposa Tânia Castro pelo incentivo, sacrifício, paciência, compreensão e carinho ao longo deste período.

À minha família, que sempre esteve comigo me apoiando em todas as horas, obrigado pelo incentivo e carinho.

Aos meus dois orientadores: Marcelo Gattass e Paulo Cezar Pinto Carvalho. O primeiro, confiou e acreditou em mim no início e nos momentos mais difíceis do doutorado. O segundo, pelas curtas, mas proveitosas discussões, pelos incentivos, pelos ensinamentos, pelas críticas, e principalmente, pela segurança na orientação. Muito obrigado aos dois.

À instituição a que pertenço: a Universidade Federal do Maranhão – UFMA; a instituição que me acolheu: a Pontifícia Universidade Católica do Rio de Janeiro – PUC-RJ; e a instituição que me adotou e me deu suporte: o Instituto de Matemática Pura e Aplicada – IMPA.

Ao Dr. Rodolfo Acatauassú Nunes pelos ensinamentos, apoio, confiança e incentivo que me foram extremamente importantes durante toda a tese.

Ao Dr. Rodolfo Acatauassú Nunes e sua equipe, pelo suporte médico, e ao pessoal do Instituto Fernandes Figueira, em particular à Dra. Márcia Cristina Bastos Boechat, pelas imagens fornecidas.

Ao Prof. Luiz Velho pelas boas idéias e por me abrir as portas do Visgraf.

À Carolina Alfaro, pelas traduções, correções e dicas em meus artigos e neste trabalho.

Ao Prof. Sidnei Paciornik pelas boas idéias durante a defesa de proposta de tese.

Aos funcionários do IMPA e do Tecgraf pelo apoio logístico.

Ao Tecgraf pela ajuda financeira que me possibilitou ir em alguns congressos.

À CAPES que me propiciou a bolsa PICDT, fundamental para a realização desse trabalho.

Aos professores da UFMA, em especial ao Anselmo Paiva e a Maria da Guia.

Aos meus amigos da minha “Ilhinha do Amor” - São Luís (MA): Mário Borges, Gutemberg Santiago, Marcos Santos, Adriana Sousa, Salete

Farias, Eveline Sá, Jeane Diniz, Evaldinólia Gilbertoni e ..., que durante este período sempre me deram apoio e incentivo.

Aos meus amigos Visgrafianos (Laboratório Visgraf - IMPA): Adailson Peixoto, Antônia Lucinelma, Beatriz Alvarez, Fábio Marcos, Gustavo Pierre, José Luiz, Lourena Karen, Nair Duarte, Paula Lucena, Perfilino Eugênio e Sérgio Estevão, pela amizade, convivência, aprendizagem, companheirismo, enfim pelos bons momentos.

Muito obrigado a todos.

Resumo

Silva, Aristófares C.; Gattass, Marcelo; Carvalho, Paulo Cezar Pinto. **Algoritmos para Diagnóstico Assistido de Nódulos Pulmonares Solitários em Imagens de Tomografia Computadorizada**. Rio de Janeiro, 2004. 140p. Tese de Doutorado — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

O presente trabalho visa desenvolver uma ferramenta computacional para sugerir sobre a malignidade ou benignidade de Nódulos Pulmonares Solitários, através da análise de medidas de textura e geometria obtidas a partir das imagens de tomografia computadorizada.

São propostos quatro grupos de métodos com o objetivo de sugerir o diagnóstico para o nódulo. Os grupos de métodos são divididos de acordo com suas características comuns. O Grupo I trata dos métodos baseados em textura adaptados para 3D, como o histograma, o Método de Dependência Espacial de Níveis de Cinza, o Método de Diferença de Níveis de Cinza e o Método de Comprimento de Primitivas de Níveis de Cinza. O Grupo II também trata da textura dos nódulos, mas utiliza quatro funções geoestatísticas denominadas semivariograma, semimadograma, covariograma e correlograma. O Grupo III descreve apenas medidas baseadas na geometria do nódulo, como a convexidade, a esfericidade e medidas baseadas na curvatura. Por fim, o Grupo IV analisa os métodos do coeficiente de Gini e do esqueleto dos nódulos, que levam em consideração tanto a geometria quanto a textura do nódulo.

Foi analisada uma amostra com 36 nódulos, sendo 29 benignos e 7 malignos, e os resultados preliminares são promissores na caracterização dos nódulos pulmonares. A maioria dos grupos de métodos propostos tem o valor da área sobre a curva ROC acima de 0.800, utilizando a Análise Discriminante Linear de Fisher e a Rede Neural Perceptron de Múltiplas Camadas. Isto significa que os métodos propostos possuem grande potencial na discriminação e classificação dos Nódulos Pulmonares Solitários.

Palavras-chave

Diagnóstico de Nódulo Pulmonar Solitário; Textura; Geometria; Análise Discriminante Linear de Fisher; Redes Neurais Perceptron de Múltiplas Camadas.

Abstract

Silva, Aristófanés C.; Gattass, Marcelo; Carvalho, Paulo Cezar Pinto. **Algorithms for Assisted Diagnosis of Solitary Lung Nodules in Computerized Tomography Images**. Rio de Janeiro, 2004. 140p. PhD. Thesis — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

The present work seeks to develop a computational tool to suggest about the malignancy or benignity of Solitary Lung Nodules by the analysis of texture and geometry measures obtained from computerized tomography images.

Four groups of methods are proposed with the purpose of suggesting the diagnosis for such nodule. The groups of methods are divided according to their common characteristics. Group I includes methods based on texture adapted for 3D, such as the histogram, the Spatial Gray Level Dependence Method, the Gray Level Difference Method and Gray Level Run Length Matrices. Group II also deals with the texture of nodules, but uses four statistical functions denominated semivariogram, semivariogram, covariogram and correlogram. Group III describes measures based only on the geometry of the nodule, such as convexity, sphericity, and measures based on the curvature. Finally, Group IV analyzes the Gini coefficient and nodule skeleton methods, which take into account both the nodule's geometry and its texture.

A sample with 36 nodules, 29 benign and 7 malignant, was analyzed and the preliminary results of this approach are very promising in characterizing lung nodules. Most groups of proposed methods have the area under the ROC curve value above 0.800, using Fisher's Linear Discriminant Analysis and Multilayer Perceptron Neural Networks. This means that the proposed methods have great potential in the discrimination and classification of Solitary Lung Nodules.

Keywords

Diagnosis of Solitary Lung Nodule; Texture; Geometry; Fisher's Linear Discriminant Analysis; Multilayer Perceptron Neural Networks.

Sumário

1	Introdução	16
1.1	Motivação	18
1.2	Objetivos	19
1.3	Contribuição	20
1.4	Trabalhos Relacionados	21
1.5	Estrutura da Tese	27
2	Conceitos Básicos	29
2.1	Imagem Médica	29
2.2	Nódulo Pulmonar Solitário	38
2.3	Técnicas para Analisar, Discriminar e Classificar	42
2.4	Validação do Modelo	55
2.5	Curva ROC (<i>Receiver Operating Characteristic</i>)	55
2.6	Resumo	62
3	Métodos Computacionais para Análise do NPS	64
3.1	Análise do Nódulo Pulmonar Baseada em Textura (Métodos Clássicos) – Grupo I	65
3.2	Análise do Nódulo Pulmonar Baseada em Textura (Funções Geoestatísticas) – Grupo II	73
3.3	Análise do Nódulo Pulmonar Baseada na Geometria – Grupo III	77
3.4	Análise do Nódulo Pulmonar Baseada em Textura e na Geometria – Grupo IV	81
3.5	Resumo	87
4	Testes e Resultados	91
4.1	Metodologia	91
4.2	Exemplo da Aplicação dos Métodos Estudados	96
4.3	Classificação	105
4.4	Conclusão	115
4.5	Resumo	117
5	Conclusão	118
5.1	Trabalhos Futuros	120
A	Sistema de Análise de Nódulo Pulmonar – SANP	134
A.1	Visualização das fatias	135
A.2	Segmentação	136
A.3	Visualização 3D	137
A.4	Tempo de Dobra	138
A.5	Estruturas do Nódulo	139

Lista de Figuras

1.1	Nódulos Pulmonares Solitários.	19
1.2	Esquema dos objetivos deste trabalho.	20
2.1	Funcionamento de uma TC (Fonte: http://www.geocities.com/siumingrd/CT).	32
2.2	Tomografia computadorizada do tórax.	32
2.3	Estrutura da imagem no formato DICOM.	36
2.4	Espaço amostral do <i>voxel</i> [56].	37
2.5	Exemplo de interpolação [56].	37
2.6	Interpolação linear [56].	38
2.7	Provável diagnóstico do NPS em relação ao coeficiente de atenuação.	41
2.8	Provável diagnóstico do NPS em relação à forma.	42
2.9	Algoritmo para diagnóstico dos NPS.	43
2.10	Procedimento de seleção de variáveis <i>passo a passo</i> .	49
2.11	Modelo de uma rede MLP (3-4-1).	50
2.12	Exemplo da técnica <i>deixa um de fora</i> .	56
2.13	Ponto de corte.	59
2.14	Relação entre a curva ROC e pontos de corte.	59
2.15	Curva ROC e a qualidade no diagnóstico.	60
3.1	Exemplo de SGLDM em Imagem 2D.	67
3.2	Algoritmo do SGLDM (ou GLDM).	68
3.3	Exemplo de GLDM em Imagem 2D.	70
3.4	Exemplo de GLRLM em Imagem 2D.	72
3.5	Algoritmo do GLRLM.	72
3.6	Semivariograma e suas características: patamar, alcance e efeito pepita.	74
3.7	Definição do vetor distância.	75
3.8	Parâmetros utilizados para cálculo do semivariograma (2D).	76
3.9	(a) Aplicação do Marching Cubes. (b) Aplicação da técnica Laplaciana.	78
3.10	Tipos de superfície baseados na curvatura.	81
3.11	Exemplo da curva de Lorenz e coeficiente de Gini ($A/(A + B)$).	83
3.12	Codificação de <i>voxels</i> de uma fatia do NPS (2D).	85
4.1	Exemplos de NPS benignos ((a) e (b)) e malignos ((c) e (d)).	96
4.2	Exemplos de NPS benignos ((a), (b), (c) e (d)) e malignos ((e), (f), (g)).	97
4.3	Histograma aplicado no exemplo da Figura 4.1.	98
4.4	SGLDM aplicado no exemplo da Figura 4.1.	98
4.5	GLDM aplicado no exemplo da Figura 4.1.	99
4.6	GLRLM aplicado no exemplo da Figura 4.1.	100
4.7	Semivariograma aplicado no exemplo da Figura 4.1.	101

4.8	Semivariograma aplicado ao nódulo representado pela Figura 4.1(a).	101
4.9	Semivariograma aplicado ao nódulo representado pela Figura 4.1(d).	102
4.10	Medidas de Geometria aplicadas no exemplo da Figura 4.2.	103
4.11	Coeficiente de Gini aplicado ao nódulo representado pela Figura 4.1(a).	104
4.12	Curva de Lorenz e coeficiente de Gini aplicados no exemplo da Figura 4.1.	104
4.13	Aplicação do algoritmo de Zhou e Toga baseada nos nódulos da Figura 4.2(a), (b), (c) e (d).	105
4.14	Distribuição dos nódulos do Grupo I usando ALDF.	106
4.15	Distribuição dos nódulos do Grupo I usando MLP.	107
4.16	Curva ROC do Grupo I utilizando a ALDF e MLP.	108
4.17	Distribuição dos nódulos do Grupo II usando ALDF.	109
4.18	Distribuição dos nódulos do Grupo II usando MLP.	109
4.19	Curva ROC do Grupo II utilizando a ALDF e MLP.	109
4.20	Distribuição dos nódulos do Grupo III usando ALDF.	110
4.21	Distribuição dos nódulos do Grupo III usando MLP.	111
4.22	Curva ROC do Grupo III utilizando a ALDF e MLP.	111
4.23	Distribuição dos nódulos do Grupo IV usando ALDF.	112
4.24	Distribuição dos nódulos do Grupo IV usando MLP.	113
4.25	Curva ROC do Grupo IV utilizando a ALDF e MLP.	113
4.26	Distribuição dos nódulos dos Grupos combinados usando ALDF.	114
4.27	Distribuição dos nódulos dos Grupos combinados usando MLP.	115
4.28	Curva ROC dos Grupos combinados utilizando a ALDF e MLP.	115
4.29	Comparação das áreas das curvas ROC de todos os grupos analisados, utilizando a ALDF e MLP.	116
4.30	Nódulo com difícil diagnóstico.	116
A.1	Interface do SANP.	135
A.2	Galeria de fatias.	136
A.3	Janela/Nível de contraste e exemplos.	136
A.4	Zoom de detalhe.	137
A.5	Barreira ao redor do nódulo.	137
A.6	Barreira aberta.	138
A.7	Processo de segmentação.	138
A.8	Janela de visualização 3D - Marching Cubes.	139
A.9	Tempo de dobra.	139
A.10	Nódulo com estruturas internas visualizadas e histograma das estruturas do nódulo.	140

Lista de Tabelas

2.1	Cálculo da variância e do valor de F.	48
2.2	Termos similares entre MLP e ALDF.	55
2.3	Relação entre o resultado de um teste diagnóstico e o verdadeiro diagnóstico	56
2.4	Cálculo da especificidade e sensibilidade para uma variável dicotômica	58
2.5	Qualidade do diagnóstico em relação à área da curva ROC.	60
3.1	Grupos de medidas estudados e suas características.	64
3.2	Resumo dos métodos e das medidas propostas.	89
4.1	% de acertos usando a ALDF e MLP para o Grupo I.	106
4.2	% de acertos usando a ALDF e MLP para o Grupo II.	108
4.3	% de acertos usando a ALDF e MLP para o Grupo III.	110
4.4	% de acertos usando a ALDF e MLP para o Grupo IV.	112
4.5	% de acertos usando a ALDF e MLP para os Grupos combinados.	114

Lista de Abreviaturas

ALDF	Análise discriminante linear de Fisher
AUC	Área sob a curva ROC
CS	Comprimento dos segmentos, medida baseada no esqueleto
CNT	Contraste, medida baseada no histograma
CONd	Contraste, medida baseado no GLDM
CONg	Contraste, medida baseado no SGLDM
CON	Convexidade, medida baseada na geometria
CORd	Correlação, medida baseada no GLDM
CORg	Correlação, medida baseada no SGLDM
CPI	Média de <i>curvedness</i> do tipo <i>pit</i> , medida baseada na geometria
CPK	Média de <i>curvedness</i> do tipo <i>peak</i> , medida baseada na geometria
CSR	Média de <i>curvedness</i> do tipo <i>saddle ridge</i> , medida baseada na geometria
CSV	Média de <i>curvedness</i> do tipo <i>saddle valley</i> , medida baseada na geometria
CUR	Curtose, medida baseada no histograma
CV	Coefficiente de variação, medida baseada no esqueleto
DICOM	<i>Digital image and communications in medicine</i>
DVP	Desvio padrão, medida baseada no histograma
ENE	Energia, medida baseada no histograma
ENT	Entropia, medida baseada no histograma
ENTd	Entropia, medida baseada no GLDM
ENTg	Entropia, medida baseada no SGLDM
ESF	Esfericidade, medida baseada na geometria
FV	Fração do volume, medida baseada na geometria
GLD	Uniformidade do nível de cinza, medida baseada no GLRLM
GLDM	Método de diferença de nível de cinza
GLRLM	Método de comprimento de primitiva de nível de cinza
HOMOd	Homogeneidade baseada no GLDM
HOMOg	Homogeneidade baseada no SGLDM
IAPI	Índice da área do tipo <i>pit</i> , medida baseada na geometria
IAPK	Índice da área do tipo <i>peak</i> , medida baseada na geometria

IASR	Índice da área do tipo <i>saddle ridge</i> , medida baseada na geometria
IASV	Índice da área do tipo <i>saddle valley</i> , medida baseada na geometria
ICE	Índice de curvatura extrínseca, medida baseada na geometria
ICI	Índice de curvatura intrínseca, medida baseada na geometria
LRE	Ênfase de primitivas longas, medida baseada no GLRLM
MED	Média, medida baseada no histograma
MLP	Rede neural perceptron múltiplas camadas
NCMAI	Número de coeficientes de atenuação ≥ 200 UH, medida baseada no histograma
NPS	Nódulo pulmonar solitário
NR	Número de ramificações, medida baseada no esqueleto
NS	Número de segmentos, medida baseada no esqueleto
NSVFC	Taxa entre o número de segmentos e o fecho convexo, medida baseada no esqueleto
QPI	Quantidade de tipo <i>pit</i> , medida baseada na geometria
QPK	Quantidade de tipo <i>peak</i> , medida baseada na geometria
QSR	Quantidade de tipo <i>saddle ridge</i> , medida baseada na geometria
QSV	Quantidade de tipo <i>saddle valley</i> , medida baseada na geometria
RLD	Uniformidade do comprimento de primitivas, medida baseada no GLRLM
ROC	<i>Receiver operator characteristic curve</i>
RP	Porcentagem de primitivas, medida baseada no GLRLM
SANP	Sistema de análise de nódulo pulmonar
SE	Erro padrão baseado na área da curva ROC
SGLDM	Método de dependência espacial de nível de cinza
SIM	Simetria, medida baseada no histograma
SMA _d	Segundo momento angular, medida baseada no GLDM
SMA _g	Segundo momento angular, medida baseada no SGLDM
SPSS	<i>Statistical package for the social sciences</i>
SRE	Ênfase de primitivas curtas, medida baseada no GLRLM
TC	Tomografia Computadorizada
UH	Unidade de Hounsfield
VAR _d	Variância, medida baseada no GLDM
VAR _g	Variância, medida baseada no SGLDM
VFC	Volume do fecho convexo, medida baseada no esqueleto

”É melhor tentar e falhar, que preocupar-se a ver a vida passar. É melhor tentar, ainda que em vão, que sentar-se fazendo nada até o final. Eu prefiro na chuva caminhar, que em dias tristes em casa me esconder. Prefiro ser feliz, embora louco, que em conformidade viver.”

Martin Luther King Jr., .