

Universidade Federal do Maranhão
Centro de Ciências Exatas e Tecnologia
Programa de Pós-Graduação em Engenharia de Eletricidade

STELMO MAGALHÃES BARROS NETTO

Métodos Computacionais para Identificação,
Quantificação e Análise de Mudanças no
Tecido da Lesão Pulmonar Através de
Imagens de Tomografia Computadorizada

São Luís

2016

STELMO MAGALHÃES BARROS NETTO

**Métodos Computacionais para Identificação,
Quantificação e Análise de Mudanças no Tecido da
Lesão Pulmonar Através de Imagens de Tomografia
Computadorizada**

Versão original

Tese apresentada ao Programa de Pós-Graduação em Engenharia de Eletricidade da UFMA como parte dos requisitos necessários para obtenção do grau de Doutor em Engenharia Elétrica.

Área de concentração: Ciência da Computação

Orientador: Prof. Dr. Aristófanés Corrêa Silva

Coorientador: Prof. Dr. Anselmo Cardoso de Paiva

São Luís

2016

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Núcleo Integrado de Bibliotecas/UFMA

Barros Netto, Stelmo Magalhães.
Métodos Computacionais para Identificação,
Quantificação e Análise de Mudanças no Tecido da Lesão
Pulmonar Através de Imagens de Tomografia Computadorizada
/ Stelmo Magalhães Barros Netto. - 2016.
182 f.

Coorientador(a): Anselmo Cardoso de Paiva.
Orientador(a): Aristófanés Corrêa Silva.
Tese (Doutorado) - Programa de Pós-graduação em
Engenharia de Eletricidade/ccet, Universidade Federal do
Maranhão, São Luís - MA, 2016.

1. Avaliação e análise temporal. 2. Detecção de
mudanças de tecidos. 3. Imagens médicas. 4. Lesão
pulmonar. I. Corrêa Silva, Aristófanés. II. Paiva,
Anselmo Cardoso de. III. Título.

Texto de Defesa de Tese de autoria de Stelmo Magalhães Barros Netto, sob o título “**Métodos Computacionais para Identificação, Quantificação e Análise de Mudanças no Tecido da Lesão Pulmonar Através de Imagens de Tomografia Computadorizada**”, apresentado ao Programa de Pós-Graduação em Engenharia de Eletricidade da UFMA, como parte dos requisitos para obtenção do título de Doutor em Engenharia Elétrica, na área de concentração Ciência da Computação, aprovado em 05 de outubro de 2016 pela comissão examinadora constituída pelos doutores:

Prof. Dr. Aristófanés Corrêa Silva
Orientador

Prof. Dr. Anselmo Cardoso de Paiva
Coorientador

Prof. Dr. João Viana da Fonseca Neto
Examinador Interno

Prof. Dr. Vicente Leonardo Paucar
Examinador Interno

Prof. Dr. Rodolfo Acatauassú Nunes
Examinador Externo - Universidade Estadual do Rio de Janeiro - UERJ

Prof. Dr. Hélio Côrtes Vieira Lopes
Examinador Externo - Pontifícia Universidade Católica do Rio de Janeiro - PUC-Rio

Agradecimentos

Agradeço a Deus pela vida que me foi dada.

Aos meus familiares por todo o suporte e apoio que me deram nessa longa jornada.

Aos meus orientadores Aristófanês Corrêa Silva e Anselmo Cardoso de Paiva pela oportunidade, pelo apoio e paciência ao me guiar ao longo dos quatro anos de desenvolvimento deste trabalho de Tese.

À Universidade Federal do Maranhão por financiar a minha capacitação, em especial a todos os professores pertencentes ao Curso de Engenharia de Alimentos, que permitiram a minha saída para a capacitação.

Aos professores Daniel Duarte Costa e Heron Silva Oliveira que me substituíram durante o doutorado.

A todos da Pós-graduação de Engenharia de Eletricidade, em especial ao professor Zair Abdelouahab, que sempre estava pronto para ajudar em qualquer problema.

A todos do Núcleo de Computação Aplicada que estavam comigo nessa longa jornada.

Ao Hospital Universitário Pedro Ernesto por disponibilizar a base de nódulos pulmonares utilizada neste trabalho.

A todos que direta e indiretamente contribuíram para o desenvolvimento deste trabalho.

Resumo

Barros Netto, Stelmo Magalhães. “Métodos Computacionais para Identificação, Quantificação e Análise de Mudanças no Tecido da Lesão Pulmonar Através de Imagens de Tomografia Computadorizada”. 2016. 164 f. Tese (Doutorado em Engenharia Elétrica) – Pós-Graduação em Engenharia de Eletricidade da UFMA, Centro de Ciências Exatas e Tecnologia, Universidade Federal do Maranhão, São Luís, Maranhão, 2016.

O câncer de pulmão é um dos tipos de câncer de maior incidência no mundo. A avaliação temporal aparece como ferramenta bastante útil quando se deseja analisar uma lesão. A análise pode ocorrer quando uma lesão maligna está em tratamento ou quando surgem lesões indeterminadas, mas essas são provavelmente benignas. O objetivo deste trabalho é desenvolver métodos computacionais para detectar, quantificar e analisar mudanças de densidade locais e globais das lesões pulmonares ao longo do tempo. Desta forma, foram desenvolvidos quatro conjuntos de métodos para realização da tarefa de detectar mudanças de densidade em lesões pulmonares. O primeiro conjunto identifica mudanças de densidade locais e foi denominado de métodos baseados em *voxel*. O segundo conjunto é composto da divergência de Jensen e do teste de hipótese com abordagens locais e globais. Com o mesmo propósito de detectar mudanças de densidade locais em lesões pulmonares, o terceiro conjunto possui um único método, a análise de componentes principais. O último conjunto também possui um único método, denominado de *quality threshold* modificado e identifica as mudanças locais de densidade. Para cumprir o objetivo deste trabalho, propõe-se uma metodologia composta de cinco etapas. A primeira etapa consiste na aquisição das imagens da lesão em diversos instantes. Duas bases de lesões foram utilizadas e dois modelos de lesões foram propostos para avaliação dos métodos. A primeira base possui 24 lesões em tratamento (base pública) e a segunda possui 13 nódulos benignos (base privada) em acompanhamento. A segunda etapa corresponde ao registro rígido das imagens da lesão. A próxima etapa é a aplicação dos quatro conjuntos de métodos propostos. Como resultado, o segundo conjunto de métodos detectou mais mudanças de densidade que o quarto conjunto, que por sua vez, este último detectou mais regiões que o primeiro conjunto e este mais que o terceiro conjunto, para a base pública de lesões. Em relação a base privada, o quarto conjunto de métodos detectou mais regiões de mudança de densidade que o primeiro conjunto. O terceiro conjunto detectou menos regiões de mudança quando comparado ao primeiro conjunto e o segundo conjunto teve o menor número de regiões detectadas. Em adição às mudanças de densidade encontradas, o modelo de classificação proposto com medidas clássicas de textura para predição do diagnóstico da lesão teve acurácia acima de 98%. Os resultados encontrados indicam que existem mudanças de densidade em ambas as bases de lesões pulmonares. Entretanto, as mudanças detectadas por cada um dos métodos propostos possuem características de intensidade e localização diferentes em ambas as bases. Essa conclusão é motivada pela alta acurácia obtida em seu diagnóstico para as bases utilizadas.

Palavras-chaves: Imagens médicas. Lesão pulmonar. Avaliação e análise temporal. Detecção de mudanças de tecidos.

Abstract

Barros Netto, Stelmo Magalhães. "Computational methods to identify, quantify and analyze tissue changes in lung lesion through computed tomography images". 2016. 164 p. Thesis (Doctorate in Electrical Engineering) – graduate program in Electrical Engineering, Centro de Ciências Exatas e Tecnologia, Federal University of Maranhão, São Luís, Maranhão, 2016.

Lung cancer is one of the most common types of cancer around the world. Temporal evaluation has become a very useful tool when to whoever needs to analyze a lung lesion. The analysis occurs when a malignant lesion is under treatment or when there are indeterminate lesions, but they are probably benign. The objective from this work is to develop computational methods to detect, quantify and analyze local and global density changes of pulmonary lesions over time. Thus, it were developed four groups of methods to perform this task. The first identifies local density changes and it has been denominated voxel-based. The second one is composed of the Jensen divergence and the hypothesis test with global and local approaches. Similarly, the third group has only one method, the principal component analysis. The last group has one method, it has been denominated modified quality threshold, and identifies the local density changes. In order to reach the objectives, it was proposed a methodology composed of five steps: The first step consists in image acquisition of the lesion at various instants. Two image databases were acquired and two models of lesions were created to evaluate the methods. The first database has 24 lesions under treatment (public database) and the second has 13 benign nodules (private database) in monitoring. The second step refers to rigid registration of the lesion images. The next step is to apply the proposed four groups of methods. As a result, the second group of methods detected more density changes than the fourth group, which in turn, this latter detected more regions than the first group and this more than the third group, for the public database. For the private database, the fourth group of density change methods detected more regions than the first group. The third group detected few regions of changes when compared to the first group and the second group had the lowest number of detected regions. In addition to the density changes found, the proposed classification model with texture features had accuracy above 98% in the diagnosis prediction. The results state that there are changes in both databases. However, the detected changes for each group of methods have different intensity and location to the databases. This conclusion is based from high accuracy that was obtained from the prediction of the lesion diagnosis from both databases.

Keywords: Medical image. Pulmonary lesion. Temporal analysis and evaluation. Detection of changes in tissues.

Lista de figuras

Figura 3.1 – Processo de varredura do cabeçote e detectores	15
Figura 3.2 – Partes de um tomógrafo computadorizado	16
Figura 3.3 – Imagem 3D do pulmão.	17
Figura 3.4 – Características da imagem 2D	17
Figura 3.5 – Estruturas do pulmão.	18
Figura 3.6 – Pulmão com a massa destacada pelo contorno.	19
Figura 3.7 – Lesão pulmonar envolta de um contorno em amarelo.	21
Figura 3.8 – Exemplo de lesão pulmonar e fundo.	21
Figura 3.9 – Componentes do registro.	23
Figura 3.10–Comportamento do gráfico do erro médio quadrático.	26
Figura 3.11–Vetores de Suporte destacados por círculos.	38
Figura 4.1 – Imagens dos resultados do método da diferença.	43
Figura 4.2 – Histograma da imagem resultante da aplicação do método da diferença absoluta.	44
Figura 4.3 – Mudanças de densidade detectadas pelo método da diferença da Figura 4.1. Em (a), têm-se as mudanças lesão-fundo e (b), têm-se as mudanças lesão-lesão.	45
Figura 4.4 – Imagens resultantes da aplicação do método da razão de imagens.	47
Figura 4.5 – Imagem ilustrativa da relação linear entre os <i>voxels</i> da imagem.	48
Figura 4.6 – Mudanças de densidade detectadas pelo método da regressão da Figura 4.5. Em (a), têm-se as mudanças lesão-fundo e (b), têm-se as mudanças lesão-lesão.	49
Figura 4.7 – Imagem <i>background</i> formada pela mediana das imagens.	49
Figura 4.8 – Imagens resultantes da aplicação do operador XOR <i>fuzzy</i>	51
Figura 4.9 – Mudanças de densidade detectadas pelo método do operador XOR <i>fuzzy</i> da Figura 4.8. Em (a), têm-se as mudanças lesão-fundo e (b), têm-se as mudanças lesão-lesão.	51
Figura 4.10–Esquema geral de aplicação dos métodos baseados em <i>voxel</i>	52
Figura 4.11–Modelos que simulam a lesão em dois diferentes instantes. Em (a), tem-se a lesão em instante t_1 e em (b), tem-se a lesão em um instante posterior t_2	53

Figura 4.12–Resultado do método da diferença para os modelos da Figura 4.11. Em (a), tem-se a imagem da diferença e em (b), a imagem binária de mudança.	58
Figura 4.13–Resultado do método da razão para os modelos da Figura 4.11. Em (a), tem-se a imagem da razão e em (b), a imagem binária de mudança.	58
Figura 4.14–Resultado do método de regressão para os modelos da Figura 4.11. Em (a), tem-se a imagem de regressão e em (b), a imagem binária de mudança.	59
Figura 4.15–Resultado do método da mediana para os modelos da Figura 4.11. Em (a), tem-se a imagem da mediana, em (b), a imagem da diferença e em (c), a imagem binária de mudança.	60
Figura 4.16–Resultado do método XOR <i>fuzzy</i> para os modelos da Figura 4.11. Em (a), tem-se a imagem XOR <i>fuzzy</i> e em (b), a imagem binária de mudança.	60
Figura 4.17–Resultado do método diferença entre imagens para o Caso 10.	65
Figura 4.18–Resultado do método diferença entre imagens para o Caso 28.	66
Figura 4.19–Resultado do método razão de imagens para o Caso 10.	67
Figura 4.20–Resultado do método razão de imagens para o Caso 28.	68
Figura 4.21–Resultado do método da imagem mediana para o Caso 10.	69
Figura 4.22–Resultado do método da imagem mediana para o Caso 28.	70
Figura 4.23–Resultado do método do operador XOR <i>fuzzy</i> para o Caso 10.	71
Figura 4.24–Resultado do método do operador XOR <i>fuzzy</i> para o Caso 28.	72
Figura 4.25–Resultado do método da regressão de imagens para o Caso 10.	73
Figura 4.26–Resultado do método da regressão de imagens para o Caso 28.	74
Figura 4.27–Resultado do método da diferença para o Caso 4.	76
Figura 4.28–Resultado do método da diferença para o Caso 12.	77
Figura 4.29–Resultado do método da razão para o Caso 4.	78
Figura 4.30–Resultado do método da diferença para o Caso 12.	78
Figura 4.31–Resultado do método da regressão para o Caso 4.	79
Figura 4.32–Resultado do método da regressão para o Caso 12.	80
Figura 4.33–Resultado do método da mediana para o Caso 4.	81
Figura 4.34–Resultado do método da mediana para o Caso 12.	81
Figura 4.35–Resultado do método do operação XOR <i>fuzzy</i> para o Caso 4.	82
Figura 4.36–Resultado do método do operação XOR <i>fuzzy</i> para o Caso 12.	83
Figura 4.37–Regiões de extração das características de textura.	84
Figura 5.1 – Método <i>bootstrap</i>	89

Figura 5.2 – Formação da amostra <i>bootstrap</i> pelo método MBB.	89
Figura 5.3 – Intervalo de variação da média e seus limites de confiança.	91
Figura 5.4 – Limite de confiança.	91
Figura 5.5 – Exemplo de duas distribuições T e Z	92
Figura 5.6 – Fluxo das etapas da estimação global das incertezas.	94
Figura 5.7 – Fluxo das etapas para comparação global da lesão.	95
Figura 5.8 – Fluxo das etapas da estimação local das incertezas.	96
Figura 5.9 – Fluxo das etapas para comparação local da lesão.	97
Figura 5.10–Resultado do teste de hipótese local para o Caso 3.	101
Figura 5.11–Resultado do teste de hipótese local para o Caso 42.	102
Figura 5.12–Resultado dos métodos baseados na incerteza para os modelos de lesões da Figura 4.11.	103
Figura 5.13–Histogramas do Caso 10 para a lesão em t_a (a) e em t_b (b).	104
Figura 5.14–Resultado do teste de hipótese local para o Caso 10.	105
Figura 5.15–Resultado do teste de hipótese local para o Caso 28.	105
Figura 5.16–Histogramas do Caso 28 para a lesão em t_a (a) e em t_b (b).	106
Figura 5.17–Histogramas para a lesão do Caso 42 nos dois instantes analisados.	107
Figura 5.18–Resultado do teste de hipótese local para o Caso 4.	108
Figura 5.19–Resultado do teste de hipótese local para o Caso 12.	108
Figura 5.20–Histogramas para a lesão do Caso 4 nos dois instantes analisados.	109
Figura 5.21–Histogramas para a lesão do Caso 12 nos dois instantes analisados.	109
Figura 5.22–Resultado do teste de hipótese local para o Caso 14.	110
Figura 5.23–Histogramas para a lesão do Caso 14 nos dois instantes analisados.	110
Figura 5.24–Regiões de extração das características de textura.	111
Figura 6.1 – Conjunto de pontos no R^2 onde a média é extraída.	115
Figura 6.2 – Extração das componentes principais.	116
Figura 6.3 – Exemplo de duas imagens da lesão.	116
Figura 6.4 – Imagem resultante da primeira aplicação do ACP.	117
Figura 6.5 – Exemplo da lesão pulmonar, seu contorno e fundo.	117
Figura 6.6 – A lesão com seus contornos em diferentes instantes.	118
Figura 6.7 – Resultado final da aplicação da ACP.	118
Figura 6.8 – Resultado final da ACP após o agrupamento.	119
Figura 6.9 – Resultado do método ACP para os modelos de lesões.	120

Figura 6.10–Resultado final da aplicação da ACP para o Caso 10.	121
Figura 6.11–Resultado final da aplicação da ACP para o Caso 28.	123
Figura 6.12–Resultado final da aplicação da ACP para o Caso 4.	124
Figura 6.13–Resultado final da aplicação do ACP para o Caso 12.	125
Figura 6.14–Regiões da extração das características de textura.	126
Figura 7.1 – Resultado do agrupamento com o <i>jump method</i> sem modificação. . . .	130
Figura 7.2 – Curva EMQ x k . Em verde, está a maior diferença que resultou no melhor valor k	131
Figura 7.3 – Resultado da aplicação do QT modificado para os Casos 10 e 28. . . .	136
Figura 7.4 – (a) e (b) são os grupos para o Caso 42 nos diferentes instantes. Em (c), têm-se as regiões de mudança.	137
Figura 7.5 – Resultado da aplicação do QT modificado para os Casos 4 e 12.	138
Figura 7.6 – Resultado da aplicação do QT modificado para o Caso 14.	140
Figura 7.7 – Regiões da extração das características.	141
Figura 8.1 – Curva do P_m em função da variação volumétrica para a base privada. .	148
Figura 8.2 – Curva do P_m em função da variação volumétrica para a base pública. .	149

Lista de tabelas

Tabela 2.1 – Alguns trabalhos relacionados descritos em termos de suas técnicas, base de imagens e resultados.	12
Tabela 4.1 – Associação de cada caso analisado com o par de instantes da lesão para a base privada.	55
Tabela 4.2 – Associação de cada caso analisado com o par de instantes da lesão para a base pública.	56
Tabela 4.3 – Percentuais de mudança da base pública para os métodos baseados em <i>voxel</i>	63
Tabela 4.4 – Percentuais de mudança base privada para os métodos baseados em <i>voxel</i>	75
Tabela 4.5 – Métricas de avaliação da classificação com MVS para os métodos baseados em <i>voxel</i>	85
Tabela 5.1 – Resultado da aplicação do teste de hipótese global.	99
Tabela 5.2 – Resultado da aplicação do teste de hipótese global.	100
Tabela 5.3 – Métricas de avaliação dos modelos de classificação com MVS para o método do teste de hipótese.	112
Tabela 6.1 – Percentuais de mudança obtidos do método ACP para a base pública.	122
Tabela 6.2 – Percentuais de mudança obtidos do método ACP para a base privada.	124
Tabela 6.3 – Métricas de avaliação da classificação com MVS para o método ACP.	126
Tabela 7.1 – Conjunto de valores das medidas que quantificam as mudanças de densidade para valor do centro do grupo do Caso 10.	134
Tabela 7.2 – Conjunto de valores das medidas que quantificam as mudanças de densidade para valor do centro do grupo do Caso 28.	135
Tabela 7.3 – Conjunto de valores das medidas que quantificam as mudanças de densidade para valor do centro do grupo do Caso 42.	136
Tabela 7.4 – Conjunto de valores das medidas que quantificam as mudanças de densidade para valor do centro do grupo do Caso 4.	138
Tabela 7.5 – Conjunto de valores das medidas que quantificam as mudanças de densidade para valor do centro do grupo do Caso 12.	139
Tabela 7.6 – Conjunto de valores das medidas que quantificam as mudanças de densidade para valor do centro do grupo do Caso 14.	140
Tabela 7.7 – Métricas de avaliação da classificação com MVS para o método QT.	141

Tabela 8.1 – Percentuais de mudança de todos os métodos para a base privada de lesões.	145
Tabela 8.2 – Percentuais de mudança de todos os métodos para a base pública de lesões.	146
Tabela 8.3 – Métodos mais e menos sensíveis às variações de densidade para a base privada.	148
Tabela 8.4 – Métodos ordenados pelo número de ocorrências dos casos da base privada.	148
Tabela 8.5 – Métodos mais e menos sensíveis às variações de densidade para a base pública.	150
Tabela 8.6 – Métodos ordenados pelo número de ocorrências dos casos da base pública.	151
Tabela 8.7 – Comparação dos resultados de classificação de todos os métodos desenvolvidos.	152

Lista de abreviaturas e siglas

ac	Acurácia
ACP	Análise de componentes principais
esp	Especificidade
DICOM	<i>Digital Imaging and Communications in Medicine</i>
fp	Falso Positivo
fn	Falso Negativo
HU	<i>Hounsfield Unit</i>
INCA	Instituto Nacional de Câncer
MBB	<i>Moving Block Bootstrap</i>
MVS	Máquina de Vetores de Suporte
NCI	<i>National Cancer Institute</i>
P_m	Percentual de Mudança
PRF	<i>Prevent Cancer Foundation</i>
RBF	<i>Radial Basis Function</i>
ROC	<i>Receiver Operating Characteristic</i>
RL	<i>Run-Length</i>
sen	Sensibilidade
TC	Tomografia Computadorizada
VVR	Variação Volumétrica Relativa
vp	Verdadeiro Positivo
vn	Verdadeiro Negativo

Sumário

1	Introdução	1
1.1	Objetivos e justificativa	2
1.2	Contribuições	3
1.3	Organização do Trabalho	4
2	Trabalhos Relacionados	5
3	Fundamentação	14
3.1	Tomografia Computadorizada	14
3.2	Imagem 3D	16
3.3	Lesões Pulmonares	18
3.3.1	Base de lesões pulmonares	19
3.3.1.1	Base privada	20
3.3.1.2	Base pública	20
3.4	Registro de Imagens	22
3.4.1	Registro Rígido de Imagens 3D	23
3.4.1.1	Transformação 3D	23
3.4.1.2	Interpolador	25
3.4.1.3	Métrica de semelhança	25
3.4.1.4	Otimizador	26
3.5	Características de textura	27
3.5.1	Matriz de cocorrência	28
3.5.1.1	Inércia	28
3.5.1.2	Energia	29
3.5.1.3	Entropia	29
3.5.1.4	Homogeneidade	29
3.5.1.5	Tom do grupo - <i>Cluster shade</i>	29
3.5.1.6	Proeminência - <i>Cluster prominence</i>	30
3.5.2	Matriz de comprimento de nível de cinza	30
3.5.2.1	Ênfase em primitivas curtas (<i>Short-run emphasis</i> - (SRE))	30
3.5.2.2	Ênfase em primitivas longas (<i>Long-run emphasis</i> - (LRE))	31

3.5.2.3	Não-uniformidade do nível de cinza (<i>Gray-level nonuniformity - (GLNU)</i>)	31
3.5.2.4	Não-uniformidade do comprimento da primitiva (<i>Run-Length Nonuniformity - (RLN)</i>)	31
3.5.2.5	Ênfase de primitiva de pequeno nível de cinza (<i>Low Grey Level Run Emphasis - (LGRE)</i>)	32
3.5.2.6	Ênfase de primitiva de grande nível de cinza (<i>High Grey Level Run Emphasis - (HGLRN)</i>)	32
3.5.2.7	Ênfase em primitivas curtas de pequeno nível de cinza (<i>Short Run, Low Grey Level Emphasis - (SRLGLE)</i>)	32
3.5.2.8	Ênfase em primitivas curtas de grande nível de cinza (<i>Short Run, High Grey Level Emphasis - (SRHGLE)</i>)	32
3.5.2.9	Ênfase em primitivas longas de pequeno nível de cinza (<i>Long Run, Low Grey Level Emphasis - (LRLGLE)</i>)	33
3.5.2.10	Ênfase em primitivas longas de grande nível de cinza (<i>Long Run, High Grey Level Emphasis - (LRHGLE)</i>)	33
3.5.3	Medidas estatísticas	33
3.6	Reconhecimento de padrões	34
3.6.1	Máquina de Vetores de Suporte	36
3.7	Validação do resultado da classificação	39
3.8	Medidas de Quantificação das Mudanças	40
4	Métodos baseados em <i>voxel</i>	42
4.1	Diferença entre imagens	42
4.2	Razão de imagens	45
4.3	Regressão de imagem	47
4.4	Imagem Mediana	48
4.5	Operador XOR <i>fuzzy</i>	50
4.6	Identificação das mudanças de densidade	51
4.7	Resultados	52
4.7.1	Base de modelos de lesões	53
4.7.2	Base de lesões	54
4.7.2.1	Base privada	54

4.7.2.2	Base pública	55
4.7.3	Registro de imagens	57
4.7.4	Estudo de casos com modelos de lesões	57
4.7.4.1	Diferença entre imagens	57
4.7.4.2	Razão de imagens	57
4.7.4.3	Regressão de imagem	59
4.7.4.4	Imagem mediana	59
4.7.4.5	Operador XOR <i>fuzzy</i>	59
4.7.5	Estudo de casos com a base pública de lesões	60
4.7.5.1	Diferença entre imagens	64
4.7.5.2	Razão de imagens	66
4.7.5.3	Imagem Mediana	68
4.7.5.4	Operador XOR <i>fuzzy</i>	70
4.7.5.5	Regressão de imagens	72
4.7.6	Estudo de casos com a base privada de lesões	74
4.7.6.1	Diferença entre imagens	75
4.7.6.2	Razão de imagens	77
4.7.6.3	Regressão de imagem	79
4.7.6.4	Imagem mediana	80
4.7.6.5	Operador XOR <i>fuzzy</i>	82
4.7.7	Classificação de lesões quanto ao diagnóstico	83
4.8	Discussão	84
4.9	Conclusão	86
5	Métodos de estimação de incerteza	87
5.1	Incerteza	87
5.2	Método <i>Bootstrap</i>	88
5.2.1	Método <i>moving block bootstrap</i>	88
5.3	Testes de hipóteses	90
5.4	Intervalo de Confiança	91
5.5	Divergência de Jensen	92
5.6	Identificação das mudanças de densidade	92
5.6.1	Análise global de mudanças de lesão	93

5.6.1.1	Divergência Global de Jensen	93
5.6.1.2	Divergência Global de Jensen do Histograma	93
5.6.1.3	Teste de hipótese global	93
5.6.2	Análise local de mudanças de lesão	95
5.6.2.1	Divergência Local de Jensen	95
5.6.2.2	Teste de hipótese local	96
5.7	Resultados	97
5.7.1	Estudo de casos	101
5.7.1.1	Estudo de casos com modelos de lesões	102
5.7.1.2	Estudo de casos com a base pública de lesões	103
5.7.1.3	Estudo de casos com a base privada de lesões	106
5.7.2	Classificação de lesões quanto ao diagnóstico	109
5.8	Discussão	111
5.9	Conclusão	112
6	Método baseado em análise de componentes principais	114
6.1	Análise de Componentes Principais	114
6.2	Identificação das mudanças de densidade	115
6.3	Resultados	119
6.3.1	Estudo de casos	119
6.3.1.1	Estudo de casos com modelos de lesões	119
6.3.1.2	Estudo de casos com a base pública de lesões	120
6.3.1.3	Estudo de casos com a base privada de lesões	123
6.3.2	Classificação de lesões quanto ao diagnóstico	125
6.4	Discussão	127
6.5	Conclusão	128
7	Método baseado em agrupamento	129
7.1	Método de agrupamento <i>quality threshold</i>	129
7.1.1	<i>Quality threshold modificado</i>	130
7.2	Identificação das mudanças de densidade	131
7.3	Resultados	132

7.3.1	Aplicação do método proposto	132
7.3.1.1	Estudo de casos com modelos de lesões	133
7.3.1.2	Estudo de casos com a base pública de lesões	133
7.3.1.3	Estudo de casos com a base privada de lesões	137
7.3.2	Classificação de lesões quanto ao diagnóstico	139
7.4	Discussão	142
7.5	Conclusão	142
8	Comparação dos métodos	144
8.1	Dificuldades de comparação com trabalhos relacionados	144
8.2	Comparação dos resultados obtidos	144
8.2.1	Quanto ao percentual de mudança	145
8.2.1.1	Base privada	147
8.2.1.2	Base pública	149
8.2.2	Quanto ao resultado da classificação	151
8.2.3	Conclusão	152
9	Conclusão geral e atividades futuras	154
9.1	Conclusão	154
9.2	Trabalhos futuros	155
9.3	Artigos	156
	Referências	157

1 Introdução

O câncer de pulmão é o segundo tipo de câncer mais comum entre homens e mulheres nos Estados Unidos, excluindo o câncer de pele ((NCI), 2011). A previsão para 2016 nos Estados Unidos é de aproximadamente 224.390 novos casos e 158.080 mortes para o câncer de pulmão, segundo a *American Cancer Society*. A última estimativa mundial em 2014 indica que o câncer de pulmão foi o mais incidente ((INCA), 2011). No Brasil, a situação não é muito diferente, pois o câncer de pulmão é o segundo tipo de câncer mais incidente entre os homens e o quarto mais incidente entre as mulheres. O último levantamento estatístico realizado no Brasil pelo Instituto Nacional de Câncer (INCA) apurou 24.490 mortes, das quais 14.811 eram homens e 9.675 eram mulheres. Além disso, a estimativa do número de novos casos é de aproximadamente mais de 28.000 em 2016 ((INCA), 2016). Devido a essas estatísticas e ao pobre prognóstico do câncer de pulmão, muitos trabalhos para detecção e avaliação precoce estão em desenvolvimento nos últimos anos. Dessa forma, muitos critérios de avaliação clínica e terapêutica foram criados, sempre com o objetivo de corrigir alguma incapacidade inerente ao critério ou ao tipo de lesão de estudo. Portanto, este trabalho visa contribuir com um conjunto de métodos que detecta as mudanças de densidade como forma de avaliar mudanças temporais.

A avaliação precoce ocorre quando o paciente se submete a programas de *screenings* antes dos sintomas ocorrerem. Como consequência da aplicação do programa de *screenings*, muitos casos de lesões indeterminadas ou provavelmente benignas são encontradas. Na ocorrência de lesões malignas, o tratamento químico ou radioterápico surge como alternativa à cirurgia. Em todos esses casos, repetidas avaliações das lesões pulmonares são necessárias para verificar o comportamento da lesão ao longo do tempo. Portanto, na tentativa de melhor caracterizar muitas das estruturas encontradas durante a aquisição das imagens, a lesão é continuamente avaliada e esse procedimento é chamado de avaliação temporal sobre a lesão, sendo que as formas mais comuns de se realizar essa avaliação é utilizando o diâmetro ou volume da lesão (TEAM, 2011; FIELD et al., 2013; LINNING et al., 2013; ZHANG et al., 2012).

A avaliação clínica ou terapêutica de lesões pulmonares em tomografia computadorizada (TC) é frequentemente conduzida usando o diâmetro e o volume (FIELD et al., 2013). Do mesmo modo, a avaliação unidirecional para lesões sólidas tem sido o principal indicador de progressão para lesões em tratamento. Essa avaliação foi denominada

de *Response evaluation criteria in solid tumors* (RECIST) (EISENHAUER et al., 2009) e foi criada para padronizar a avaliação das mudanças no tamanho da lesão. Por essa razão, o RECIST é utilizado por grupos farmacêuticos e especialistas com o objetivo de melhorar a resposta ao tratamento do câncer. Outro critério utilizado no passado e anterior ao RECIST é denominado de *World Health Organization* (WHO) (JAFJE, 2006). Ao contrário do RECIST, o WHO faz uso de medidas 2D sobre a lesão. Ambos os critérios citados são utilizados por especialistas e grupos farmacêuticos para avaliar a resposta de substâncias desenvolvidas para tratamento do câncer. Entretanto, esses critérios não produzem bons resultados em determinadas situações e em outros casos precisam de alguma adaptação (BYRNE; NOWAK, 2004; LENCIONI; LLOVET, 2010; EDELINE et al., 2012; FOURNIER et al., 2014; LIU et al., 2014).

O RECIST apresenta alguns problemas. Esse critério pode não reproduzir perfeitamente a complexidade da morfologia da lesão ou seu comportamento ao longo do tempo. Em muitos casos, ele não consegue revelar pequenas mudanças, provendo poucas contribuições para o monitoramento terapêutico. Em oposição a esse critério, outros trabalhos exploram a textura dos tecidos da lesão como forma de avaliar as mudanças no tecido, principalmente em casos onde existe a estabilidade no volume ou no diâmetro da lesão ao longo tempo (PURI et al., 2005; KIM et al., 2014; LAI et al., 2007; ZHANG et al., 2006; HARRISON et al., 2010; ZHANG, 2012). Em vista do que foi exposto, este trabalho visa revelar mudanças na lesão ocorridas ao longo do tempo, com base nas mudanças de densidade.

1.1 Objetivos e justificativa

O objetivo geral deste trabalho é desenvolver um conjunto de métodos para identificar, quantificar e analisar as variações de densidade (tecido) do nódulo ou massa pulmonar ao longo do tempo. Assim, os conjuntos de métodos propostos são aplicados quando existem lesões malignas em tratamento químico ou radioterápico e quando são diagnosticadas como indeterminadas em um exame inicial, mas são prováveis benignos.

Os métodos propostos fornecem suporte aos especialistas na análise temporal das lesões pulmonares. O especialista pode avaliar como as mudanças de densidade encontradas influenciam no comportamento da lesão ao longo do tempo, estando a mesma em tratamento ou não. Assim, supõe-se que, ao determinar as alterações de densidade e as

suas localizações nas lesões pulmonares, essas informações podem ser valiosas ferramentas ao especialista durante a sua análise. Pode-se destacar que as informações podem dar suporte no auxílio à tomada de decisões em relação à administração da droga, como, a dosagem; ou se a mesma é eficiente para o tratamento e etc. Do mesmo modo, quando nódulos pulmonares indeterminados são encontrados, mas há a chance de serem benignos no futuro, as informações de mudança de densidade também podem complementar a avaliação realizada pelo especialista. Desta forma, acredita-se que as informações de localização das mudanças de densidade, combinadas com uma medida de avaliação das mudanças, podem servir de parâmetros para dar suporte ao especialista.

Outro objetivo deste trabalho é quantificar as mudanças de densidade encontradas. Para quantificar essas mudanças, foram usadas medidas que visam contar os *voxels* que mais se modificaram no tempo. As principais medidas foram o percentual de mudança e o percentual de mudança intra-cluster. Do mesmo modo, para complementar o estudo sobre as mudanças de densidade, um outro conjunto de medidas foi utilizado para caracterizar as mudanças de densidade, as medidas descritivas da estatística e medidas de textura. Portanto, as medidas propostas foram importantes para avaliar o grau de mudança das densidades da lesão, e, conseqüentemente, avaliar as transformações da lesão ao longo do tempo. Ao mesmo tempo, o intuito da quantificação das mudanças de densidade é de contribuir com a determinação da caracterização em maligno ou benigno das lesões pulmonares em avaliação.

A determinação da natureza da lesão pulmonar foi realizada pela extração de características estatísticas e de textura obtidas nas regiões de mudança. Essa análise é apenas uma forma de caracterizar a lesão, caso o seu diagnóstico não seja definido inicialmente. Logo, a partir dessas características, é possível confirmar o seu diagnóstico avaliando as mudanças de densidade em futuras aquisições da mesma lesão.

1.2 Contribuições

Este trabalho apresenta as seguintes contribuições:

- Apresenta uma forma de visualizar as mudanças de densidade na lesão pulmonar em exames diferentes para a análise do especialista;

- Quatro conjuntos de métodos foram desenvolvidos para identificação, quantificação e análise das mudanças de densidade para as lesões pulmonares:
 - Um conjunto de métodos de realce das mudanças de densidade foi adaptado para análise de imagens 3D da lesão pulmonar, que antes eram utilizados em imagens 2D de sensoriamento remoto.
 - O desenvolvimento de métodos para análise local e global das mudanças de densidade em lesões pulmonares. Os métodos locais fazem uso de técnicas e métodos baseados nos conceitos de inferência estatística, como o teste de hipótese e a divergência de Jensen;
 - A adaptação da análise de componentes principais para determinar as diferenças dos valores de densidade da lesão pulmonar;
 - A modificação do *quality threshold* (QT) para a análise de mudanças de densidade ao longo do tempo.
- A utilização conjunta de duas formas de quantificar as mudanças de densidade no tempo, o percentual de mudança e o percentual de mudança intra-grupo.

1.3 Organização do Trabalho

Este trabalho é organizado da seguinte forma: o Capítulo 2 trata da fundamentação teórica necessária para sustentação da metodologia de detecção de mudanças em lesões pulmonares. É composto pelo método de aquisição das imagens do pulmão, pela sua representação digital, pelos conceitos relacionados aos tipos de lesão pulmonares, registro de imagens, pelas características de textura e pelo classificador máquina de vetores de suporte. O Capítulo 3 trata de todos os métodos baseados em *voxel*. O Capítulo 4 contém a definição dos métodos baseados em incerteza, dos quais a divergência de Jensen e o teste de hipótese fazem parte. O Capítulo 5 possui o método de análise de componentes principais (ACP). No Capítulo 6, o último método proposto utiliza o algoritmo de agrupamento *quality threshold*. Por fim, o Capítulo 7 possui uma comparação dos resultados obtidos e a proposta de trabalhos futuros.

2 Trabalhos Relacionados

A ideia de medir as transformações geométricas e de textura sofridas pela lesão pulmonar ao longo do tempo não é nova. Muitos trabalhos demonstram essa trajetória na busca de novas características que sejam úteis na avaliação de mudanças dos tecidos da lesão ao longo do tempo. A seguir, serão descritos alguns trabalhos que realizam essas detecções de mudanças para a lesão pulmonar analisado ao longo do tempo.

O trabalho de [El-Baz et al. \(2005\)](#) apresenta uma metodologia para avaliar o crescimento ou a redução do nódulo pulmonar, com o intuito de prever a sua evolução para câncer de pulmão. A metodologia envolve várias etapas. A primeira é a extração da região do pulmão pela modelagem de distribuição de nível de cinza das fatias de TC, usando uma combinação linear de gaussianas. As etapas seguintes são as classificações dos diversos tecidos; registro rígido, usando uma combinação de informações da segmentação; informação mútua, como medidas de similaridade e; registro não-rígido, para diminuir os efeitos de movimento do paciente. Os resultados dessa metodologia para uma base privada com quatro exames mostraram que é possível realçar o progresso exibido pelo nódulo pulmonar ao longo do tempo para quatro pacientes em estudo por nove meses.

Usando uma modalidade de imagem diferente para detectar um estágio anterior ao câncer pulmonar, [Shen et al. \(2006\)](#) desenvolveram uma ferramenta para análise temporal-espacial em imagens de ressonância magnética (RM) de perfusão, com o objetivo de encontrar padrões que caracterizem a angiogênese. Tal ferramenta inclui tanto algoritmos de segmentação quanto algoritmos de registro baseados em pontos de marcação em imagens sequenciais. O padrão utilizado nessa metodologia para a caracterização da angiogênese foi o perfil de densidade de regiões coincidentes para a lesão em instantes diferentes. O resultado deste estudo, para uma base proprietária com 80 imagens de RM, foi a criação de uma ferramenta para auxílio ao especialista na avaliação das mudanças sofridas pela lesão ao longo do tempo.

[Reeves et al. \(2006\)](#) utilizaram o casamento de forma baseada em conhecimento, registro de imagem, limiarização adaptativa, segmentação pleural, detecção da localização e segmentação, para estimar o tamanho e o volume do nódulo, verificando a sua mudança em exames diferentes. Foi verificado que a reamostragem isotrópica dos *voxels* melhora significativamente a detecção da mudança do volume do nódulo. O percentual de mudança

do volume foi melhorado de 11,54% para 9,35% durante a análise de 50 nódulos pulmonares proprietários.

Zheng et al. (2007) propuseram um *framework* para analisar o crescimento de tumores ao longo do tempo. Esse *framework* é capaz de iterativamente segmentar o tumor utilizando o algoritmo 2D *graph-cut* e registrar o pulmão de forma não-rígida, para que seja feita a análise de mudanças sofridas pelo tumor. Na etapa seguinte, o tumor é registrado de forma rígida com o intuito de não deformá-lo e conseqüentemente manter a geometria. Apenas 5 conjuntos de exames de TC da base proprietária, obtidos de forma sequencial, foram analisados. A segmentação produziu tumores cujos volumes possuíam erros de média e variância percentuais de 0,8 e 0,6, respectivamente.

Uma análise temporal de características de textura e geometria de nódulos malignos e benignos foi realizada por Hadjiiski et al. (2007). A análise é composta por uma segmentação automática usando modelo de contorno ativo e extração de características de textura 3D, como as estatísticas de comprimento de nível de cinza (RLS), características de nível de cinza e morfológicas em 3D. Essas características foram apresentadas a um classificador discriminante linear para classificação em malignos e benignos, usando uma técnica de reamostragem *leave-one-case-out* e uma média de 5 características foram selecionadas do conjunto de treinamento para 113 pares de nódulos pulmonares disponíveis, que formam a base proprietária. O resultado da classificação mostrou uma área sobre a curva ROC (*Receiver Operating Characteristic*) de $0,85 \pm 0,04$ e as características mais frequentemente selecionadas foram as características de mudança de intervalo e 4 estatísticas RLS, superior ao resultado obtido pelo classificador quando todas as características das imagens eram utilizadas, que nesse caso resultou em uma área sobre a curva ROC de $0,78 \pm 0,05$.

O trabalho de Lindell et al. (2007) avalia um conjunto de características histológicas, de localização, taxa de crescimento e características do TC ao longo de 5 anos de exames. Cada exame foi realizado anualmente e foi descoberto que o *screening* é uma importante ferramenta de obtenção de diagnóstico de câncer, especialmente em mulheres. A média do tamanho de tumores encontrados foi de 16,4 mm e o intervalo foi de 5,5 a 52,5 mm. A média do tempo de dobra foi de 518 dias. Dos 48 nódulos analisados da base privada, 15 tiveram tempo de dobra maior que 400 dias.

A reprodutibilidade de medidas de mudança de volume obtidas de maneira manual e por *software* foram investigadas por Honda et al. (2009). Analisando 45 cânceres pulmonares, 3 especialistas avaliaram pares de exames obtidos em instantes diferentes, se

os tumores cresceram, permaneceram ou decresceram em volume. A avaliação do volume dos tumores foi realizada em 3 instantes diferentes por *software* e classificados da mesma forma que a realizada pelos especialistas. Os exames foram então divididos em 3 níveis de concordância entre os especialistas. A variação de volume e a variação de volume relativa foram calculadas da mediana dos volumes encontrados por *software*. Os 45 tumores da base privada foram classificados como pertencentes ao 1º nível de concordância por *software* e apenas 21 tumores pertenceram ao mesmo nível, quando classificados pelos especialistas. A conclusão dessa investigação é a verificação de boa reprodutibilidade dos resultados obtidos por *software*, mesmo quando a variação do volume é pequena.

Estudos foram promovidos para detecção de trocas de volume do nódulo pulmonar, como os de [Duindam, Hoop e Ginneken \(2009\)](#), que desenvolveram três métodos para mensurar as mudanças de volume do nódulo pulmonar em exames sequenciais. Uma das três medidas é o casamento de uma esfera sobre o nódulo pulmonar para o qual é calculada a densidade dos tecidos sobre essa esfera. O segundo método consiste em realizar a segmentação dos nódulos pulmonares utilizando limiarização, processamento morfológico e *component labeling* nos exames sequenciais. O último método utiliza o registro não-rígido para transformar o primeiro exame no segundo exame e, conseqüentemente, os nódulos pulmonares do primeiro exame para os do segundo exame. Os resultados obtidos para os três métodos mostraram que os mesmos não foram capazes de detectar simultaneamente situações de estagnação ou variação no volume do nódulo pulmonar para uma base pública de imagens chamada de VOLCANO'09.

Em [Tao et al. \(2009\)](#), um estudo foi proposto para avaliar a performance do registro automático de nódulos pulmonares de três exames sequencias de TC. O estudo consistiu do registro de nódulos pulmonares obtidos no instante T1 com os obtidos no instante T2; e o nódulo pulmonar em T2 com o instante T3, sempre calculando o percentual de registro entre cada par registrado. A taxa de registro dos nódulos pulmonares foi de 92,7%, para a base pública de imagens do *National Lung Screening Trial*, com 145 lesões pulmonares. Como resultado, a técnica se mostrou invariante ao tamanho do nódulo pulmonar.

[Sofka e Stewart \(2010\)](#) apresentam um algoritmo capaz de mapear nódulos pulmonares de um exame obtido em um determinado instante para as suas respectivas posições em um segundo exame posterior, mesmo que o segundo exame tenha sido obtido em um procedimento completamente diferente. Uma vez que o mapeamento (registro) é realizado, uma análise no crescimento do nódulo pulmonar pode ser verificada. A base

pública utilizada é denominada de *RIDER*, com 84 nódulos pulmonares. A média e a variância da distância entre os nódulos pulmonares mapeados foram calculadas. Os resultados mostraram que a metodologia é mais eficiente no alinhamento dos nódulos, que os métodos de registro *Demons Diffeomorphic* e das técnicas de registro global-para-local.

Xue, Wong e Wong (2010) também mostram um algoritmo iterativo de segmentação e registro guiado por imagem. Esse trabalho mostra como é possível melhorar o tratamento e o diagnóstico de lesões de pequenos nódulos pulmonares periféricos em imagens de tomografia computadorizada sequenciais. A metodologia consiste basicamente em 3 etapas. Na primeira etapa, a segmentação de todas as lesões suspeitas é realizada para todos os exames analisados ao longo do tempo. As estruturas segmentadas na primeira etapa são utilizadas na segunda etapa, que compreende o registro deformável longitudinal das estruturas correspondentes em tempos diferentes. Este resultado é então utilizado em um algoritmo de agrupamento que segmenta regiões semelhantes das estruturas correspondentes em tempos diferentes. Essas novas regiões segmentadas são novamente utilizadas na segunda etapa, que continua nas etapas seguintes, até que a diferença entre as regiões segmentadas fique abaixo de um limiar. Foram utilizados exames proprietários de 20 pacientes e a metodologia foi capaz de detectar 32% das mudanças sofridas pelos tecidos.

Chen et al. (2011) desenvolveram um novo método para segmentar automaticamente os nódulos pulmonares e registrá-los em exames de TC sequenciais feitos em diferentes instantes. A segmentação automática nesse método é realizada através da análise das estruturas de intensidade local e dos vasos sanguíneos. O método utiliza os registros afim e o não-rígido para compensar a deformação do nódulo pulmonar. O método foi aplicado a 14 exames de 3 pacientes, que formam uma base privada, e 96,7% dos nódulos foram detectados, enquanto 83,3% deles tiveram correspondência no registro.

Uma análise semelhante foi realizada por Henschke et al. (2012), em que um *screening* repetidamente feito no ano avalia o tempo de dobra para cada tipo celular do câncer. Os resultados mostraram que a mediana do tempo de dobra para 111 cânceres da base privada foi de 98 dias. Em 50% dos casos, o tempo de dobra foi menor que 100 dias e para 3% dos casos, o tempo de dobra foi maior que 400 dias. Como previsto, nódulos pulmonares subsólidos apresentaram tempo de dobra muito maior que os sólidos.

O trabalho de Linning et al. (2013) comparou a detecção de mudanças de volume com uma maneira de verificar modificações no crescimento do nódulo pulmonar. Análises quantitativas do volume dos componentes sólidos do nódulo pulmonar foram realizadas na

tentativa de verificar a ocorrência de modificações. O percentual de mudança do volume total foi comparado ao do volume dos componentes sólidos. Os resultados mostraram que para os 21 casos de nódulos pulmonares malignos da base privada, o percentual de mudança dos componentes sólidos (115,78%-418,91%; $130,45 \pm 119,48$) cresceu com um percentual maior que o do volume do nódulo pulmonar (78,56%-105,73%; $42,34 \pm 32,17$).

Aerts et al. (2014) demonstraram que o câncer exibe diferentes fenótipos que são detectados por meio de imagem. O termo *radiomics* é utilizado quando se deseja determinar o fenótipo exibido por uma lesão. O fenótipo é obtido a partir da extração de um grande número de características. O trabalho obteve 440 características baseadas em intensidade, forma e textura. A metodologia foi aplicada em várias bases públicas, totalizando 1019 pacientes com câncer nas regiões do pulmão, pescoço ou cabeça. Os resultados mostram que muitas características que anteriormente não foram consideradas significantes na determinação do prognóstico mostraram-se promissoras nessa tarefa, independente da base de câncer utilizada.

Nagarajan et al. (2015) realizaram um estudo de mudanças volumétricas da matriz de cartilagem patelar em imagens de tomografia computadorizada de raio-X de contraste de fase. A ideia é capturar mudanças na organização de condrócitos induzidas por osteoartrite através da avaliação de características geométricas e topológicas. As características geométricas foram derivadas do método de índice de escala e as topológicas foram dos funcionais de Minkowski. Para 1392 volumes de interesse de uma base proprietária, as características foram extraídas e, através da regressão de vetores de suporte, foi realizada uma classificação em saudável ou não-saudável. A área sobre a curva ROC (*Receiver Operating Characteristic*) para as características geométricas foi $0,9 \pm 0,09$ e foi melhor que as características de Minkowski: volume ($0,54 \pm 0,02$), superfície ($0,72 \pm 0,06$), distância média ($0,74 \pm 0,06$) e Euler ($0,78 \pm 0,04$). Os resultados sugerem que as características geométricas são bons indicadores para o diagnóstico, a progressão e a resposta terapêutica da osteoartrite.

A variabilidade das medidas do tamanho da lesão em imagens de tomografia computadorizada sob condição de nenhuma mudança foi estudada por McNitt-Gray et al. (2015). O objetivo foi verificar o impacto da variabilidade das medidas com dois paradigmas de leitura. Foi utilizada uma base de 32 pacientes com câncer de pulmão de não-pequenas células obtidas no intervalo de 15 minutos e as medidas foram realizadas por cinco especialistas em duas fases. A diferença percentual média (DPM) para as medidas de

tamanho da lesão foram ($2,8 \pm 22,2\%$) em 1D e ($23,4 \pm 105,0\%$) em 3D para as leituras, independentes do exame no tempo. Os valores de DPM para leituras sequenciais no tempo para um exame foram reduzidos para ($2,52 \pm 14,2\%$) em 1D e ($7,4 \pm 44,2\%$) em 3D. A conclusão foi que existe variação nas medidas da lesão e a mesma se reduz quando ela é realizada sequencialmente no tempo.

A incerteza dos valores de volume obtidos por dois diferentes métodos (baseado em modelo e baseado em segmentação) foi avaliada sobre 29 nódulos sintéticos com tamanhos, densidade e formas variadas, segundo o trabalho de [Li et al. \(2015\)](#). A aquisição foi realizada dez vezes sob diferentes condições e os volumes foram reconstruídos pela aproximação de filtro-casado e por segmentação. Através da aplicação da regressão sobre as medidas de erro, foram obtidos os fatores que afetam a estimação do volume para cada método. Os resultados mostraram que, para ambos os métodos, os fatores que influenciam na estimação do volume foram o tamanho do nódulo, a forma e a espessura da imagem.

A partir do conceito de *radiomics*, [Coroller et al. \(2016\)](#) propuseram um conjunto de características para predição da resposta patológica em pacientes com câncer de pulmão não-pequenas células localmente avançado após quimioterapia neoadjuvante. Um conjunto de quinze características foram selecionadas a partir da base com 127 pacientes. Os resultados obtidos foram comparados com a avaliação convencional que usa o volume e o diâmetro da lesão. Os resultados da predição mostraram que as características do *radiomics* são melhores para a predição que as características convencionais. Da mesma forma que o trabalho anterior, que usa *radiomics*, este trabalho usa um grande conjunto diversificado de característica e seleciona as que mais corroboram com a tarefa de predição.

[Bak et al. \(2016\)](#) analisaram imagens de nódulos de opacidade de vidro-fosco para determinar mudanças futuras ou crescimento. As mudanças foram classificadas de três formas: 1 - Desenvolvimento de componente sólido; 2 - Crescimento do componente de vidro-fosco e; 3 - Nenhuma mudança no tamanho. Logo, as análises basearam-se na avaliação do tamanho da lesão, volume, densidade e valores de atenuação do exame de TC inicialmente adquiridos. Os resultados mostraram que a avaliação com base nas medidas propostas não eram estatisticamente diferentes entre as classes, mas os valores de atenuação eram estatisticamente diferentes. Portanto, eles concluíram os valores de atenuação podem ajudar a prever mudanças em suas componentes, bem como o crescimento da lesão pulmonar.

Trabalhos como os de [Zheng et al. \(2007\)](#), [Duindam, Hoop e Ginneken \(2009\)](#), [Reeves et al. \(2006\)](#), [Honda et al. \(2009\)](#) e [Linning et al. \(2013\)](#) ainda estão relacionados à medição volumétrica do nódulo pulmonar, como medida de variabilidade da lesão ao longo do tempo. Mesmo quando é possível verificar aonde estas mudanças ocorreram, as técnicas acima não conseguem verificar simultaneamente mudanças dos tecidos internamente e externamente à lesão ao longo do tempo.

Os trabalhos apresentados neste capítulo e resumidos na Tabela 2.1 apresentaram resultados interessantes quando se está analisando a variação do volume da lesão pulmonar ao longo do tempo. Entretanto, verifica-se, da leitura acima, que muitos deles não usam uma base pública de imagens de lesões pulmonares. A base pública facilita a comparação e a efetividade das técnicas na avaliação das mudanças. Ao mesmo tempo, observa-se também que a maioria dos trabalhos não apresenta simultaneamente resultados para identificação das mudanças, realce das mesmas, quantificação e análise para conclusão do diagnóstico. Conclui-se que, em virtude de a análise temporal de mudanças de densidade ser ainda um campo em aberto, esta pesquisa se concentrou em estudar e adaptar métodos e técnicas que avaliam as mudanças de intensidade (densidade do *voxel*) ao longo do tempo.

Tabela 2.1 – Alguns trabalhos relacionados descritos em termos de suas técnicas, base de imagens e resultados.

Trabalho	Técnicas	Base de imagens	Resultados
El-Baz et al. (2005)	Algoritmo EM, registro B-splines, mudança de volume	Proprietária: 4 exames com intervalo de aquisição de 9 meses.	Variação do volume no intervalo de 0,3872 a 0,8400 mm.
Shen et al. (2006)	pMRI, ferramenta de análise espaço-temporal, maximização da correlação, estratégia baseada em modelo	Proprietária	-
Reeves et al. (2006)	Laplaciano da gaussiana, limiarização adaptativa, registro de corpo rígido	Proprietária: 50 nódulos pulmonares	Redução do desvio padrão do percentual de mudança do volume de 11,54% para 9,35%.
Zheng et al. (2007)	registro B-splines, algoritmo 2D graph-cut	Proprietária: 6 exames de 3 pacientes com 12 meses de diferença.	Média e variância do erro na segmentação, respectivamente, de 0,8 e 0,6.
Hadjiiski et al. (2007)	Modelo de contorno ativo 3D, run-length, morfologia 3D, característica de nível de cinza e classificador linear discriminante	Proprietária: 103 pares de nódulos com intervalo de tempo de 2 a 16 meses.	curva ROC de 0.85 ± 0.04 .
Lindell et al. (2007)	Software de análise de mudança morfológica e tempo de dobra do volume (VDT)	Proprietária: 61 exames e 48 nódulos	15 dos 48 nódulos tiveram VDT maior que 400 dias. 11 dos 13 cânceres estavam nas mulheres.
Honda et al. (2009)	Software de análise de mudança morfológica e tempo de dobra do volume (VDT)	Proprietária: 45 cânceres	Inconsistência na classificação realizada pelo especialista, boa reprodutibilidade por software, mesmo quando a variação do volume é pequena.
Duindam, Hoop e Ginneken (2009)	Segm, Sphere, Reg	Pública: 50 pares de nódulos	Os 3 métodos possuem boa correlação entre si, mas não conseguem detectar simultaneamente situações de estagnação ou variação no volume do nódulo pulmonar.
Tao et al. (2009)	Software de análise	Pública: 145 nódulos pulmonares	Taxa de casamento de 92,5% dos nódulos pulmonares.
Sofka e Stewart (2010)	Registro baseado em características, descritor de contexto de forma, algoritmo LRR, MVS	Proprietária: 67 pares de nódulos; e Pública: RIDER com 84 nódulos pulmonares.	Erro de mapeamento foi de 1,70 mm.

Fonte: Autor

Tabela 1: Alguns trabalhos relacionados descritos em termos de suas técnicas, base de imagens e resultados.

Trabalho	Técnicas	Base de imagens	Resultados
Xue, Wong e Wong (2010)	Deformação de forma livre, agrupamento 4D	Proprietária: 20 exames de 2 a 3 meses de intervalo.	Erro médio do registro no intervalo de 0,9 a 4,1 mm
Chen et al. (2011)	blob-like shape, blob structure enhancement, registro B-splines	Proprietária: 14 exames com 180 nódulos pulmonares.	96,7% dos nódulos pulmonares foram detectados e 83,3% deles tiveram correspondência no casamento.
Henschke et al. (2012)	Software de análise de mudança morfológica e tempo de dobra do volume (VDT)	Proprietária: 111 cânceres	50% dos casos, o VDT foi menor que 100 dias e 3% dos casos, o VDT foi maior que 400 dias.
Linning et al. (2013)	Software de análise	Proprietária: 21 exames com 21 nódulos malignos.	Percentual de troca no volume dos componentes sólidos: (115,78–418,91%, 130,45±119,48).
Aerts et al. (2014)	Radiomics, análise de características de intensidade, textura e forma no tempo	Pública: 1019 pacientes de 3 tipos de cânceres	Descoberta de características que são úteis para o prognóstico
Nagarajan et al. (2015)	Análise volumétrica de condrócitos, com características geométricas e topológicas de Minkowsk	Proprietária: 1392 VOI's	As características geométricas são bons indicadores para o diagnóstico, progressão e resposta terapêutica da osteoartrite
McNitt-Gray et al. (2015)	Variabilidade das medidas do tamanho da lesão ao longo do tempo	Pública: RIDER com 32 nódulos pulmonares	Existe variação das medidas da lesão independente do tempo e diminui sequencialmente no tempo
Li et al. (2015)	Incerteza da estimação dos volumes obtidos pelas técnicas de filtro-casado e segmentação	Proprietária: 29 phantom nódulos pulmonares	O Tamanho e a forma influenciam nas medidas de volume

Fonte: Autor

3 Fundamentação

Apresenta-se neste capítulo o embasamento teórico necessário para compreensão dos métodos e das técnicas propostos neste trabalho de tese. O capítulo é composto pelos conceitos relacionados à forma de aquisição das imagens de tomografia computadorizada do pulmão, características das imagens geradas, dos nódulos e das massas pulmonares e da teoria de registro das lesões segmentadas.

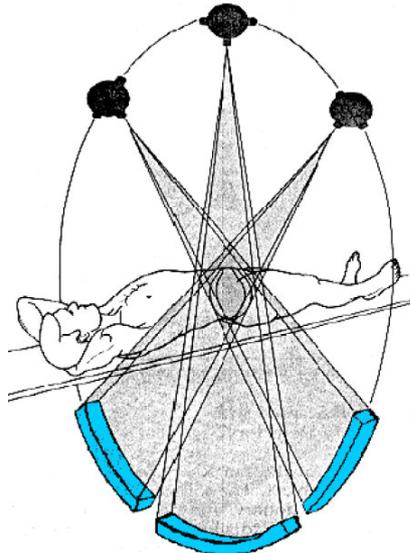
3.1 Tomografia Computadorizada

A palavra tomógrafo é derivada da palavra grega “Tomos”, que significa corte ou fatia, e “grafos”, que significa desenhar uma imagem ou gráfico. Assim, trata-se de um dispositivo de aquisição de imagens que emprega os mesmos princípios da radiografia convencional. O objetivo desse dispositivo é criar uma representação anatômica dos diversos órgãos do paciente em várias imagens. A geração das imagens é baseada na quantidade de atenuação sofrida pela radiação incidente nos diversos tecidos do paciente. O tomógrafo surgiu em meados de 1969, através dos trabalhos de Sir Godfrey Hounsfield (SOARES, 2006). Para tornar esse dispositivo possível de funcionar, Hounsfield projetou um sistema eletrônico de detecção do feixe de radiação e necessitava de um processo de formação das imagens. Em colaboração com Allan M. Cormack, toda a matemática necessária para a reconstrução das imagens tomográficas foi desenvolvida, revolucionando a forma de sua aquisição. Por essa razão, o dispositivo foi denominado tomógrafo computadorizado (TC) e a imagem produzida é chamada de tomografia computadorizada.

O funcionamento do TC é basicamente semelhante ao do tomógrafo convencional e pode ser visto na Figura 3.1. Um feixe de raios X estreito é gerado por um cabeçote e registrado por um detector ligado diametralmente. Conseqüentemente, a radiação que atravessa o paciente sofre uma atenuação. A atenuação depende da densidade e do número atômico médio do tecido analisado, sendo que sua intensidade é dada em unidades de Hounsfield (*Hounsfield Unit* - HU). Ao mesmo tempo, a construção da imagem seccional é obtida por um movimento de rotação do conjunto cabeçote-detector. Após um movimento completo de rotação, é iniciado um novo processo de varredura, que se inicia depois de um deslocamento da mesa de alguns décimos de milímetro. A obtenção das imagens finaliza quando a mesa passa por todo eixo axial do corpo do paciente. Ao realizar seu

funcionamento, pode-se perceber que vários elementos estão envolvidos e isso faz com que o TC possua características que o diferencia do tomógrafo convencional.

Figura 3.1 – Processo de varredura do cabeçote e detectores.



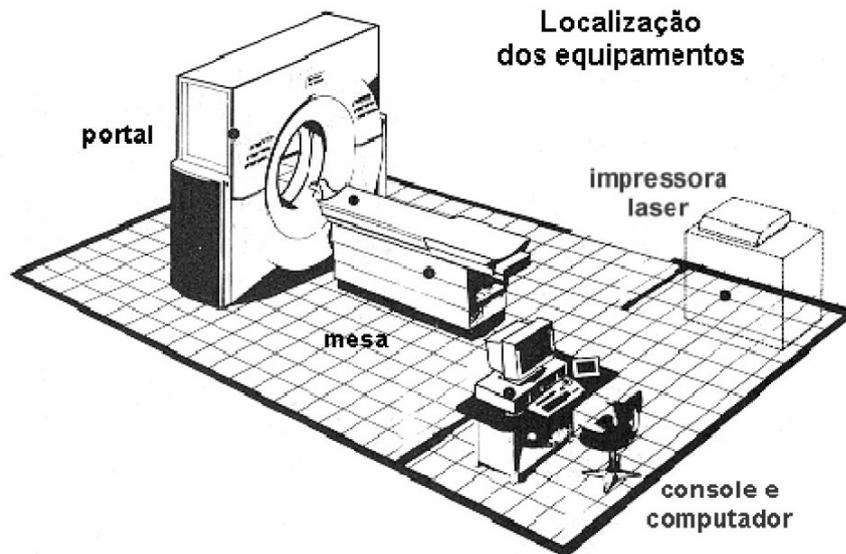
Fonte: SOARES (2006)

O tomógrafo computadorizado pode ser dividido, conforme a visualização da Figura 3.2, em três grandes partes:

- O portal é composto pelo cabeçote contendo a ampola de raios X e pelo detector das radiações atenuadas, que está diametralmente oposto;
- A eletrônica de controle e de aquisição dos dados é responsável pela monitoração dos sensores e por enviar comandos aos dispositivos servo-mecânicos. Dessa forma, a transdução dos sinais captados pelo detector são enviados ao computador para a reconstrução da imagem;
- Console de comando e computador, isto é, o *software* encarregado de gerenciar toda estrutura do tomógrafo computadorizado e da formação das imagens; e o *hardware* no qual este programa é executado, respectivamente.

As características essenciais do aparelho de TC são a não sobreposição de estruturas, melhor contraste (menos radiação espalhada), detectores otimizados e processamento das imagens para produção da imagem final. Essas características o difere da radiografia simples e facilita a detecção de nódulos inferiores a 1 cm. Em razão dessas características, são formadas imagens bidimensionais do pulmão de alta qualidade, formando uma matriz de elementos.

Figura 3.2 – Partes de um tomógrafo computadorizado.



Fonte: SOARES (2006)

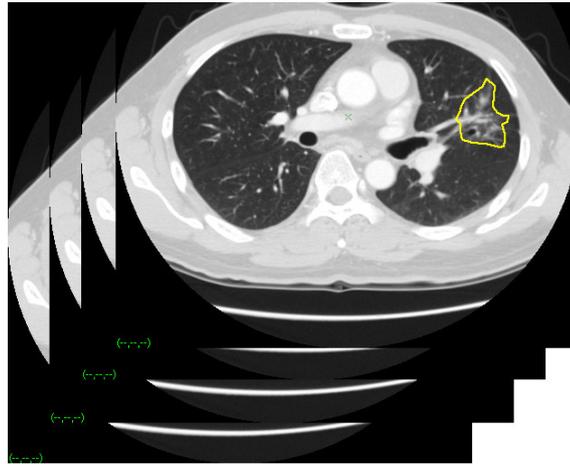
A matriz de imagem seccional formada do cálculo computacional do *software* do aparelho está em nível de cinza. Esse valor está em uma escala de contraste de Hounsfield, que usa como referência diversas substâncias, tais como a água (entre 0 e ± 10), o osso cortical denso (entre 1000 e 3000) e o ar (com valor igual a -1000). O campo de reconstrução das imagens é chamado de *Field Of View* (FOV) e é dado pelo produto entre o tamanho da matriz da imagem e a dimensão do *pixel*, definindo o tamanho do objeto em estudo. Este objeto de estudo é tridimensional, portanto, existe um conjunto de imagens que representam este objeto.

3.2 Imagem 3D

O TC gera um conjunto de imagens bidimensionais. Cada imagem bidimensional é formada por uma matriz de valores correspondentes à intensidade dos *pixels* (abreviação de *picture elements*), como se pode observar nas Figuras 3.3 e 3.4. Esse conjunto de imagens forma uma imagem tridimensional, que é tratada como uma coleção de dados volumétricos organizados em uma matriz tridimensional. Similarmente, cada elemento dessa matriz tridimensional é denominado *voxel* (abreviação para *volume elements*) e a imagem como um todo possui algumas características.

Em aplicações médicas, algumas características da imagem são essenciais. Pode-se citar como exemplos as dimensões do *pixel* ou *voxel* e a origem da imagem no sistema de

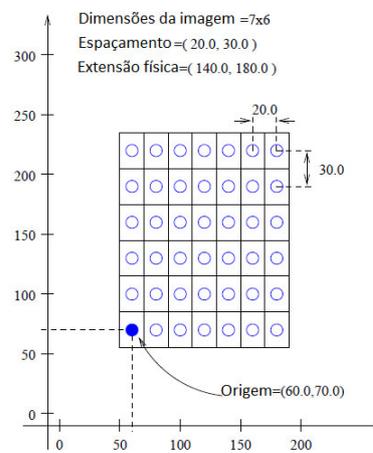
Figura 3.3 – Conjunto de imagens bidimensionais que formam uma imagem 3D do pulmão.



Fonte: Autor

referência do mundo. A Figura 3.4 possui alguns desses elementos essenciais para o caso de uma imagem 2D, que possui 7 *pixels* de largura e 6 *pixels* de altura. Isso corresponde a uma extensão física da imagem de 140 (7 x 20) milímetros de largura e 180 (6 x 30) milímetros de altura, caso as dimensões do *pixel* ou do espaçamento sejam 20 mm de largura e 30 mm de altura. O mesmo processo pode ser usado para imagens 3D, lembrando que a imagem e o *voxel* possuem 3 dimensões. Todo o conjunto de características da imagem citadas são importantes para qualquer que seja o processamento na área médica. A partir delas, muitas informações podem ser extraídas sobre a imagem ou da lesão pulmonar em estudo, a exemplo, pode-se citar o tamanho da lesão, o volume e etc.

Figura 3.4 – Características da imagem 2D no sistema de referência da imagem e do mundo.

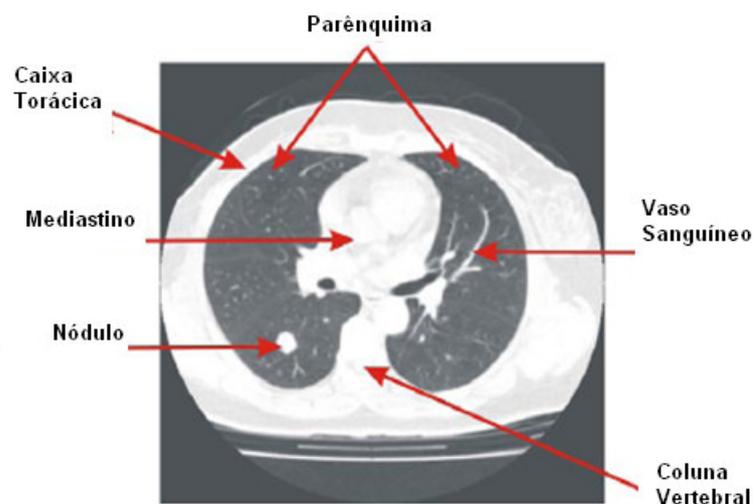


Fonte: Ibanez et al. (2003)

3.3 Lesões Pulmonares

O nódulo pulmonar é uma lesão de forma esférica, menor que 3 cm e não acompanhada de quaisquer lesões sugestivas de metástase ou invasão de estruturas vizinhas. Pode ser classificado em maligno, benigno ou indeterminado, de acordo com o diagnóstico dado pelo especialista. Essa última classificação é dada quando não se tem certeza se o mesmo é maligno ou benigno (OLIVEIRA et al., 2010). Outra divisão encontrada, levando em conta as densidades presentes na lesão, permite a classificação em nódulos sólidos e subsólidos. Nódulos sólidos são aqueles completamente visíveis no parênquima pulmonar. Os subsólidos também são denominados de não-sólidos ou de vidro-fosco (GODOY; NAIDICH, 2009), por serem parcialmente visíveis na região do parênquima. Neste trabalho, faz-se uso apenas de nódulos sólidos.

Figura 3.5 – Estruturas do pulmão.



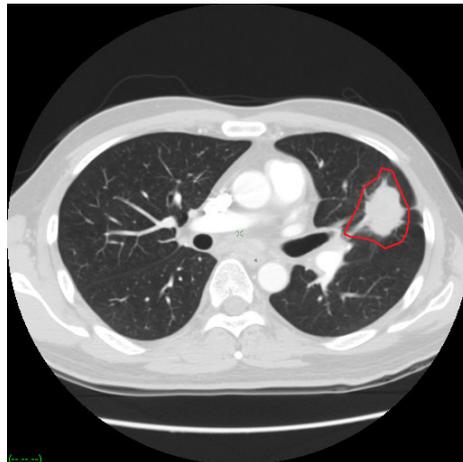
Fonte: Silva (2004)

Nódulos sólidos possuem alto contraste com relação ao parênquima e a imagem da Figura 3.5 possui um exemplo desse tipo de nódulo e das estruturas presentes na imagem de TC do pulmão. Quando achadas precocemente, podem exibir dimensões inferiores a 1 cm. A grande preocupação nesses casos é a busca de características que possam indicar algum sinal de malignidade, como a forma e as suas características nas bordas. Se as informações obtidas dessas características ainda não forem suficientes para a determinação da malignidade, o especialista observa se existirão mudanças no nódulo pulmonar em um intervalo de tempo e esse procedimento é chamado neste trabalho de análise temporal da lesão pulmonar (OLIVEIRA et al., 2010; BAGLIO; GAZZANA; SILVA, 2009; LIBBY;

HENSCHKE; YANKELEVITZ, 1995). Como se pode perceber, apesar de a lesão possuir contraste em relação ao parênquima, suas características iniciais podem não definir a sua natureza. Ao contrário desse tipo de lesão, as massas são estruturas que permitem melhor identificação.

Outra estrutura de interesse nas imagens de TC são as massas pulmonares. São estruturas maiores que o nódulo pulmonar, isto é, maiores que 3 cm. As massas são geralmente densas e, da mesma forma que os nódulos, podem ser diagnosticadas como malignas ou benignas (RINGEL, 2008). Um exemplo de massa pulmonar pode ser visto na Figura 3.6, onde o contorno em vermelho define os seus limites em uma única imagem de TC.

Figura 3.6 – Pulmão com a massa destacada pelo contorno.



Fonte: Autor

Ao longo do texto, usou-se o termo lesão pulmonar para referenciar tanto os nódulos quanto as massas pulmonares.

3.3.1 Base de lesões pulmonares

As imagens para a análise temporal das mudanças das lesões pulmonares foram adquiridas a partir de uma base pública de imagens, criada pela *Prevent Cancer Foundation* (PCF) em conjunto com o *National Cancer Institute* (REEVES et al., 2009). A segunda base privada foi criada com nódulos pulmonares benignos. Portanto, as duas bases são utilizadas com o objetivo de observar o comportamento das mudanças de densidade em situações distintas de avaliação pelo especialista.

A base pública é caracterizada pela presença de lesões malignas em tratamento, enquanto a outra, privada, é caracterizada por nódulos indeterminados, mas provavelmente benignos e não estão em tratamento. Logo, acredita-se que esta diversidade nas bases seja essencial para a busca de importantes resultados, pois em uma, as lesões são passíveis de grandes variações no volume e densidade devido ao tratamento. Enquanto na outra, sem tratamento, e por ser constituída por prováveis nódulos benignos, são mais estáveis em densidade e volume.

3.3.1.1 Base privada

O conjunto de lesões pulmonares desta base privada foi adquirido no Hospital Universitário Pedro Ernesto - Rio de Janeiro. O tomógrafo utilizado é o modelo GE HiSpeed helicoidal com voltagem do tubo de 140 KVP e corrente de 150 mA. A base consiste exclusivamente de nódulos pulmonares que foram considerados inicialmente como indeterminados, mas com alta probabilidade de serem benignos. A mesma é constituída de 16 pacientes em estado de avaliação das lesões e possui o total de 40 lesões pulmonares. Todos os nódulos pulmonares foram segmentados manualmente pelo especialista com anos de experiência. O acompanhamento do comportamento da lesão pelo especialista foi anual, com duração de 6 anos em alguns casos. As imagens estão no formato DICOM¹ e a espessura do *voxel* é de 1,0 ou 2,0 mm. O número de imagens por lesão está no intervalo de 3 a 19 imagens, com média de 10,25 imagens. O volume médio dos nódulos pulmonares foi 0,51 cm^3 .

3.3.1.2 Base pública

A base pública é formada por casos de pacientes em tratamento, então, as imagens são de exames antes e depois da aplicação dos medicamentos.

A base possui 24 lesões obtidas em diferentes tempos de aquisição. A maioria dos casos da base são massas pulmonares, formando um conjunto de 58 exames (REEVES et al., 2009). As imagens estão no formato DICOM com a espessura do *voxel* variando de 2,5 a 7,5 mm. O número de fatias das lesões está no intervalo de 5 a 132, com média de 14,53, o que indica a presença de grandes lesões pulmonares. O volume médio das lesões foi de

¹ *Digital Imaging and Communications in Medicine*. - Padrão de armazenamento de imagens

Figura 3.7 – Imagem de TC de um paciente e lesão pulmonar envolta de um contorno em amarelo.

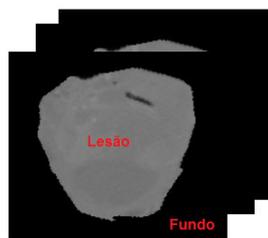


Fonte: Autor

164,3 cm^3 . Como complemento, a base de imagens oferece uma marcação (contorno) para cada lesão de interesse sobre o exame e as medidas volumétricas, que podem ser úteis em estudos comparativos, conforme se pode observar pela Figura 3.7.

A imagem 3D da lesão pulmonar foi obtida através da segmentação da mesma. A segmentação foi realizada a partir da detecção dos pontos internos de uma marcação. Essa marcação é definida pelo especialista ao longo das imagens de TC para a lesão. A Figura 3.8 possui um exemplo da imagem segmentada. Na figura, percebe-se que a lesão possui duas regiões distintas, a região da própria lesão e a região do fundo. Chama-se de fundo da imagem, um valor de marcação que define quais são os *voxels* que não pertencem à lesão e que foi colocado para tornar a imagem regular. Dessa forma, tem-se uma imagem 3D da lesão com dimensões proporcionais aos limites da lesão pulmonar.

Figura 3.8 – Exemplo de imagem resultante da segmentação da lesão pulmonar e do fundo da imagem.



Fonte: Autor

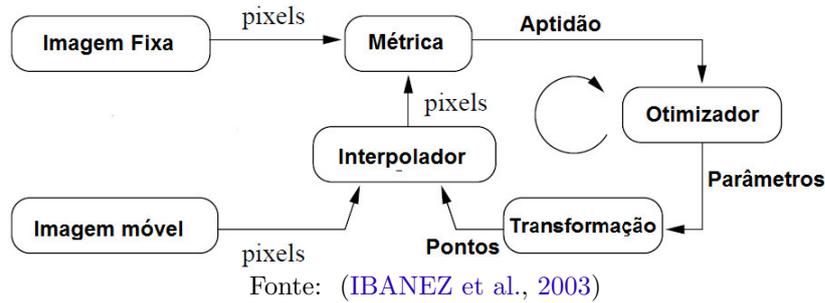
3.4 Registro de Imagens

Devido aos diferentes protocolos de aquisição, os exames da mesma lesão pulmonar consistem em imagens com características distintas. Um exemplo dessas características são as dimensões das imagens, as dimensões dos *voxels* e etc. Outros fatores podem produzir exames com características diferentes, por exemplo, os movimentos respiratórios realizados pelo paciente e as diferenças da sua própria disposição sobre o tomógrafo nos instantes de aquisição. Logo, o registro rígido é uma etapa fundamental para uniformizar essas características e corrigir os desvios de alinhamento entre o par de imagens da lesão que estão em análise.

O registro de imagens consiste em encontrar o melhor alinhamento que mapeia ou transforma o conjunto de pontos de uma imagem, chamada de móvel, para outro conjunto de pontos de uma imagem, chamada fixa. O mapeamento precisa da definição de uma métrica que verifica a semelhança entre os pontos das imagens móvel e fixa e de uma função de interpolação responsável por obter os valores dos *pixels* da imagem móvel que foram mapeados para posições fora da matriz da imagem. Tal mapeamento da imagem móvel para a imagem fixa é realizado por uma transformação. Os parâmetros desta transformação são ótimos quando a métrica de semelhança entre as imagem móvel e fixa é submetida a um otimizador, que busca encontrar o melhor alinhamento (PANDAV; KHOT, 2015).

Para que a imagem móvel se torne mais semelhante possível à imagem fixa, cada um dos pontos da imagem fixa precisa ter um correspondente na móvel. Entretanto, cada um dos pontos correspondentes entre as imagens estão no espaço tridimensional do objeto, e não no espaço da imagem. Logo, os pontos da imagem móvel são modificados por uma transformação para o sistema de referência do mundo. Essa nova imagem gerada no sistema de referência do mundo tem cada um dos seus pontos interpolados. Da mesma forma, a imagem fixa também vai para o mesmo sistema de referência do mundo. Assim, com ambas as imagens no mesmo sistema de referência do mundo, é possível comparar a correspondência entre os pontos das imagens fixa e móvel através de uma métrica de similaridade. Os parâmetros da transformação são encontrados através do uso de um método de otimização, que encontra o melhor alinhamento entre as imagens, com base no melhor valor para a métrica de similaridade. A Figura 3.9 contém cada um dos componentes descritos para o processo de registro de imagens e as suas interações.

Figura 3.9 – Componentes do registro.



Os métodos de registro de imagens podem ser classificados quanto à função de transformação, em rígidos e não-rígidos. Nos métodos rígidos, as transformações não produzem deformações sobre a imagem, enquanto os não-rígidos podem produzir distorções nas imagens. Os métodos não-rígidos podem ser subdividido em modelo afim, modelo linear elástico, modelo de fluido viscoso, modelo de elementos finitos, modelo de função de base radial, modelo de fluxo óptico e etc. Portanto, o método aplicado depende do resultado final desejado, isto é, com ou sem distorção das imagens registradas.

Na próxima seção, mostra-se, resumidamente, cada um dos componentes utilizados no registro rígido de imagens. Esse método foi escolhido com o intuito de não provocar nenhuma deformação nas imagens das lesões pulmonares em análise.

3.4.1 Registro Rígido de Imagens 3D

É caracterizado pela aplicação de transformações de corpo rígido, como a translação e a rotação, sobre um conjunto de pontos em 3D da imagem.

3.4.1.1 Transformação 3D

A Equação 2 realiza uma translação 3D do ponto x em relação aos eixos cartesianos, através da especificação do vetor $t = [t_x, t_y, t_z]$ e de uma rotação R , dada pela Equação 1, que define a rotação com ângulos Θ_x , Θ_y e Θ_z sobre os respectivos eixos coordenados x , y

e z . A rotação R é composta pelas rotações $R_z(\Theta_z)$, $R_y(\Theta_y)$ e $R_x(\Theta_x)$, sobre os eixos z , y e x , respectivamente.

$$R = \begin{bmatrix} \cos(\theta_z) & -\sin(\theta_z) & 0 \\ \sin(\theta_z) & \cos(\theta_z) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos(\theta_y) & 0 & \sin(\theta_y) \\ 0 & 1 & 0 \\ -\sin(\theta_y) & 0 & \cos(\theta_y) \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta_x) & -\sin(\theta_x) \\ 0 & \sin(\theta_x) & \cos(\theta_x) \end{bmatrix} \quad (1)$$

Uma transformação de corpo rígido 3D é definida pela Equação 2 e é composta por uma translação do vetor x rotacionado.

$$x' = R\mathbf{x} + \mathbf{t} \quad (2)$$

Uma outra forma de determinar rotações é através do conceito matemático denominado quatérnio. O mesmo pode ser definido como um vetor q unitário, consistindo de quatro elementos - q_0, q_x, q_y, q_z , que, quando aplicado a outro vetor, produz um novo. Um outro operador importante é o versor que, quando aplicado a um vetor, provoca uma mudança em sua direção mantendo o mesmo comprimento. O versor produz uma rotação Θ sobre um eixo representado por um vetor $\hat{\omega}$. Basicamente, seus componentes são definidos pela parte escalar $q_0 = \cos(\Theta/2)$ e o vetor $[q_x, q_y, q_z]$, pela parte imaginária (SONKA; FITZPATRICK, 2000).

Dado um vetor x , o quatérnio $q = [q_0, q_x, q_y, q_z]$ que produz uma rotação sobre x é definido pela Equação 3.

$$x' = q \diamond x \diamond q^{-1} \quad (3)$$

sendo \diamond representa a operação do quatérnio sobre o vetor x .

A Equação 3 pode ser expandida na forma

$$R = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2q_0q_1 - 2q_3q_2 & 2q_0q_2 + 2q_3q_1 & 0 \\ 2q_0q_1 + 2q_3q_2 & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2q_1q_2 - 2q_3q_0 & 0 \\ 2q_0q_2 - 2q_3q_1 & 2q_1q_2 + 2q_3q_0 & q_0^2 - q_1^2 - q_2^2 + q_3^2 & 0 \\ 0 & 0 & 0 & q_0^2 + q_1^2 + q_2^2 + q_3^2 \end{bmatrix} \times x \quad (4)$$

O vetor x é tratado como um quatérnio de parte real nula. Como o quatérnio utilizado para rotações é unitário, este faz com que se transforme em um versor que provoca uma rotação pura, isto é, sem o problema chamado *gimbal lock*. As componentes do quatérnio formam parte dos parâmetros da transformação a serem encontrados pelo

otimizador. Como os componentes do versor não formam um espaço vetorial, é necessário um otimizador especializado para a tarefa de otimização.

3.4.1.2 Interpolador

Quando os pontos de uma imagem são transformados de um espaço da imagem para outro espaço e são trazidos de volta após alguma transformação, esses pontos podem cair em posições que estão fora da matriz da imagem. Portanto, necessita-se encontrar o valor de intensidade do *pixel* ou *voxel* para a posição mapeada.

A escolha do método de interpolação pode influenciar na suavidade do espaço da métrica de similaridade. Os métodos mais comuns utilizados para interpolação são o linear, *BSpline* e o vizinho mais próximo. Este último será brevemente explicado a seguir, pois foi o utilizado neste trabalho.

A interpolação pelo vizinho mais próximo utiliza a noção de que dado um ponto, os seus vizinhos mais próximos tendem a ser mais semelhantes. Matematicamente, dado um conjunto de pontos (x_i, y_i) , com $y = f(x)$, se x_0 é ponto a interpolar, o y interpolado obedece a Equação 5, que procura minimizar a distância em x .

$$y = f(\min_{x_i}((x_0 - x_i)^2)) \quad (5)$$

3.4.1.3 Métrica de semelhança

A métrica é utilizada pelo otimizador para avaliar quantitativamente os parâmetros da transformação. Dependendo do método de otimização, determina-se algumas derivadas da métrica de similaridade em função dos parâmetros da transformação.

As métricas de similaridade mais comuns são a média quadrática, a correlação normalizada e a informação mútua.

A média quadrática é descrita pela Equação 6. Dado um número N de pontos na imagem, a média quadrática é o somatório do quadrado da diferença entre o valor de intensidade da imagem móvel M , interpolada no ponto x' após a transformação T , pelo valor de intensidade no ponto x , na imagem fixa F . Como se pode observar, a métrica possui algumas características especiais, como a simplicidade de ser calculada e amplitude

do raio de variação. O seu valor ótimo é atingido quando passa a ser o mais próximo possível do valor zero.

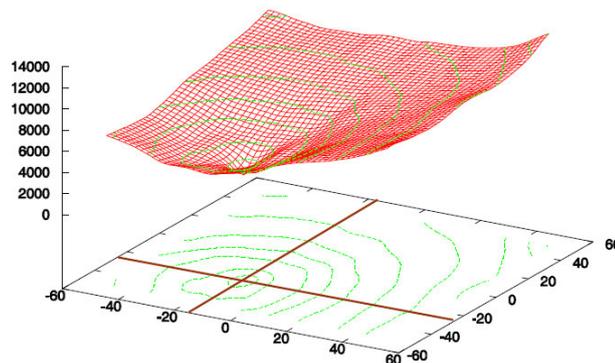
$$E(x|F, M, T) = \frac{1}{N} \sum_i^N (F(x_i) - M(x'_i))^2 \quad (6)$$

3.4.1.4 Otimizador

O otimizador busca os parâmetros da transformação que produzem o melhor alinhamento entre a imagem fixa e móvel com base na métrica de similaridade. Os métodos mais usuais são gradiente descendente, o gradiente descendente de passo regular, o quasi-Newton, entre outros.

O gradiente descendente de passo regular é o método utilizado neste trabalho. O método é uma variante do gradiente descendente. O mesmo consiste em minimizar uma função objetivo, que neste caso é o erro médio quadrático (EMQ). O gráfico do EMQ possui o comportamento da Figura 3.10, caso o número de parâmetros seja dois. Com base no espaço dos parâmetros, direções são obtidas através do gradiente do EMQ, de maneira a minimizar o erro.

Figura 3.10 – Comportamento do gráfico do erro médio quadrático.



Fonte: [Stewart e Ibanez \(2004\)](#)

O método parte de um ponto inicial x_0 e então é calculado o gradiente nesse ponto. Como o gradiente é a direção de maior variação para EMQ, o novo ponto x é somado com o negativo do gradiente. Em adição, uma taxa de aprendizado α é utilizada na multiplicação do gradiente obtido, limitando o tamanho do deslocamento do ponto inicial. O resultado

do produto é chamado de etapa ou passo e pode ser visto na Equação 7, que possui o valor de x em uma iteração k (GONÇALVES, 2011).

$$x^{k+1} = x^k - \alpha \cdot \nabla EMQ(x^k) \quad (7)$$

Dependendo do valor do α , o tamanho do passo pode ser grande ou pequeno e isso pode prejudicar a localização do valor ótimo para o EMQ. Para resolver esse problema, o método foi modificado de maneira que caso ocorra uma mudança de direção, a taxa de aprendizagem se reduzirá pela metade, de maneira a reduzir o tamanho do passo e melhorar a localização do mínimo.

3.5 Características de textura

A caracterização de objetos a partir de suas propriedades mais perceptíveis é algo que continuamente se faz, e mesmo o computador, não é diferente. Entretanto, tem-se o papel de selecionar as características que o computador utilizará na tarefa de discriminação entre os objetos, mas, para isso, algum método de aprendizagem de máquina precisa ser utilizado. As propriedades mais utilizadas na caracterização de lesões pulmonares são baseadas na forma ou geometria e na textura. Entretanto, como se está avaliando apenas mudanças de densidade ao longo do tempo, apenas algumas características clássicas de textura serão apresentadas a seguir.

A textura é algo que intuitivamente se consegue observar, contudo, nenhum conceito preciso foi dado a ela. Textura é um conjunto de elementos mutuamente relacionados formando blocos de textura que são agrupados em primitivas chamadas de *texels* (*texture elements*) e dependente da escala. Pode ser descrita pela sua tonalidade ou estrutura espacial (SONKA; HLAVAC; BOYLE, 1998).

A tonalidade está relacionada às propriedades do pixel, enquanto que a estrutura descreve a espacialidade das primitivas. Quando a textura é fraca, pode-se descrevê-la através das frequências dos tipos de primitivas que aparecem na vizinhança. Já a textura forte, pode ser avaliada melhor por meio da avaliação da frequência dos pares de primitivas por possuir definição exata (SONKA; HLAVAC; BOYLE, 1998).

Uma melhor descrição das texturas pode ser dada pela estatística. A estatística de descrição da textura tem como objetivo medir as propriedades lineares das primitivas,

usando o pixel para avaliar suas correlações, e “dizer” se a mesma é uma textura grossa com largas primitivas e baixas frequências ou textura fina de primitivas menores e altas frequências. Essas correlações são feitas por diversos cálculos, como o coeficiente de autocorrelação, a análise de Fourier e a matriz de coocorrência (SONKA; HLAVAC; BOYLE, 1998).

A textura será utilizada neste trabalho na análise da relação do seu diagnóstico com as mudanças das imagens da lesão pulmonar.

3.5.1 Matriz de coocorrência

As medidas de textura foram calculadas sobre a matriz de coocorrência (HARALICK; SHANMUGAM; DINSTEN, 1973) do volume, também chamada de características de Haralick 3D, que é encontrada segundo a Equação 8, onde v é a função que retorna um dos 26 vizinhos tridimensionais de um *voxel* de acordo com o índice α e P_{xyz} é o valor do *voxel* de coordenadas x, y, z .

$$Co(i, j) = \{quant(P_{xyz}, v(P_{xyz}, \alpha)) \mid P_{xyz} = i, v(P_{xyz}, \alpha) = j, \alpha \in \{1, \dots, 26\}\} \quad (8)$$

3.5.1.1 Inércia

A inércia é a medida da quantidade de variação local de níveis de cinza em uma imagem. É dada pela Equação 9, na qual G é a quantidade de níveis de cinza e $Co(i, j)$ é o elemento da matriz de coocorrência situado na linha i e coluna j .

$$Iner = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} Co(i, j) (i - j)^2 \quad (9)$$

3.5.1.2 Energia

A energia, dada pela Equação 10, é a medida de não uniformidade de uma imagem. Em uma imagem uniforme, existem poucas transições de níveis de cinza. Nesse caso, a matriz de coocorrência possui baixas entradas de alta magnitude.

$$Ene = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} Co(i, j)^2 \quad (10)$$

3.5.1.3 Entropia

A entropia ou grau de dispersão de níveis de cinza mede a quantidade de variação local da imagem. A entropia é calculada pela equação:

$$Ent = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} -Co(i, j) \ln(Co(i, j)) \quad (11)$$

3.5.1.4 Homogeneidade

Quanto menor a variação dos valores dos pixels da imagem, maior é o valor da homogeneidade, que é obtida através da equação:

$$Hom = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{Co(i, j)}{1 + (i - j)^2} \quad (12)$$

3.5.1.5 Tom do grupo - *Cluster shade*

É uma medida de enviesamento da matriz Co e um grande valor significa uma grande assimetria. É calculada pela equação:

$$Tom = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} Co(i, j) \left(i + j - \sum_{i=0}^{G-1} iCo(i, j) - \sum_{j=0}^{G-1} jCo(i, j) \right)^3 \quad (13)$$

3.5.1.6 Proeminência - *Cluster prominence*

É considerada também uma medida de assimetria. Um grande valor significa menor simetria ou maior achatamento, enquanto que um valor pequeno significa a presença de um grande pico em torno da média. É calculada pela equação:

$$Pro = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} Co(i, j) \left(i + j - \sum_{i=0}^{G-1} iCo(i, j) - \sum_{j=0}^{G-1} jCo(i, j) \right)^4 \quad (14)$$

3.5.2 Matriz de comprimento de nível de cinza

É denominada na literatura por *run-length (RL)* (CHU; SEHGAL; GREENLEAF, 1990). Uma primitiva *run-length* é um conjunto colinear de *pixels* conectados contendo todos o mesmo nível de cinza. Tais primitivas podem ser caracterizadas basicamente pelo seu comprimento, sua inclinação e o seu nível de cinza.

As matrizes são do tipo $M_\theta(a, r)$, onde θ é o ângulo de inclinação das primitivas e cada elemento (a, r) indica o número de vezes que uma primitiva de nível de cinza a e comprimento r ocorre dentro da imagem, conforme a Equação 15.

$$M_\theta(a, r) = \text{quant} \{ (x, y) \mid v(x, y) = v(x + s, y + t) = a, v(x + u, y + k) \neq a \}, \\ 0 \leq s < u \ \& \ 0 \leq t < k, u = r \cos \theta \ \& \ k = r \sin \theta \quad (15)$$

Geralmente, várias matrizes são computadas para diversos ângulos, sendo que os mais utilizados são $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$ e simétricos. A dimensão das matrizes é dada por $LxNr$, onde L é o número de níveis de cinza da imagem e Nr é o número de comprimentos distintos de primitivas.

3.5.2.1 Ênfase em primitivas curtas (*Short-run emphasis* - (SRE))

Um valor alto significa grandes quantidades de primitivas curtas com o mesmo coeficiente de atenuação. É definida por:

$$SRE = \frac{1}{K} \sum_{a=0}^{G-1} \sum_{r=1}^{N_r} \frac{M(a, r)}{r^2} \quad (16)$$

3.5.2.2 Ênfase em primitivas longas (*Long-run emphasis* - (LRE))

Enfatiza as primitivas longas. Um valor alto significa grandes quantidades de primitivas longas com o mesmo coeficiente de atenuação. É definida pela equação:

$$LRE = \frac{1}{K} \sum_{a=0}^{G-1} \sum_{r=1}^{N_r} M(a, r)r^2 \quad (17)$$

onde M é a matriz, a é o nível de cinza, r é a primitiva e K é o número total de primitivas na matriz.

3.5.2.3 Não-uniformidade do nível de cinza (*Gray-level nonuniformity* - (GLNU))

Mede a não-uniformidade do nível de cinza e quanto maior seu valor, maior é a presença de grandes valores em M ; enquanto que um pequeno valor é resultado de um valor de M igualmente distribuído com o nível de cinza. É dada pela equação:

$$GLNU = \frac{1}{K} \sum_{r=1}^{N_r} \left[\sum_{a=0}^{G-1} M(a, r) \right]^2 \quad (18)$$

3.5.2.4 Não-uniformidade do comprimento da primitiva (*Run-Length Nonuniformity* - (RLN))

Mede a não-uniformidade das primitivas. Se as primitivas são igualmente distribuídas ao longo do seu comprimento, a função terá valores baixos. Primitivas longas contribuem mais para a função. É dada pela equação:

$$RLN = \frac{1}{K} \sum_{a=0}^{G-1} \left[\sum_{r=1}^{N_r} M(a, r) \right]^2 \quad (19)$$

3.5.2.5 Ênfase de primitiva de pequeno nível de cinza (*Low Grey Level Run Emphasis* - (LGRE))

É sensível a presença de muitas primitivas de pequeno nível de cinza. É dada pela equação:

$$LGRE = \frac{1}{K} \sum_{r=1}^{N_r} \sum_{a=0}^{G-1} \frac{M(a, r)}{(a+1)^2} \quad (20)$$

3.5.2.6 Ênfase de primitiva de grande nível de cinza (*High Grey Level Run Emphasis* - (HGLRN))

É sensível a presença de muitas primitivas de pequeno nível de cinza. É dada pela equação:

$$HGLRN = \frac{1}{K} \sum_{r=1}^{N_r} \sum_{a=0}^{G-1} M(a, r) (a+1)^2 \quad (21)$$

3.5.2.7 Ênfase em primitivas curtas de pequeno nível de cinza (*Short Run, Low Grey Level Emphasis* - (SRLGLE))

É uma medida que combina características do *SRE* e do *LGRE* e é sensível à presença de primitivas curtas de pequeno nível de cinza. É dada pela equação:

$$SRLGLE = \frac{1}{K} \sum_{a=0}^{G-1} \sum_{r=1}^{N_r} \frac{M(a, r)}{r^2 (a+1)^2} \quad (22)$$

3.5.2.8 Ênfase em primitivas curtas de grande nível de cinza (*Short Run, High Grey Level Emphasis* - (SRHGLE))

É uma medida que combina características do *SRLGE* e do *LRHGE* e é sensível à presença de primitivas curtas de grande nível de cinza. É dada pela equação:

$$SRHGLE = \frac{1}{K} \sum_{a=0}^{G-1} \sum_{r=1}^{N_r} \frac{M(a, r) (a+1)^2}{r^2} \quad (23)$$

3.5.2.9 Ênfase em primitivas longas de pequeno nível de cinza (*Long Run, Low Grey Level Emphasis* - (LRLGLE))

É sensível à presença de primitivas longas de grande nível de cinza. É dada pela equação:

$$LRLGLE = \frac{1}{K} \sum_{a=0}^{G-1} \sum_{r=1}^{N_r} \frac{M(a, r)r^2}{(a+1)^2} \quad (24)$$

3.5.2.10 Ênfase em primitivas longas de grande nível de cinza (*Long Run, High Grey Level Emphasis* - (LRHGLE))

É sensível à presença de primitivas longas de grande nível de cinza. É dada pela equação:

$$LRHGLE = \frac{1}{K} \sum_{a=0}^{G-1} \sum_{r=1}^{N_r} M(a, r)r^2 (a+1)^2 \quad (25)$$

3.5.3 Medidas estatísticas

As medidas utilizadas para a caracterização da lesão quanto ao diagnóstico foram: média da diferença, média da diferença dos valores positivos, média da diferença dos valores negativos e máxima diferença sobre a região de mudança.

A média da diferença da região de mudança é a média dos valores da diferença dos *voxels* correspondentes da lesão em diferentes instantes e é dada pela Equação 26. N é o número de *voxels* da região onde as mudanças ocorreram e $voxVal(t)$ é o valor do *voxel* no instante t .

$$MD = \frac{\sum_{i=1}^N [voxVal(t_2) - voxVal(t_1)]}{N} \quad (26)$$

A média da diferença dos valores positivos considera apenas os valores positivos da diferença dos *voxels* e é dada pela Equação 27. N_p é o número de *voxels* positivos da região onde as mudanças ocorreram.

$$MPD = \frac{\sum_{i=1}^{N_p} [voxVal(t_2) - voxVal(t_1)]}{N_p} \quad (27)$$

A média da diferença dos valores negativos considera apenas os valores negativos da diferença dos *voxels* e é dada pela Equação 28. N_n é o número de *voxels* negativos da região onde as mudanças ocorreram.

$$MND = \frac{\sum_{i=1}^{N_n} [voxVal(t_2) - voxVal(t_1)]}{N_n} \quad (28)$$

A máxima diferença é definida pelo maior valor da diferença dos *voxels* e é dada pela Equação 29.

$$MAXD = \max_{i=1, \dots, N} [voxVal(t_2) - voxVal(t_1)] \quad (29)$$

3.6 Reconhecimento de padrões

O Reconhecimento de Padrões (RP) é realizado em muitas áreas de aplicação – navegações, migrações, identificação de predadores, etc. – com razoável eficiência. O desenvolvimento de métodos capazes de emulação das mais variadas formas de reconhecimento de padrões tem evoluído ao longo do tempo. Esse desenvolvimento surge da necessidade de construção de sistemas automatizados inteligentes, que estão se tornando a principal tendência na indústria e em outros campos de atividades semelhantes. Nesses sistemas, os objetos são representados de modo adequado ao tipo de processamento no qual estão sujeitos. Tais representações são chamadas de padrões, isto é, objetos e padrões são intercambiáveis com similar significado (Sá, 2001).

O RP é um conjunto de métodos responsáveis pela descrição e classificação de objetos. Desta forma, muitos são os métodos desenvolvidos, mas a pesquisa ainda encontra-se fértil. Desde a análise em sinais, imagens, tabelas de valores e etc.

Uma noção muito útil no reconhecimento de padrões é o da similaridade. Os objetos são reconhecidos pela similaridade de atributos comuns. A similaridade é estabelecida, em termos de um conceito mais abstrato, pela associação entre um objeto e um conceito alvo. Por exemplo, reconhece-se uma maçã pela sua imagem e pelo conjunto de características idealizadas, formando um protótipo e, através deste protótipo, pode-se distingui-la de uma laranja (Sá, 2001).

A avaliação da similaridade de padrões está fortemente relacionada à tarefa de reconhecimento de padrões propostos. Logo, extrai-se características dos objetos que se

quer reconhecer, com o objetivo de gerar protótipos que facilitem classificação dos objetos em classes (variável categórica). Em geral, essa classificação não é perfeita e erros por muitas causas podem ocorrer:

- As características usadas são inadequadas ou insuficientes;
- Os padrões simples usados para projetar o classificador não são suficientemente representativos;
- O classificador não é eficiente o bastante para separar as classes por possuir inadequados protótipos;
- Há uma intrínseca sobreposição das classes que nenhum classificador pode resolver.

O RP objetiva o mapeamento do espaço de representação (vetor de características) ao espaço de interpretação (vetor de protótipos ou classes) e é chamado de hipótese (Sá, 2001).

A hipótese pode ser:

- Supervisionada, dirigida a conceito ou hipótese indutiva: encontra no espaço de representação uma hipótese correspondente à estrutura do espaço de interpretação;
- Não supervisionada, dirigida a dados ou hipótese dedutiva: encontra uma estrutura no espaço de interpretação correspondente à estrutura no espaço de representação.

A hipótese é derivada do uso de métodos de aprendizagem, que podem ser estatístico, de aproximação (minimização do erro) ou de natureza estrutural.

A tarefa de reconhecimento requer um conjunto de etapas:

- Aquisição do padrão, que pode ter muitas formas: aquisição de um sinal ou imagem e coleção de dados;
- Extração de características, na forma de medidas, extração de primitivas, etc;
- Pré-processamento. Em alguns casos, os valores não estão diretamente prontos para o uso em alguns classificadores. Por exemplo, em algumas aplicações de rede neurais, é usual padronizar as características em algum modo (colocando os dados no intervalo $[0,1]$);
- A classificação propriamente dita pelo uso de um núcleo de classificação.

3.6.1 Máquina de Vetores de Suporte

A Máquina de Vetores de Suporte (MVS) ([VAPNIK, 1998](#)) é um método de aprendizagem supervisionada usado para estimar uma função que classifique dados de entrada em duas classes. A ideia básica do funcionamento da MVS é construir um hiperplano como superfície de decisão, de tal maneira que a margem de separação entre as classes seja máxima. O objetivo do treinamento através de MVS é a obtenção de hiperplanos que dividam as amostras de modo que sejam otimizados os limites de generalização.

As MVS são consideradas sistemas de aprendizagem que utilizam um espaço de hipóteses de funções lineares em um espaço de muitas dimensões. Os algoritmos de treinamento das MVS possuem forte influência da teoria de otimização e de aprendizagem estatística. Em poucos anos, as MVS vêm demonstrando sua superioridade frente a outros classificadores em uma grande variedade de aplicações ([CRISTIANINI; SHAWE-TAYLOR, 2000](#)).

Em casos em que o conjunto de amostras é composto por duas classes separáveis, um classificador MVS é capaz de encontrar um hiperplano baseado em um conjunto de pontos denominados “vetores de suporte”, o qual maximiza a margem de separação entre as classes. Por hiperplano, entende-se uma superfície de separação de duas regiões num espaço multidimensional, onde o número de dimensões possíveis pode ser infinito. Mesmo quando as duas classes não são separáveis, a MVS é capaz de encontrar um hiperplano através do uso de conceitos pertencentes a teoria da otimização. O hiperplano ótimo (linha central) não somente separa as duas classes, mas mantém a maior distância possível com relação aos pontos da amostra.

Dado o conjunto amostral de treinamento (x_i, y_i) , sendo $x_i \in R^N$ o vetor de entrada, y_i a classificação correta das amostras e $i = 1, \dots, n$ o índice de cada ponto amostral. O objetivo da classificação é estimar a função $f : R^N \rightarrow \{\pm 1\}$, que separe corretamente os exemplos de teste em classes distintas.

A etapa de treinamento estima a função $f(x) = (w \cdot x) + b$, procurando pelos valores do vetor w e da variável b , tais que a seguinte relação seja satisfeita:

$$y_i ((w \cdot x_i) + b) \geq 1 \quad (30)$$

onde w é o vetor normal ao hiperplano de decisão e b é o corte ou distância da função f em relação a origem. Os valores ótimos de w e b são obtidos ao minimizar a seguinte equação, de acordo com a restrição dada pela Equação 30 (BISHOP, 2006):

$$\phi(w) = \frac{w^2}{2} \quad (31)$$

MVS ainda possibilita encontrar um hiperplano que minimize a ocorrência de erros de classificação nos casos em que uma perfeita separação entre as duas classes não seja possível. Isso graças a inclusão de variáveis de folga, que permitem que as restrições presentes na Equação 30 sejam obtidas.

O problema de otimização passa a ser, então, a minimização da Equação 31, de acordo com a restrição imposta pela Equação 30. A variável C é um parâmetro de treinamento que estabelece um equilíbrio entre a complexidade do modelo e o erro de treinamento. Seu valor deve ser selecionado pelo usuário. Assim, define-se o seguinte sistema:

$$\phi(w, \xi) = \frac{w^2}{2} + C \sum_{i=1}^N \xi_i \quad (32)$$

para

$$y_i((w \cdot x_i) + b) + \xi_i \geq 1 \quad (33)$$

Através da teoria dos multiplicadores de Lagrange, chega-se a Equação 34. O objetivo é encontrar os multiplicadores de Lagrange α_i ótimos que satisfaçam a Equação 32:

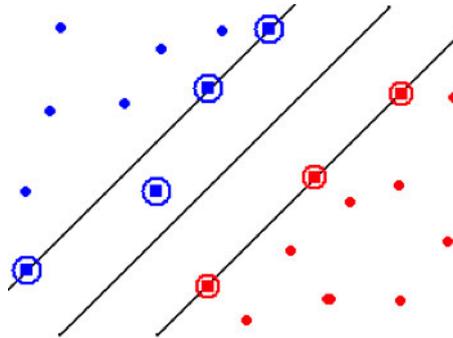
$$w(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j x_i x_j \phi(x_i x_j) \quad (34)$$

sujeito a

$$\sum_{i=1}^N \alpha_i x_i = 0, \quad 0 \leq \alpha_i \leq C \quad (35)$$

Apenas os pontos da restrição presente na Equação 30 sejam exatamente iguais a unidade têm correspondentes $\alpha \neq 0$. Esses pontos são chamados de vetores de suporte, pois se localizam geometricamente sobre as margens. Tais pontos possuem fundamental importância na definição do hiperplano ótimo, pois os mesmos delimitam a margem do conjunto de treinamento.

Figura 3.11 – Vetores de Suporte destacados por círculos.



Fonte: Junior et al. (2009)

A Figura 3.11 destaca os pontos que representam os vetores de suporte. Os pontos além da margem não influenciam decisivamente na determinação do hiperplano, enquanto que os vetores de suporte, por terem pesos não nulos, são decisivos.

Para que a MVS possa classificar amostras que não são linearmente separáveis, é necessária uma transformação não-linear que transforme o espaço entrada (dados) para um novo espaço (espaço de características). Esse espaço deve apresentar dimensão suficientemente grande, e através dele, a amostra pode ser linearmente separável. Assim, o hiperplano de separação é definido como uma função linear de vetores retirados do espaço de características ao invés do espaço de entrada original. Essa construção depende do cálculo de uma função K de núcleo de um produto interno (HAYKIN; ENGEL, 2001). A função K pode realizar o mapeamento das amostras para um espaço de dimensão muito elevada sem aumentar a complexidade dos cálculos.

Substituindo a Equação 34 pela equação a seguir, com a utilização do núcleo, tem-se:

$$w(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j K(x_i x_j) \quad (36)$$

Uma importante família de funções de núcleo é a função de base radial, muito utilizada em problemas de reconhecimento de padrões e também utilizada neste trabalho. A função de base radial é definida por:

$$R(x_i, x_j) = e^{-\gamma(x_i - x_j)^2} \quad (37)$$

onde γ é um parâmetro informado externamente. A variável C representa a penalidade por um erro durante a classificação.

Em conjunto com as características de textura, o MVS será utilizado na tarefa de classificação de lesões pulmonares.

3.7 Validação do resultado da classificação

Fazer reconhecimento de padrão é um processo imperfeito que resulta mais em probabilidade de se estar certo do que em certeza. Embora possam existir várias medidas para verificar o desempenho de um classificador qualquer, a medida mais importante é o desempenho do mesmo a partir da classificação de novos casos (conjunto de testes). O desempenho do classificador, medido através do conjunto de teste, é uma boa indicação de seu desempenho real.

Em problemas ligados à área de saúde, a estrutura básica dos testes de classificação é para determinar quão bem um teste discrimina a presença ou ausência de uma doença. Nesses tipos de problemas, existe a presença de uma variável preditora (resultado do teste) e uma variável resultante (a presença ou ausência da doença).

A variável preditora do teste é geralmente dividida em duas, definindo os critérios de corte que confirmam ou não a presença da doença. A variável resultante do teste diagnóstico é a presença ou ausência da doença, confirmada por um exame considerado "padrão ouro".

Quando se avalia um teste diagnóstico, quatro situações são possíveis:

1. O teste é positivo e o paciente tem a doença - Verdadeiro Positivo (vp);
2. O teste é positivo, mas o paciente não tem a doença - Falso Positivo (fp);
3. O teste é negativo e o paciente tem a doença - Falso Negativo (fn); e
4. O teste é negativo e o paciente não tem a doença - Verdadeiro Negativo (vn).

Geralmente, utilizam-se algumas estatísticas descritivas para avaliar o desempenho do classificador, como Sensibilidade (sen), Especificidade (esp) e Acurácia (ac).

A sensibilidade de um teste é definida pela proporção de pessoas com a doença de interesse que têm o resultado do teste positivo. Indica quão bom é o teste para identificar os indivíduos doentes:

$$sen = \frac{vp}{vp + fn} \quad (38)$$

A especificidade de um teste é a proporção de pessoas sem a doença que têm o teste negativo. Indica quão bom é o teste para identificar os indivíduos não doentes:

$$esp = \frac{vn}{vn + fp} \quad (39)$$

A taxa de classificação correta (Acurácia) é definida como a razão entre o número de casos na amostra em estudo que foram classificados corretamente e o número total de casos na amostra em estudo:

$$ac = \frac{vp + vn}{vp + vn + fp + fn} \quad (40)$$

V_1

3.8 Medidas de Quantificação das Mudanças

A variação volumétrica relativa (VVR) é a razão entre a variação de volume entre dois instantes e o volume do instante inicial. Esta medida é utilizada para avaliar as mudanças do volume da lesão ao longo do tempo e é definida como:

$$VVR = \frac{(V_2 - V_1)}{V_1} \quad (41)$$

O percentual de mudança P_m é a razão do número de *voxels* que mudaram (N_m) pelo número total de *voxel* da imagem N_T (Equação 42).

$$P_m = \frac{N_m}{N_T} \quad (42)$$

Este trabalho utiliza o P_m para quantificar as mudanças detectadas por cada um dos métodos baseados em *voxel*, para a análise de componentes principais, para a abordagem local usando o teste de hipótese e para o método com QT.

As modificações intra-grupo foram determinadas pelo percentual de mudança (P_i) para o grupo i . Essa medida quantifica as mudanças detectadas pelos grupos correspondentes da lesão no par de instantes. A medida é definida como a razão do número de *voxels* do grupo i no instante t_2 (N_2) e o número de *voxels* no instante t_1 (N_1), conforme a Equação 43.

$$P_i = \frac{N_2}{N_1} \quad (43)$$

Outra forma de quantificar as mudanças de densidade encontradas é através da comparação do número de *voxels* do grupo i ($N_i(t)$) e o número total de *voxels* do volume (N_T), denominada de relevância ($R_i(t)$). A relevância quantifica o percentual de *voxels* do grupo i do instante t em relação ao volume.

$$R_i(t) = \frac{N_i(t)}{N_T} \quad (44)$$

4 Métodos baseados em *voxel*

É um conjunto de métodos cujas operações são realizadas a nível de *voxel*. Todos os métodos pertencentes a este conjunto foram derivados da análise de mudanças em imagens aéreas para sensoriamento remoto (İLSEVER; ÜNSALAN, 2012): 1) diferença entre imagens; 2) razão de imagens; 3) regressão de imagens; 4) operador XOR fuzzy; e 5) mediana das imagens. Portanto, esse conjunto foi adaptado para identificação das mudanças de densidade em lesões pulmonares.

4.1 Diferença entre imagens

O método de diferença entre imagens visa detectar variações de *voxels* entre duas imagens de um mesmo objeto, obtidas em instantes diferentes. Com isso, o método localiza as regiões onde ocorreram mudanças significativas no valor de *voxel*, indicando que essas regiões foram as que mais sofreram mudanças ao longo do tempo.

As etapas do algoritmo e as adaptações realizadas para este trabalho são descritas abaixo:

- Etapa 1:

Esta etapa é realizada independente da dimensão dos dados de entrada para o método, isto é, independente da imagem ser 2D ou 3D.

A operação de diferença absoluta *voxel a voxel* é aplicada para o volume inteiro nas duas imagens 3D obtidas em instantes de tempo t_a e t_b , conforme a Equação 45:

$$I_d(x, y, z) = |I_b(x, y, z) - I_a(x, y, z)| \quad (45)$$

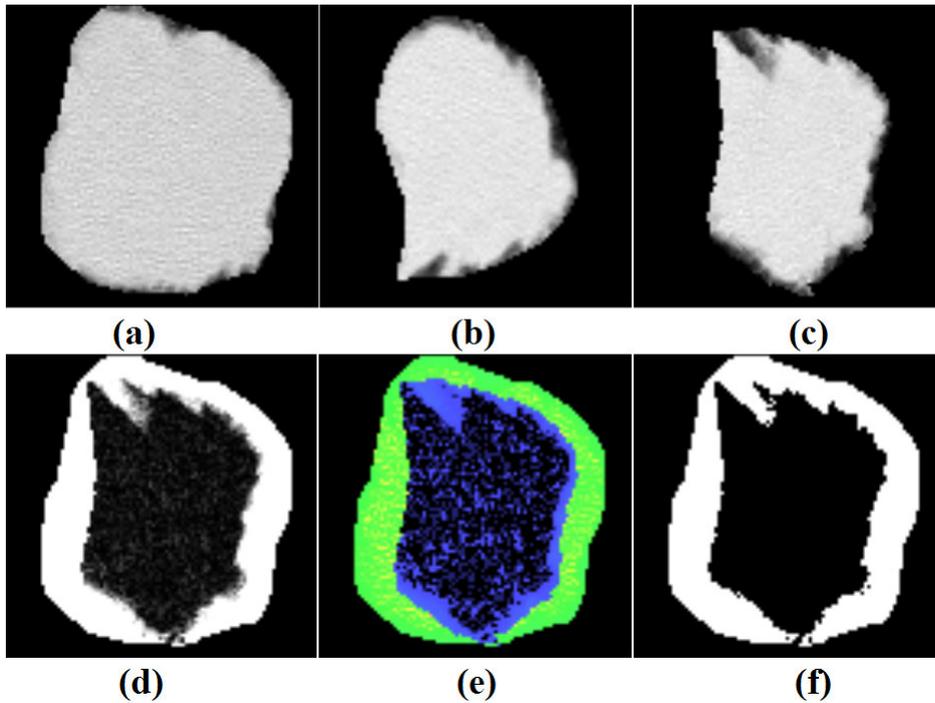
onde $I_a(x, y, z)$ e $I_b(x, y, z)$ são os *voxels* das imagens obtidas nos instantes t_a e t_b , respectivamente, e $I_d(x, y, z)$ são os valores de *voxel* da imagem resultante da diferença absoluta.

Um exemplo da imagem resultante da diferença absoluta da imagem no instante t_a com a imagem no instante t_b é dado na Figura 4.1(d) e a correspondente imagem colorida é dada na Figura 4.1(e).

- Etapa 2:

No trabalho de İlsever e Ünsalan (2012), antes da aplicação da Equação 45, existe a normalização da imagem obtida no instante t_b com relação a imagem obtida

Figura 4.1 – Imagens representativas das lesões pulmonares. Em (a), tem-se uma lesão pulmonar obtida no instante t_a . Em (b), tem-se a imagem em t_b . Em (c), tem-se a imagem em (b) registrada com a imagem em (a). Em (d) tem-se a imagem da diferença absoluta entre a lesão em (a) e a lesão em (c). Uma imagem colorida da imagem diferença em (d) é dada em (e). A imagem em (f) é a imagem binária que realça as mudanças encontradas entre as imagem em (a) e (c).



Fonte: Autor

em t_a pela Equação 46. Entretanto, escolheu-se não normalizá-la, pois a imagem normalizada torna-se muito semelhante à imagem obtida no instante t_a . Além disso, deseja-se nenhuma interferência sobre os valores dos *voxels*.

$$\tilde{I}_b(x, y, z) = \frac{\sigma_a}{\sigma_b} (I_b(x, y, z) - \mu_b) + \mu_a \quad (46)$$

onde μ_a , σ_a e μ_b , σ_b são as médias e os desvios padrões das imagens $I_a(x, y, z)$ e $I_b(x, y, z)$, respetivamente.

- Etapa 3:

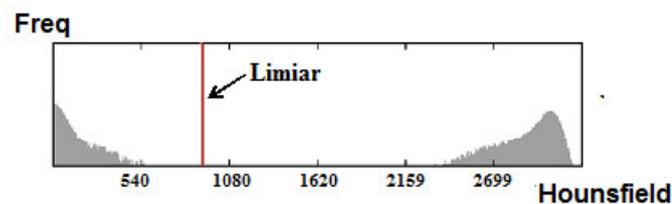
Esta etapa consiste em encontrar um limiar T , que indica o que realmente se tornou diferente entre as imagens t_a e t_b das lesões pulmonares. Esse limiar é obtido através da aplicação do método de Otsu (OTSU, 1979) para limiarização do histograma da imagem diferença absoluta. Um exemplo do histograma e limiar encontrado da imagem diferença absoluta da Figura 4.1(d) pode ser visto na Figura 4.2. A partir desse limiar, pode-se binarizar a imagem diferença absoluta. Essa binarização

evidencia as grandes mudanças (grande variação do valor do *voxel*) e as pequenas mudanças (pequena variação do valor do *voxel*), exemplificada na Figura 4.1(f). Entretanto, em vez de se encontrar um limiar, achou-se mais robusto binarizar a imagem através do algoritmo de agrupamento *k-means* (KANUNGO et al., 2002) com dois grupos. Conseqüentemente, um grupo evidencia as grandes mudanças e o outro as pequenas mudanças. A modificação foi realizada devido ao limiar encontrado pela técnica de limiarização de Otsu, em alguns casos, não separar estas regiões.

- Etapa 4:

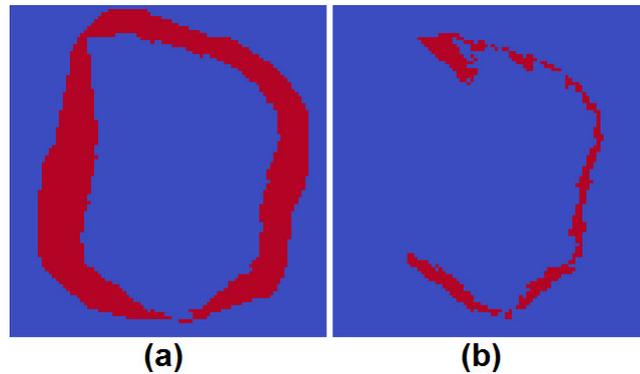
Este método foi modificado quanto a sua aplicação. O método tem o papel de detectar dois tipos de mudanças de densidade. O primeiro tipo envolve a detecção das mudanças lesão-fundo. Esse tipo de mudança é caracterizado pelas mudanças de densidade que ocorreram entre a densidade do parênquima ou do próprio fundo da imagem e as densidades da lesão. Ou seja, as densidades da lesão são aquelas cujos valores são diferentes do parênquima e do próprio fundo da imagem. Semelhantemente, o segundo tipo envolve a detecção das mudanças lesão-lesão. Esse tipo de mudança é o mais importante, pois está relacionado com *voxels* que permaneceram ao longo do tempo, mas mudaram de tecido. Um exemplo da aplicação das duas aplicações está na Figura 4.3, onde em (a) tem-se o resultado da primeira e em (b) o da segunda aplicação. Desta forma, o método é aplicado duas vezes, onde a primeira vez detecta as mudanças de densidade lesão-fundo e a segunda as mudanças lesão-lesão, até que não reste nenhum *voxel* para a análise.

Figura 4.2 – Histograma da imagem resultante da aplicação do método da diferença absoluta.



Fonte: Autor

Figura 4.3 – Mudanças de densidade detectadas pelo método da diferença da Figura 4.1. Em (a), têm-se as mudanças lesão-fundo e (b), têm-se as mudanças lesão-lesão.



Fonte: Autor

4.2 Razão de imagens

Da mesma forma que o método da diferença entre imagens, o objetivo do método da razão de imagens é realçar as mudanças sofridas pelos *voxels* entre duas imagens da mesma cena, obtidas em instantes diferentes. Esse método encontra a fase da razão entre os valores dos *voxels* das imagens. A fase indica o quão semelhantes são os *voxels* correspondentes entre as duas imagens.

As etapas do algoritmo e as adaptações realizadas para este trabalho são descritas abaixo:

- Etapa 1:

O método da razão de imagens, originalmente, utiliza a Equação 47, onde $I_a(x, y, z)$ e $I_b(x, y, z)$ são os *voxels* das imagens obtidas nos instantes t_a e t_b , respectivamente; e $I_r(x, y, z)$ são os *voxels* da imagem resultante da razão dos valores dos *voxels*, enquanto a imagem $\tilde{I}_b(x, y)$ é a imagem $I_b(x, y)$ normalizada, conforme a Equação 46.

$$I_r(x, y) = \arctan \left(\frac{I_a(x, y)}{\tilde{I}_b(x, y)} \right) - \frac{\pi}{4} \quad (47)$$

$$I_r(x, y, z) = \arctan \left(\left| \frac{I_a(x, y, z)}{I_b(x, y, z)} \right| \right) \quad (48)$$

A Equação 47 foi alterada de maneira que o arco tangente do módulo da razão fosse calculado. Dessa forma, os valores obtidos pela Equação 48 estão no intervalo de $[0, \pi/2]$. Um exemplo de resultado está na Figura 4.4(a) e realçada através da aplicação de cores na imagem da Figura 4.4(c). Valores de *voxel* da imagem I_r

próximos de $\pi/4$ pertencem aos *voxels* semelhantes entre as imagem em t_a e t_b . De maneira oposta, valores de *voxel* da imagem I_r próximos de zero radianos ou $\pi/2$ pertencem aos *voxels* cujos valores nas imagens em t_a e t_b são diferentes nas suas coordenadas (x, y, z) correspondentes.

- Etapa 2:

Novamente, não houve a normalização da imagem em t_b pelo motivo já explicado no método de diferença entre imagens.

- Etapa 3:

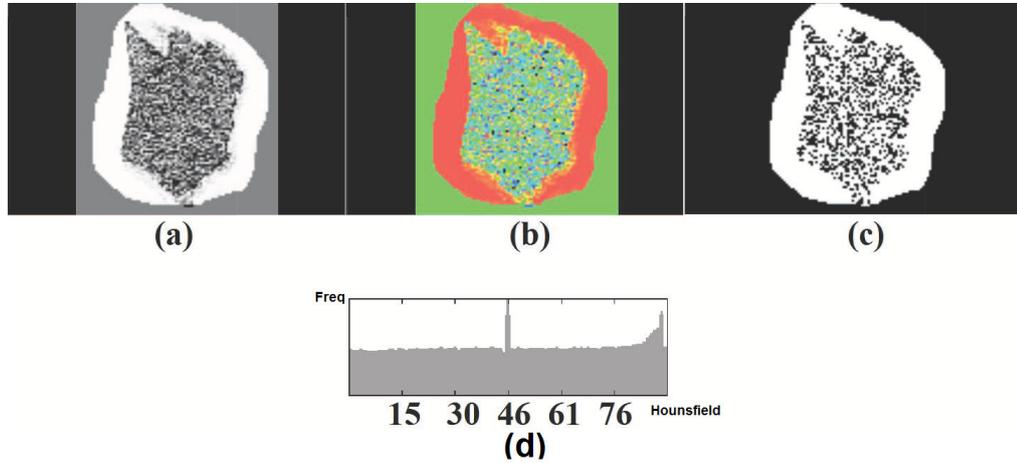
Esta etapa tem a função de binarizar a imagem em grande mudança e em pequena mudança no valor do *voxel*. Originalmente, essa etapa utiliza a técnica de limiarização proposta por Otsu ou Kapur sobre o histograma (Figura 4.4(d)) da imagem I_r . Com intuito de melhorar essa etapa, modificou-se a mesma. Usou-se um limiar inicial e final, que inclui $\pi/4$, para definir uma faixa de valores da razão, nos quais os *voxels* são considerados semelhantes. Portanto, tais limiares foram determinados de forma que o intervalo de variação do histograma da imagem razão fique dividido em 3 partes. Essas 3 partes representam, respectivamente, uma região ou grupo de semelhança, isto é, todos os *voxels* cujos valores da razão estão próximo de $\pi/4$ são semelhantes. As outras duas partes nos extremos do histograma da imagem razão formam um grupo onde se encontram os *voxels* com as maiores diferenças da imagem razão e geram a imagem binária da Figura 4.4(c).

- Etapa 4:

Este método não necessitou da dupla aplicação do mesmo. Isto é resultado da sua inerente capacidade de encontrar as diferenças de densidade. Portanto, o método é capaz de encontrar tanto as mudanças lesão-lesão quanto as mudanças lesão-fundo.

A Figura 4.4(a) mostra uma imagem resultante da aplicação deste método para as lesões pulmonares da Figura 4.1(a) e (c). Na Figura 4.4(a), verificam-se as mudança pontuais internas à lesão e na Figura 4.4(b), como ficou distribuído os valores de *voxel* da imagem razão pelo seu histograma. A Figura 4.4(c) possui o resultado do realce das mudanças entre as imagens em t_a e t_b aplicando a técnica de agrupamento *k-means* com dois grupos.

Figura 4.4 – Na Figura (a), tem-se a imagem resultante da razão das imagens da Figura 4.1(a) e (c). A imagem colorida da imagem em (a) é dada em (b). A imagem resultante do agrupamento da Figura (a) está na Figura (c). O histograma da imagem da Figura (a) é dado na Figura(d).



Fonte: Autor

4.3 Regressão de imagem

O método de regressão de imagem tem como objetivo gerar uma imagem de regressão, com base nas imagens de um mesmo objeto, obtidas em instantes diferentes. Essa imagem de regressão pode ser gerada em qualquer instante, entre os instantes inicial e final de aquisição das imagens. Então, a imagem de regressão é utilizada como parâmetro de comparação com qualquer outra imagem dada. Logo, encontra-se as diferenças entre os *voxels* das imagens.

- Etapa 1:

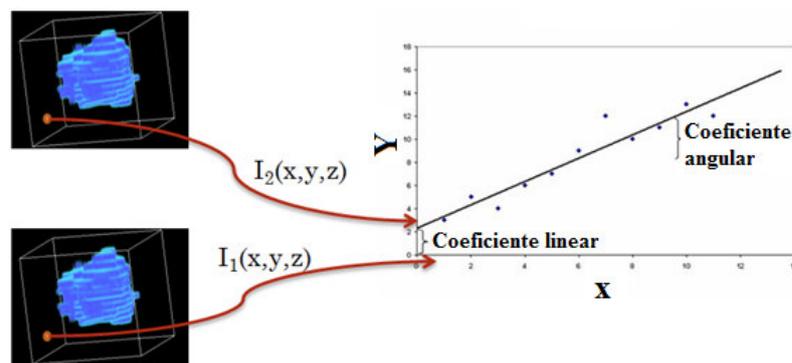
O método de regressão de imagem consiste em encontrar uma imagem \hat{I}_b em t_b estimada a partir da imagem I_a em t_a . A equação da reta relaciona os *voxels* nas coordenadas (x, y, z) da imagem estimada em t_b com os *voxels* correspondentes da imagem obtida em t_a . A mesma é obtida através do método de regressão linear por mínimo quadrático (İLSEVER; ÜNSALAN, 2012). Ilustra-se na Figura 4.5 como cada valor de *voxel* de coordenadas (x, y, z) da imagem I_b é colocado no eixo da variável dependente, para ser correlacionada com os valores de *voxels* de coordenadas (x, y, z) da imagem I_a , que estão no eixo independente. Então, obtém-se a reta de regressão.

O método de regressão linear encontra o coeficiente angular (a) e o coeficiente linear (b) da Equação 49, que relaciona os *voxels* das imagens I_a e I_b .

$$\hat{I}_b(x, y, z) = aI_a(x, y, z) + b \quad (49)$$

A Figura 4.5 ilustra a distribuição dos *voxels* da imagem em t_b (variável dependente) com os *voxels* da imagem em t_a . Nessa figura, visualiza-se que cada um dos N *voxels* pertencentes às lesões são exibidas como um ponto $(I_a(x, y, z), I_b(x, y, z))$, definindo uma nuvem de pontos a partir da qual será extraída a equação da reta de regressão. A partir dessa equação da reta, gera-se uma imagem regressão \hat{I}_b que pode ser comparada a imagem I_b .

Figura 4.5 – Imagem ilustrativa da relação linear entre os *voxels* da imagem em t_b (\hat{I}_b) com os *voxels* da imagem em t_a (\hat{I}_a).



Fonte: Autor

- Etapa 2:

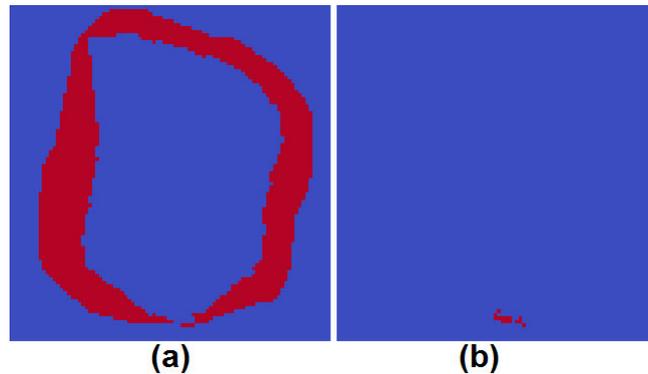
O passo anterior também foi aplicado duas vezes. As aplicações têm o objetivo de encontrar as mudanças lesão-fundo e lesão-lesão. O resultado da primeira aplicação está na Figura 4.6(a) e da segunda em (b).

4.4 Imagem Mediana

O método da imagem mediana consiste em gerar uma imagem que é a mediana dos valores de *voxel* das outras imagens que representam um objeto, em diversos instantes diferentes. A imagem mediana gerada é utilizada na comparação com todas as outras imagens dadas.

O método da imagem mediana foi aplicada às imagens 3D, da mesma forma que o método original, isto é, sem nenhuma mudança.

Figura 4.6 – Mudanças de densidade detectadas pelo método da regressão da Figura 4.5. Em (a), têm-se as mudanças lesão-fundo e (b), têm-se as mudanças lesão-lesão.



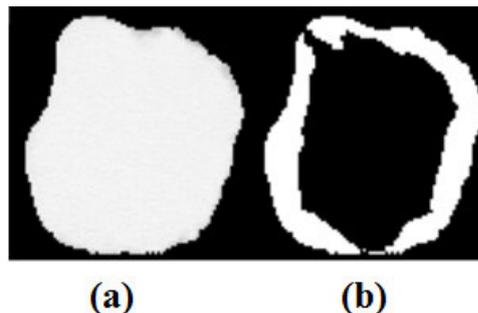
Fonte: Autor

Este método consiste em utilizar duas ou mais imagens $I_j, j = a, b, \dots, ab, \dots$, para a formação de uma imagem *background*, a partir da mediana dos valores de *voxel* de todas as imagens. A Equação 50 mostra como é obtida a imagem mediana resultante. A partir da imagem *background*, o método de diferença entre imagens deve ser aplicado a cada imagem I_j como um dos operandos e a imagem mediana como outro operando, seguindo os mesmos passos já descritos para o método de diferença entre imagens. Logo, para N imagens formadoras da imagem *background*, obtêm-se N imagens diferença.

$$I_{bg}(x, y, z) = Med(I_a(x, y, z), I_b(x, y, z), \dots, I_{ab}(x, y, z)) \quad (50)$$

Como se está interessado neste trabalho apenas na verificação de mudanças I_b da lesão pulmonar em diante, o número de imagens diferença será $(N - 1)$. A Figura 4.7 mostra a imagem *background* gerada pelas imagens das lesões em t_a e t_b da Figura 4.1(a) e (c), a imagem diferença entre I_b e I_{bg} , já binarizada, em grande mudança e pequena mudança para o valor do *voxel*.

Figura 4.7 – Em (a), tem-se a imagem *background* formada pela mediana das imagens da Figura 4.1(a) e (c). Em (b), tem-se a binária de mudança da imagem em (a).



Fonte: Autor

O método encontra tanto as mudanças lesão-lesão quanto as mudanças lesão-fundo em uma única aplicação do mesmo.

4.5 Operador XOR *fuzzy*

Este método utiliza o conceito do operador ou-exclusivo da lógica booleana, com a noção de incerteza, para transformar o valor do *voxel* das imagens de um mesmo objeto. O resultado desse método realça regiões diferentes e deixa menos evidente as semelhanças.

Novamente, esse método foi aplicado sem nenhuma mudança significativa para imagens em 3D.

O método necessita que as duas imagens sejam binárias, ou ainda, que seja associada a cada imagem uma função de semelhança (İLSEVER; ÜNSALAN, 2012). A Equação 51 mostra como calcular o operador XOR *fuzzy*. As Equações 52, 53 e 54 mostram a simplificação da Equação 51.

$$I(x, y, z) = g_a(x, y, z) \oplus g_b(x, y, z) = \bar{g}_a(x, y, z) \bullet g_b(x, y, z) + g_a(x, y, z) \bullet \bar{g}_b(x, y, z) \quad (51)$$

$$\bar{g}(x, y, z) = 1 - g(x, y, z) \quad (52)$$

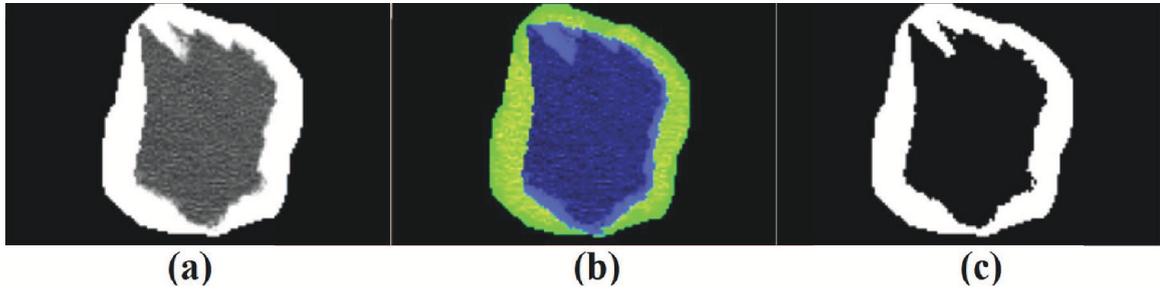
$$g_a(x, y, z) \bullet g_b(x, y, z) = g_a(x, y, z) * g_b(x, y, z) \quad (53)$$

$$g_a(x, y, z) + g_b(x, y, z) = \text{Max}(g_a(x, y, z), g_b(x, y, z)) \quad (54)$$

onde $I(x, y, z)$ é o valor do *voxel* da imagem resultante da operação XOR *fuzzy* e g_i são as funções de semelhança para a imagem I .

O operador XOR *fuzzy* realça as diferenças encontradas *voxel* a *voxel* entre um par de imagens, ou seja, quanto mais diferente o par de *voxels* for, mais próximo de 1 será; e quanto mais semelhante o par de *voxels*, mais próximo de 0 será o seu valor. Dadas as imagens das lesões da Figura 4.1(a) e (c), a imagem resultante da operação XOR *fuzzy* é dada na Figura 4.8(a) e a sua imagem colorida que realça as mudanças é dada na Figura 4.8(b). Para binarizar a imagem resultante em grande mudança e pequena mudança no valor do *voxel* entre as imagens, a técnica de agrupamento *k-means* foi utilizada na Figura 4.8(a), resultando na imagem da Figura 4.8(c).

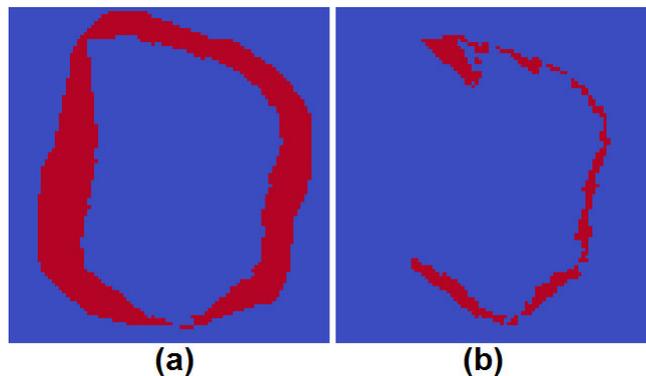
Figura 4.8 – Em (a), tem-se a imagem resultante da aplicação do operador XOR *fuzzy* das imagens da Figura 4.1 - (a) e (c). A imagem em (b) é a imagem colorida de realce de (a). Em (c), tem-se a binária de mudança da imagem em (a).



Fonte: Autor

O passo anterior também foi aplicado duas vezes. As aplicações têm o objetivo de encontrar as mudanças lesão-fundo e lesão-lesão. O resultado da primeira aplicação está na Figura 4.9(a) e da segunda em (b).

Figura 4.9 – Mudanças de densidade detectadas pelo método do operador XOR *fuzzy* da Figura 4.8. Em (a), têm-se as mudanças lesão-fundo e (b), têm-se as mudanças lesão-lesão.



Fonte: Autor

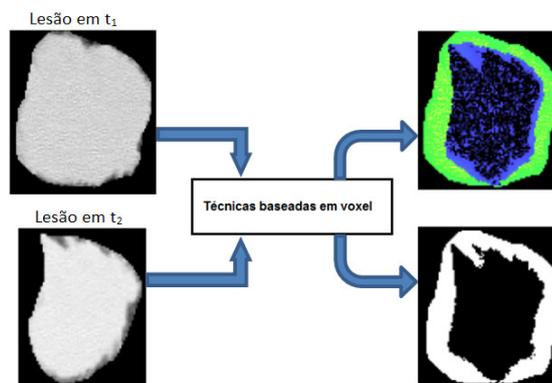
4.6 Identificação das mudanças de densidade

O conjunto de métodos propostos foi utilizado para detecção das mudanças de densidade das lesões pulmonares. O primeiro passo para essa tarefa é realizar o registro das imagens das lesões pulmonares. Isso é necessário, pois as imagens das lesões pulmonares devem possuir as mesmas dimensões em 3D para aplicação dos métodos de detecção de mudanças e são obtidas através do registro rígido (Seção 3.4). Após a realização do registro, cada um dos métodos propostos são aplicados para geração de um conjunto de imagens. As imagens geradas são os realces das mudanças de densidade que ocorreram ao longo

do tempo. Em seguida, quantificam-se as mudanças detectadas e realiza-se a análise de textura sobre estas regiões. O objetivo da análise é verificar se as mudanças encontradas são indicadoras do diagnóstico da lesão. Portanto, ao final da execução de todas as etapas citadas, tem-se uma avaliação de todas as regiões que sofreram mudanças e um diagnóstico final.

A aplicação dos métodos baseados em *voxel* (1- diferença entre imagens, 2- razão de imagens, 3- regressão de imagens, 4- operador XOR *fuzzy* e 5- mediana das imagens) requer apenas que as imagens 3D da lesão estejam registradas. De acordo com a descrição dos métodos na Seção 4, todos eles produzem uma imagem binária final, onde valores em 1 são indicativos de mudança em um *voxel*, enquanto que o valor 0 indica que nenhuma mudança ocorreu naquele *voxel*. A partir da imagem binária, o percentual de mudança é obtido através da Equação 42, que indica qual foi o percentual de *voxels* do volume que sofreram mudanças. Dessa forma, além da visualização das mudanças locais em cada *voxel* da lesão, como pode ser visto na Figura 4.10, pode-se expressar o tamanho dessa mudança.

Figura 4.10 – Esquema geral de aplicação dos métodos baseados em *voxel*.



Fonte: Autor

4.7 Resultados

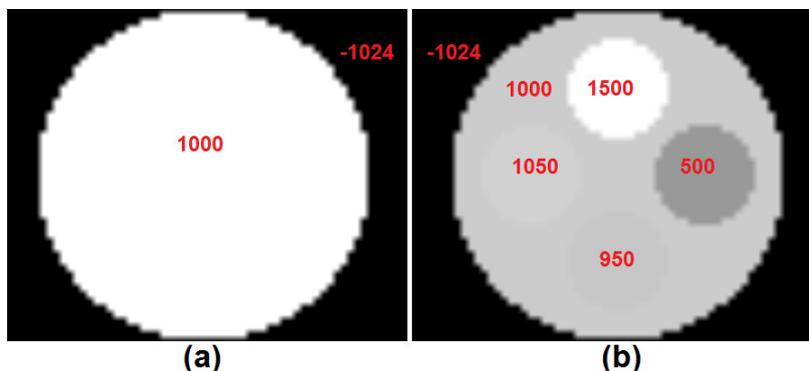
Esta seção contém os resultados numéricos e visuais provenientes da aplicação do método proposto. Dessa forma, pode-se verificar quão efetivo foi o método proposto na verificação das mudanças de densidade, quando comparado à volumetria.

4.7.1 Base de modelos de lesões

Uma base de modelos de lesões foi utilizada para simular mudanças de densidade ao longo do tempo para uma lesão pulmonar. A base é composta de dois modelos. O primeiro modelo simula a lesão em um instante inicial. Este é uma esfera de raio 25, que está em uma imagem cujas dimensões em *voxels* é (51 x 51 x 51) e a dimensão do *voxel* é (1mm x 1mm x 1mm). O valor interno de densidade para o modelo é 1000 HU e o valor de fundo é -1024 HU. Da mesma forma, um segundo modelo foi criado para simular a lesão em um instante posterior.

O segundo modelo é uma esfera com as mesmas características do primeiro modelo. Entretanto, alterações foram realizadas em sua estrutura para simular mudanças de densidade ocorridas em um instante posterior. As alterações foram produzidas com a inclusão de quatro esferas de raio 7,5, todas internas à esfera inicial. Cada uma das esferas possui uma densidade diferente e foram localizadas em pontos distintos, conforme a Figura 4.11. Na figura, pode-se visualizar a posição de cada esfera e a sua correspondente densidade em vermelho. Portanto, as diferentes densidades foram utilizadas para simular pequenas mudanças (aumento e diminuição de 50 HU) e grandes mudanças (aumento e diminuição de 500 HU) do valor de densidade inicial (1000 HU). Logo, a partir dos dois modelos, pode-se verificar a sensibilidade do método proposto em detectar pequenas e grandes mudanças ao longo do tempo.

Figura 4.11 – Modelos que simulam a lesão em dois diferentes instantes. Em (a), tem-se a lesão em instante t_1 e em (b), tem-se a lesão em um instante posterior t_2 .



Fonte: Autor

A partir dos modelos gerados, são obtidos os resultados para as mudanças simuladas. Assim, são apresentados os resultados para cada um dos métodos propostos. O objetivo do uso dos modelos é verificar a sensibilidade às mudanças de densidade. Logo, como

se sabe exatamente quantos *voxels* mudam de densidade para a lesão simulada, pode-se verificar o quanto cada método proposto se distancia do valor real de mudança. O número de *voxels* que se modificam em cada esfera das mudanças é de 1791. Esse valor, em termos de P_m , corresponde a 1,35%. Logo, se os métodos propostos atingirem um P_m próximo ao quádruplo do valor anteriormente citado, estes são declarados sensíveis às mudanças de densidade.

4.7.2 Base de lesões

Para testar os métodos propostos, utilizam-se duas bases de imagens: uma privada e outra pública, com comportamentos distintos, as quais têm seus os detalhes estão na Seção 3.3.1.

Para ambas as bases de lesões, criou-se uma identificação para análise dos exames em tempos diferentes. O propósito da identificação é facilitar a compreensão dos testes realizados. Os pares de lesões analisados foram formados e identificados da seguinte forma: se para um determinado paciente existirem 4 exames da mesma lesão, obtidos em diferentes instantes t_i , então existem 6 combinações de pares para este paciente ($t_1 - t_2$, $t_1 - t_3$, $t_1 - t_4$, $t_2 - t_3$, $t_2 - t_4$ e $t_3 - t_4$). Todas as lesões possuem como padrão de identificação o seguinte conjunto de caracteres: ZZ00XX, onde ZZ identifica a base (SS para pública e NT para privada) e XX é o número de identificação da lesão. Logo, nesse exemplo, em que o paciente possui 4 exames da mesma lesão, cada exame em um instante diferente foi rotulado conforme as Tabelas 4.1 e 4.2. Cada rótulo possui a identificação padrão mais os caracteres "_N", onde N é a ordem de aquisição do exame.

4.7.2.1 Base privada

Esta base contém 16 pacientes. Desses, apenas 5 pacientes foram utilizados nas análises. A razão está na estabilidade ao longo do tempo e no número de exames disponíveis para cada paciente. Logo, foram utilizadas apenas as lesões consideradas estáveis em um período de 3 anos, definidas pelo especialista. Então, das 40 lesões pulmonares disponíveis, apenas 15 lesões foram analisadas, formando 17 pares de lesões. Cada um desses pares está identificado pela coluna denominada **Caso**, na Tabela 4.1.

Tabela 4.1 – Associação de cada caso analisado com o par de instantes da lesão para a base privada. A Coluna **Caso** identifica cada par de instantes da lesão. As Colunas **Lesão em t_1** e **Lesão em t_2** identificam cada um dos elementos do par de instantes da lesão. A Coluna **Registro** indica qual foi o elemento do par considerado como imagem móvel e fixa. A primeira lesão antes de ”/” indica a lesão fixa.

Caso	Par de lesões		Registro
	Lesão em t_1	Lesão em t_2	
1	NT003_1	NT003_2	NT003_2/NT003_1
2	NT003_1	NT003_3	NT003_3/NT003_1
3	NT003_2	NT003_3	NT003_3/NT003_2
4	NT003_3	NT003_4	NT003_3/NT003_4
5	NT003_1	NT003_4	NT003_4/NT003_1
6	NT003_2	NT003_4	NT003_4/NT003_2
7	NT006_1	NT006_2	NT006_1/NT006_2
8	NT006_1	NT006_3	NT006_1/NT006_3
9	NT006_1	NT006_4	NT006_1/NT006_4
10	NT006_2	NT006_3	NT006_3/NT006_2
11	NT006_3	NT006_4	NT006_3/NT006_4
12	NT006_2	NT006_4	NT006_4/NT006_2
13	NT008_1	NT008_2	NT008_2/NT008_1
14	NT010_1	NT010_2	NT010_1/NT010_2
15	NT012_2	NT012_3	NT012_3/NT012_2
16	NT012_3	NT012_4	NT012_3/NT012_4
17	NT012_2	NT012_4	NT012_4/NT012_2

Fonte: Autor

4.7.2.2 Base pública

Esta base é pública e possui marcações para as lesões. A partir da marcação disponível para essa base de imagens, foram segmentadas todas as 24 lesões disponíveis nos 58 exames, formando os 46 pares de lesões. Cada um desses pares está identificado pela coluna denominada **Caso**, na Tabela 4.2.

Tabela 4.2 – Associação de cada caso analisado com o par de instantes da lesão para a base pública. A Coluna **Caso** identifica cada par de instantes da lesão. As Colunas **Lesão em t_1** e **Lesão em t_2** identificam cada um dos elementos do par de instantes da lesão. A Coluna **Registro** indica qual foi o elemento do par considerado como imagem móvel e fixa. A primeira lesão antes de ”/” indica a lesão fixa.

Caso	Par de lesões		Registro
	Lesão em t_1	Lesão em t_2	
1	SS0001.1.mhd	SS0001.2.mhd	SS0001.1/SS0001.2
2	SS0001.1.mhd	SS0001.3.mhd	SS0001.1/SS0001.3
3	SS0001.2.mhd	SS0001.3.mhd	SS0001.2/SS0001.3
4	SS0002.1.mhd	SS0002.2.mhd	SS0002.2/SS0002.1
5	SS0003.1.mhd	SS0003.2.mhd	SS0003.1/SS0003.2
6	SS0009.1.mhd	SS0009.2.mhd	SS0009.2/SS0009.1
7	SS0009.1.mhd	SS0009.3.mhd	SS0009.3/SS0009.1
8	SS0009.2.mhd	SS0009.3.mhd	SS0009.3/SS0009.2
9	SS00010.1.mhd	SS00010.2.mhd	SS00010.2/SS00010.1
10	SS00010.1.mhd	SS00010.3.mhd	SS00010.3/SS00010.1
11	SS00010.2.mhd	SS00010.3.mhd	SS00010.3/SS00010.2
12	SS0011.1.mhd	SS0011.2.mhd	SS0011.2/SS0011.1
13	SS0012.1.mhd	SS0012.2.mhd	SS0012.2/SS0012.1
14	SS0013.1.mhd	SS0013.2.mhd	SS0013.3/SS0013.1
15	SS0013.1.mhd	SS0013.3.mhd	SS0013.3/SS0013.2
16	SS0013.2.mhd	SS0013.3.mhd	SS0013.2/SS0013.1
17	SS0014.1.mhd	SS0014.2.mhd	SS0014.1/SS0014.2
18	SS0015.1.mhd	SS0015.2.mhd	SS0015.2/SS0015.1
19	SS0016.1.mhd	SS0016.2.mhd	SS0016.1/SS0016.2
20	SS0019.1.mhd	SS0019.2.mhd	SS0019.1/SS0019.2
21	SS0020.1.mhd	SS0020.2.mhd	SS0020.1/SS0020.2
22	SS0021.1.mhd	SS0021.2.mhd	SS0021.2/SS0021.1
23	SS0022.1.mhd	SS0022.2.mhd	SS0022.2/SS0022.1
24	SS0023.1.mhd	SS0023.2.mhd	SS0023.1/SS0023.2
25	SS0023.1.mhd	SS0023.3.mhd	SS0023.1/SS0023.3
26	SS0023.1.mhd	SS0023.4.mhd	SS0023.4/SS0023.1
27	SS0023.2.mhd	SS0023.3.mhd	SS0023.2/SS0023.3
28	SS0023.2.mhd	SS0023.4.mhd	SS0023.4/SS0023.2
29	SS0023.3.mhd	SS0023.4.mhd	SS0023.4/SS0023.3
30	SS0024.1.mhd	SS0024.2.mhd	SS0024.1/SS0024.2
31	SS0025.1.mhd	SS0025.2.mhd	SS0025.1/SS0025.2
32	SS0025.1.mhd	SS0025.3.mhd	SS0025.3/SS0025.1
33	SS0025.2.mhd	SS0025.3.mhd	SS0025.3/SS0025.2
34	SS0026.1.mhd	SS0026.2.mhd	SS0026.1/SS0026.2
35	SS0026.1.mhd	SS0026.3.mhd	SS0026.1/SS0026.3
36	SS0026.1.mhd	SS0026.4.mhd	SS0026.4/SS0026.1
37	SS0026.2.mhd	SS0026.3.mhd	SS0026.3/SS0026.2
38	SS0026.2.mhd	SS0026.4.mhd	SS0026.4/SS0026.2
39	SS0026.3.mhd	SS0026.4.mhd	SS0026.4/SS0026.3
40	SS0027.1.mhd	SS0027.2.mhd	SS0027.2/SS0027.1
41	SS0028.1.mhd	SS0028.2.mhd	SS0028.2/SS0028.1
42	SS0029.1.mhd	SS0029.2.mhd	SS0029.2/SS0029.1
43	SS0030.1.mhd	SS0030.2.mhd	SS0030.1/SS0030.2
44	SS0031.1.mhd	SS0031.2.mhd	SS0031.2/SS0031.1
45	SS0031.1.mhd	SS0031.3.mhd	SS0031.1/SS0031.3
46	SS0031.2.mhd	SS0031.3.mhd	SS0031.2/SS0031.3

Fonte: Autor

4.7.3 Registro de imagens

O registro rígido foi realizado utilizando as seguintes configurações: 1) a métrica de semelhança foi a média dos quadrados das diferenças; 2) a função de interpolação utilizada foi o vizinho mais próximo; 3) a função de transformação foi o versor rígido 3D; e 4) a função de otimização utiliza o gradiente descendente de passo regular, específico para o espaço de parâmetros da função de transformação.

O registro rígido foi obtido configurando a função de otimização para executar 800 iterações, com o comprimento do passo mínimo de 0,0001 e máximo de 0,2. Tais valores de configuração são os mais comuns na tarefa de otimização.

4.7.4 Estudo de casos com modelos de lesões

Aplica-se nesta seção cada um dos métodos propostos aos modelos de lesões e apresentam-se os resultados encontrados. A partir dos resultados, verifica-se a aproximação para o percentual de mudança real das mudanças.

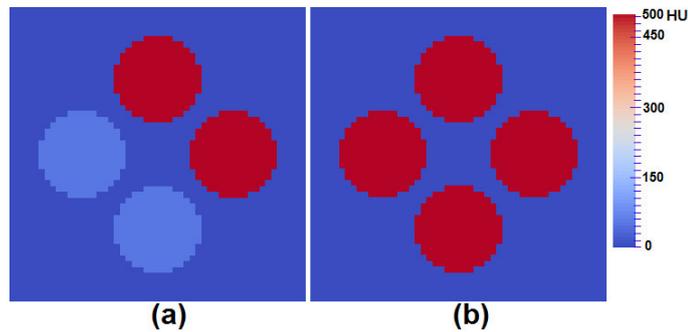
4.7.4.1 Diferença entre imagens

O método teve a melhor sensibilidade para detecção das mudanças de densidade. Ele detectou tanto as pequenas mudanças quanto as grandes mudanças de densidade, como pode ser observado da Figura 4.12(a), correspondente a diferença dos modelos. A figura apresenta as quatro esferas de mudança em destaque do fundo da imagem. O valor máximo da escala presente na figura e realçado em vermelho pela palheta representa as mudanças de maior intensidade. Da mesma forma, a imagem da Figura 4.12(b) é a imagem binária da diferença e confirma que todas as alterações de densidade foram detectadas. Como consequência, o valor de P_m encontrado é igual a 5,4%, valor previamente calculado para as quatro esferas de mudança.

4.7.4.2 Razão de imagens

O método proposto não foi promissor na detecção de todas as mudanças presentes no modelo. Embora o método tenha gerado a imagem da razão com as mudanças das

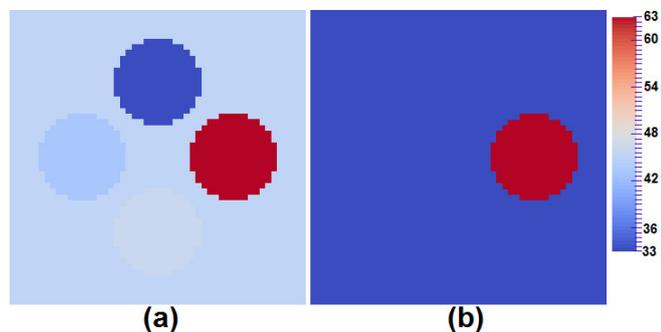
Figura 4.12 – Resultado do método da diferença para os modelos da Figura 4.11. Em (a), tem-se a imagem da diferença e em (b), a imagem binária de mudança.



Fonte: Autor

quatro esferas destacadas do fundo, a imagem binária com as mudanças detectadas apresentou apenas uma única esfera de mudança. As esferas destacadas na imagem da razão possuem valores distantes do centro da escala (45°). Este resultado pode ser visualizado na Figura 4.13. Nela, verifica-se que a única esfera de mudança detectada é a de redução de densidade na imagem binária em vermelho, isto é, a que apresenta densidade 500 HU. Como uma única esfera foi detectada, o P_m encontrado corresponde a 1,35%. O resultado é devido a definição dos limiares para determinar quais são os *voxels* de mudança. Dessa forma, a esfera de 500 HU foi a única que ficou dentro do limiar de mudança. Logo, conclui-se que o método realça todas as mudanças, mas não foi eficaz em detectar tanto as pequenas mudanças quanto as grandes mudanças, visto que a esfera de densidade 1500 HU não foi detectada.

Figura 4.13 – Resultado do método da razão para os modelos da Figura 4.11. Em (a), tem-se a imagem da razão e em (b), a imagem binária de mudança.

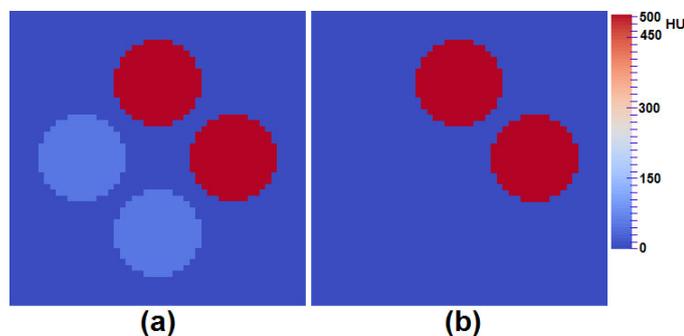


Fonte: Autor

4.7.4.3 Regressão de imagem

O método detectou apenas as grandes mudanças. Contudo, o método produz a imagem de regressão com todas as mudanças em destaque. Este resultado pode ser visto na Figura 4.14(a). Da mesma forma, a imagem na Figura 4.14(b) possui as duas esferas de mudança detectadas em vermelho. O método detectou o aumento e a diminuição de 500 HU do valor inicial (1000 HU). Logo, o P_m encontrado é igual a 2,7% e o método apenas detectou mudanças de grande magnitude de densidade.

Figura 4.14 – Resultado do método de regressão para os modelos da Figura 4.11. Em (a), tem-se a imagem de regressão e em (b), a imagem binária de mudança.



Fonte: Autor

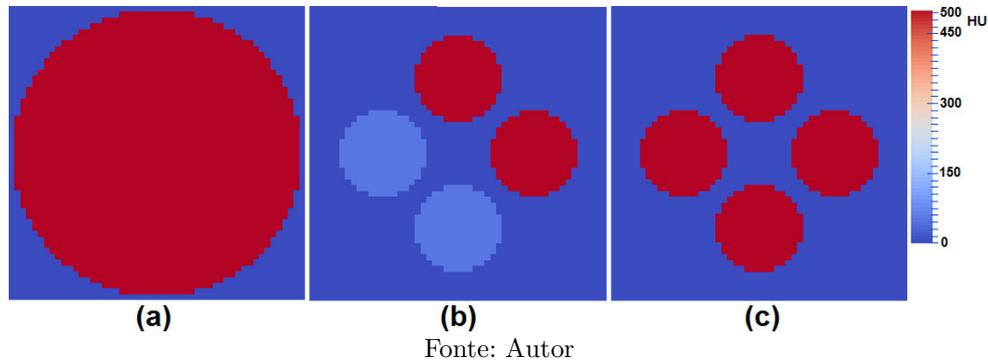
4.7.4.4 Imagem mediana

O método teve a mesma eficiência do método da diferença. Ele detectou todas as esferas de mudança. As quatro esferas que foram realçadas como mudança pelo método estão na Figura 4.15(b). A figura apresenta a imagem da diferença para a imagem mediana da Figura 4.15(a). As quatro esferas foram detectadas na última etapa do método, conforme a Figura 4.15(c). O sucesso da detecção de todas as mudanças de densidade está na similaridade com o método da diferença entre imagem. Portanto, o valor de P_m (5,4%) encontrado pelo método é igual ao valor real.

4.7.4.5 Operador XOR *fuzzy*

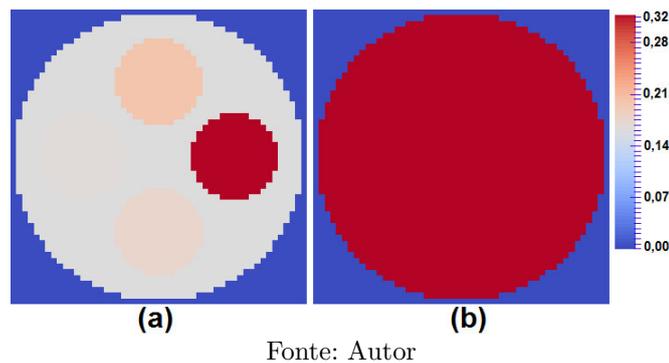
O método apresentou o pior desempenho quando comparado aos métodos baseados em *voxel* anteriores. Ele realçou todas as quatro esferas de mudança. Entretanto, a última etapa de realce das mudanças falha em detectar cada uma das esferas de mudança. Essa

Figura 4.15 – Resultado do método da mediana para os modelos da Figura 4.11. Em (a), tem-se a imagem da mediana, em (b), a imagem da diferença e em (c), a imagem binária de mudança.



falha está relacionada com a proximidade dos valores resultantes para as esferas de mudança que estão próximos do valor de fundo. Assim, o *k-means* não consegue gerar um grupo separados para as mudanças. Este resultado pode ser visto na Figura 4.16. Nela encontra-se a imagem *fuzzy* em (a) e em (b), a imagem binária de mudança. Logo, o P_m igual a 49,1% se distancia muito do valor real (5,4%). Portanto, como pode ser visto na Figura e pelo resultado de P_m , o método teve o pior desempenho na última etapa para as mudanças dos modelos utilizados.

Figura 4.16 – Resultado do método XOR *fuzzy* para os modelos da Figura 4.11. Em (a), tem-se a imagem XOR *fuzzy* e em (b), a imagem binária de mudança.



A seguir, serão apresentados os resultados visuais de cada método para ambas as bases de lesões pulmonares, com relação aos dois menores valores de *VVR*.

4.7.5 Estudo de casos com a base pública de lesões

Para a base pública, a partir das imagens registradas e dos pares de lesões formados, cada um dos métodos para análise de mudanças de densidade é aplicado e as medidas

que quantificam essas mudanças são obtidas. Sempre que um resultado da aplicação dos métodos é obtido, uma imagem binária de realce das mudanças identifica quais os *voxels* apresentaram modificações mais relevantes. A Tabela 4.3 mostra cada uma dessas medidas obtidas para os métodos propostos para cada par de lesões. Baseado nesta tabela, constata-se que:

- As medidas de percentual de mudança de densidade (P_m) para o método da diferença entre imagens não foram muito diferentes daquelas obtidas pelo método da mediana e do operador XOR *fuzzy*;
- Os Casos 1, 2, 42 e 45 foram detectados como os casos que possuíram um dos 4 maiores valores de P_m por cada um dos métodos. Da mesma forma, o Caso 42 foi detectado pelos métodos da diferença entre imagens e mediana como o caso de maior variação de densidade ao longo do tempo. Esse resultado é concordante com o obtido pelo Reeves et al. (2009), que apresenta esse caso como o de maior variação de volume relativo positivo (VVR);
- Todos os métodos indicaram o Caso 2 como um dos casos de maior mudança, o que também é concordante com Reeves et al. (2009), pois apresentou o caso de maior VVR negativo;
- A partir da análise estatística da correlação de Pearson entre as medidas obtidas de P_m com a medida de referência, o VVR, constata-se que existe uma pequena correlação entre as medidas, para um nível de significância de 5%. A correlação entre as duas medidas foi realizada com o valor de VVR em módulo, em razão do P_m obtido apenas indica o quanto o valor da densidade mudou para a lesão em dois instantes diferentes.
 - Primeiro se analisa a correlação entre o P_m obtido do método de diferença entre imagens com o VVR. A correlação de Pearson obtida é igual a 0,4832 com um valor-p de 0,0007. Com este resultado, rejeita-se a hipótese nula de não existência de correlação entre as variáveis analisadas para o nível de significância de 5%;
 - A próxima medida de P_m analisada com relação ao VVR é a obtida com o método da razão. A correlação de Pearson obtida neste caso foi de 0,3561 e o valor-p foi de 0,0151, ou seja, inferior ao nível de significância de 5%. Portanto,

rejeita-se a hipótese nula de não existência de correlação entre as variáveis analisadas;

- A imagem mediana, quando correlacionada com o VVR, teve coeficiente de correlação igual a 0,4389 e valor-p de 0,0023, resultando na correlação das variáveis. O coeficiente de correlação para o operador XOR *fuzzy* foi de 0,2934 e o valor-p obtido foi de 0,0480. O valor-p obtido é muito próximo do nível de significância, o que traduz em fraca correlação;
 - A regressão foi a única medida cujo coeficiente de correlação teve valor negativo, com o seu valor igual a -0,1357. Esse valor inconsistente é justificado pelo valor-p de 0,3687, que é superior ao nível de significância de 5%, indicando que a hipótese nula de não correlação entre as medidas não pode ser rejeitada.
- Percebe-se que o método de diferença entre imagens é o método que mais teve correlação com o valor de VVR, mostrando-se mais sensível às variações de volume da lesão. Assim, conclui-se que os resultados com baixos valores de coeficiente de correlação de Pearson mostram que as mudanças nos valores de densidade da lesão não são bons indicativos para uma mudança no volume. Mesmo a lesão não sofrendo grande variações de volume, pode ocorrer grande variação dos valores de densidade, como no Caso 10. Nesse caso, o VVR é igual a -1,61, mas o valor de P_m é de 16,36 para o método de diferença entre imagens.

As Tabelas 4.3 e 4.4 contêm os valores de P_m encontrados pela aplicação de cada um dos métodos para a base pública e privada, respectivamente.

Tabela 4.3 – Conjunto de valores de percentuais de mudança para os métodos baseados em *voxel*, aplicados a cada caso da base pública de lesões.

Caso	P_m (%)					
	Diferença	Razão	Mediana	XOR	Regressão	VVR
1	21,92	33,75	22,56	21,86	19,65	-48,30
2	29,45	36,57	30,23	29,37	33,30	-70,10
3	18,96	31,04	20,15	18,96	16,95	-42,17
4	8,24	11,23	7,87	7,88	5,57	133,60
5	9,01	12,97	11,25	8,97	8,76	-29,09
6	21,56	24,94	19,71	21,22	9,99	257,66
7	20,02	21,28	16,83	19,75	7,14	281,26
8	17,45	19,86	16,06	16,65	21,28	6,60
9	17,62	21,98	16,18	17,74	24,03	22,48
10	16,36	18,19	15,34	17,48	19,66	-1,61
11	19,89	23,22	19,40	20,14	23,86	-19,67
12	23,47	27,85	23,01	23,21	12,34	121,24
13	12,85	19,76	12,58	17,21	12,03	75,19
14	18,87	20,74	18,52	18,75	7,95	185,52
15	17,92	20,35	17,53	14,44	8,89	162,29
16	15,77	18,86	14,66	15,24	17,68	8,86
17	4,12	4,63	3,99	3,24	4,93	-17,94
18	18,60	21,01	18,53	18,53	6,51	275,08
19	22,27	27,74	22,19	22,30	22,18	-34,20
20	22,32	19,31	23,96	32,90	26,28	-13,72
21	9,97	11,81	10,06	8,19	10,36	-45,58
22	21,19	25,24	20,14	20,89	27,73	55,84
23	17,46	20,83	16,51	17,69	22,92	32,52
24	13,10	12,90	12,40	10,19	14,07	-4,77
25	14,04	14,31	13,43	13,46	16,64	11,68
26	12,78	13,51	12,14	13,69	14,92	-7,74
27	15,33	15,47	14,27	13,20	17,24	17,27
28	11,58	11,86	11,21	9,43	13,06	-3,12
29	11,94	12,76	11,35	9,51	13,42	-17,39
30	18,46	23,89	17,13	19,44	19,04	17,29
31	9,52	17,56	8,87	8,70	10,53	-5,29
32	5,84	10,16	5,74	4,66	5,62	4,55
33	6,81	11,02	6,66	5,58	6,78	10,38
34	10,43	12,26	10,42	9,07	13,78	-11,22
35	15,93	17,98	15,16	12,68	12,48	30,35
36	17,81	19,85	17,35	17,50	9,82	102,14
37	12,35	14,19	11,95	12,21	10,76	46,83
38	17,80	19,91	17,16	17,60	10,07	127,69
39	12,16	15,28	11,61	10,51	12,50	55,08
40	21,80	34,50	20,69	21,24	21,12	26,74
41	17,82	20,05	17,09	17,48	9,30	95,76
42	31,82	32,62	30,36	31,88	5,47	521,34
43	8,94	11,83	8,71	8,69	9,02	-17,12
44	12,54	13,69	11,51	13,53	18,13	26,81
45	27,47	29,08	27,47	27,50	33,52	-46,43
46	20,28	21,32	20,79	20,55	21,98	-57,75

Fonte: Autor

A seguir, serão apresentados dois casos com os menores valores de VVR para cada método. Tais casos, devido à pequena variação volumétrica, representam as dificuldades de observação das mudanças da lesão pulmonar pelo especialista.

4.7.5.1 Diferença entre imagens

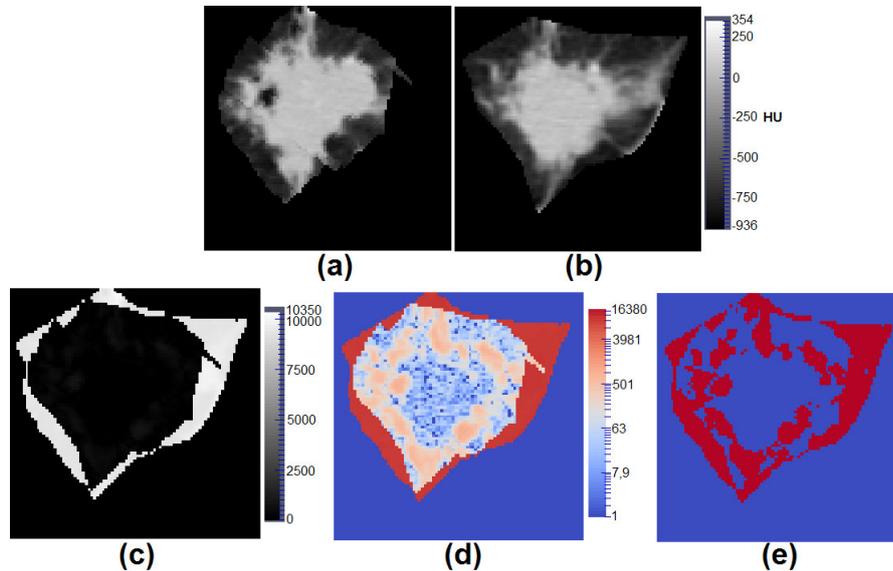
A diferença entre imagens teve um bom desempenho na detecção de mudanças internas e externas à lesão pulmonar. As mudanças internas são aquelas onde regiões volumétricas das lesões se interceptam e possuem pouca diferença devido ao processo de registro. Neste caso, como se pode perceber pela Figura 4.17(a), o método consegue identificar a pequena variação da densidade do *voxel*, apesar das poucas regiões.

Quando os pares de lesões são visualmente diferentes, ou seja, quando o volume de uma das lesões é maior que o outro, uma região da lesão de maior volume não intercepta a lesão de menor volume. Esta é a mudança externa, isto é, a lesão pulmonar pode ter aumentado ou diminuído ao longo do tempo.

O Caso 10 apresentou o menor valor de VVR , isto é, -1,61%, indicando que esta lesão sofreu pequena redução em seu volume. Entretanto, o método proposto detectou que essa lesão sofreu variações nos seus valores de *voxel* por todo o seu volume. Na Figura 4.17, podem-se visualizar estas mudanças em seus *voxels*. Logo, embora esse caso apresente pequena variação volumétrica, o valor de P_m encontrado foi 16,36%.

As imagens da lesão do Caso 10 estão na Figura 4.17(a) e (b), para os dois instantes analisados. Nesse caso, pode-se ver na Figura 4.17(c) que o método foi capaz de detectar tanto mudanças internas quanto externas. Na Figura 4.17(d), essas mudanças foram transformadas em cores e na Figura 4.17(e), têm-se as regiões com as principais mudanças.

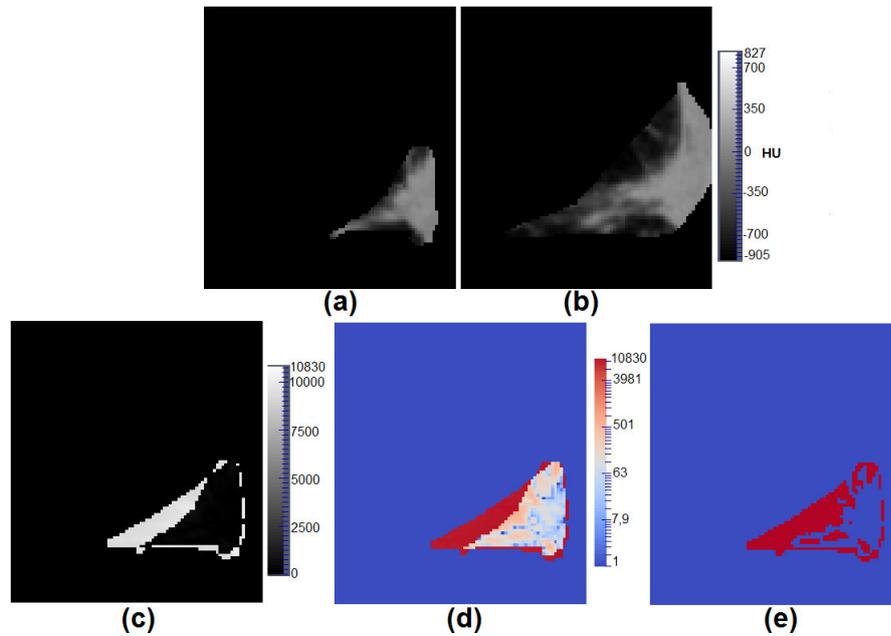
Figura 4.17 – Caso 10 apresentou o menor valor de VVR . Em (a), tem-se a lesão pulmonar em um instante e em (b), tem-se a lesão pulmonar em um instante anterior. A imagem em (c) é a diferença das imagens em (a) e (b). Em (d), tem-se a imagem colorida da imagem diferença em (c). Já a imagem em (e) é a imagem resultante da binarização com k -means da imagem em (c), que enfatiza as mudanças ocorridas para a lesão pulmonar ao longo do tempo.



Fonte: Autor

O segundo menor valor de VVR , $-3,12$, pertence ao Caso 28 e novamente ocorreu uma pequena redução do volume. Porém, mudanças nos valores de *voxel* foram detectadas ao longo do volume da lesão e podem ser vistas nas imagens da Figura 4.18(e). O valor de P_m encontrado para esse caso foi igual a $11,58\%$. Para se entender o motivo dessas mudanças, pode-se visualizar a imagem diferença entre as lesões, dada na Figura 4.18(c), e melhor vista na imagem da Figura 4.18(d), na qual foi aplicada um conjunto de cores para realce das diferenças.

Figura 4.18 – Caso 28 apresentou o segundo menor valor de VVR . Em (a), tem-se a lesão pulmonar em um instante e em (b), tem-se a lesão pulmonar em um instante anterior. A imagem em (c) é a diferença das imagens em (a) e (b). Em (d), tem-se a imagem colorida da imagem diferença em (c). Já a imagem em (e) é a imagem resultante da binarização com k -means da imagem em (c), que enfatiza as mudanças ocorridas para a lesão pulmonar ao longo do tempo.

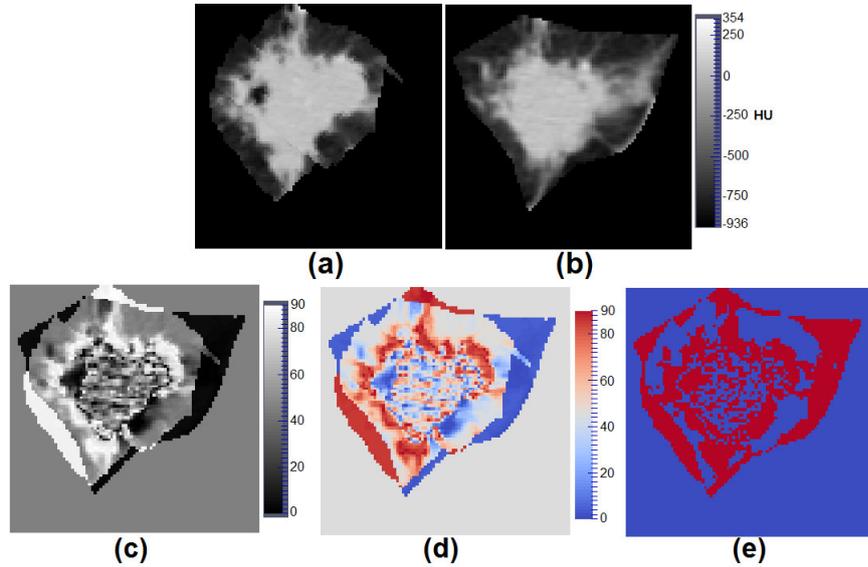


Fonte: Autor

4.7.5.2 Razão de imagens

Para esse método, o Caso 10 teve percentual de mudança igual a 18,19%. Essas modificações podem ser vistas na Figura 4.19(c), (d) e (e), que mostram, respectivamente, a imagem da razão, a imagem razão com realce colorido e a imagem das mudanças sofridas pela lesão pulmonar para as imagens da Figura 4.19(a) e (b). É possível observar que esse método encontra com maior facilidade mudanças internas pontuais de *voxel*, diferentemente do método de diferença entre imagens, que não detectou essas mudanças. Por isso, os valores de P_m para o método da razão são ligeiramente diferentes para cada caso.

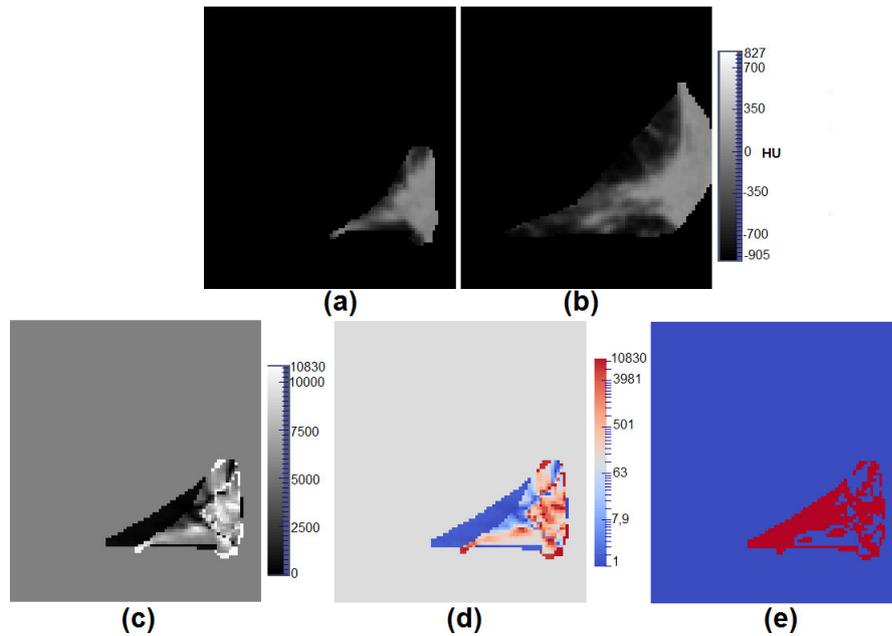
Figura 4.19 – Caso 10 apresentou o menor valor de VVR . Em (a), tem-se a lesão pulmonar em um instante e em (b), tem-se a lesão pulmonar em um instante anterior. A imagem em (c) é a razão das imagens em (a) e (b). Em (d), tem-se a imagem colorida da imagem razão em (c). Já a imagem em (e) é a imagem resultante da binarização com k -means da imagem em (c), que enfatiza as mudanças ocorridas para a lesão pulmonar ao longo do tempo.



Fonte: Autor

O segundo caso de menor VVR é exibido na Figura 4.20(a) e (b), que corresponde ao Caso 28. A lesão pulmonar reduziu seu volume ao longo do tempo, provocando tanto mudanças internas quanto externas. Um exemplo dessas mudanças está nas Figuras 4.20(c), (d) e (e), que representam, respectivamente, a imagem da razão, a imagem razão com realce colorido e a imagem das mudanças sofridas pela lesão pulmonar. O percentual de mudança de 11,86% foi encontrado, que é um valor maior que o obtido pelo método da diferença entre imagens, para o mesmo caso.

Figura 4.20 – Caso 28 apresentou o segundo menor valor de VVR . Em (a), tem-se a lesão pulmonar em um instante e em (b), tem-se a lesão pulmonar em um instante anterior. A imagem em (c) é a razão das imagens em (a) e (b). Em (d), tem-se a imagem colorida da imagem razão em (c). Já a imagem em (e) é a imagem resultante da binarização com k -means da imagem em (c), que enfatiza as mudanças ocorridas para a lesão pulmonar ao longo do tempo.



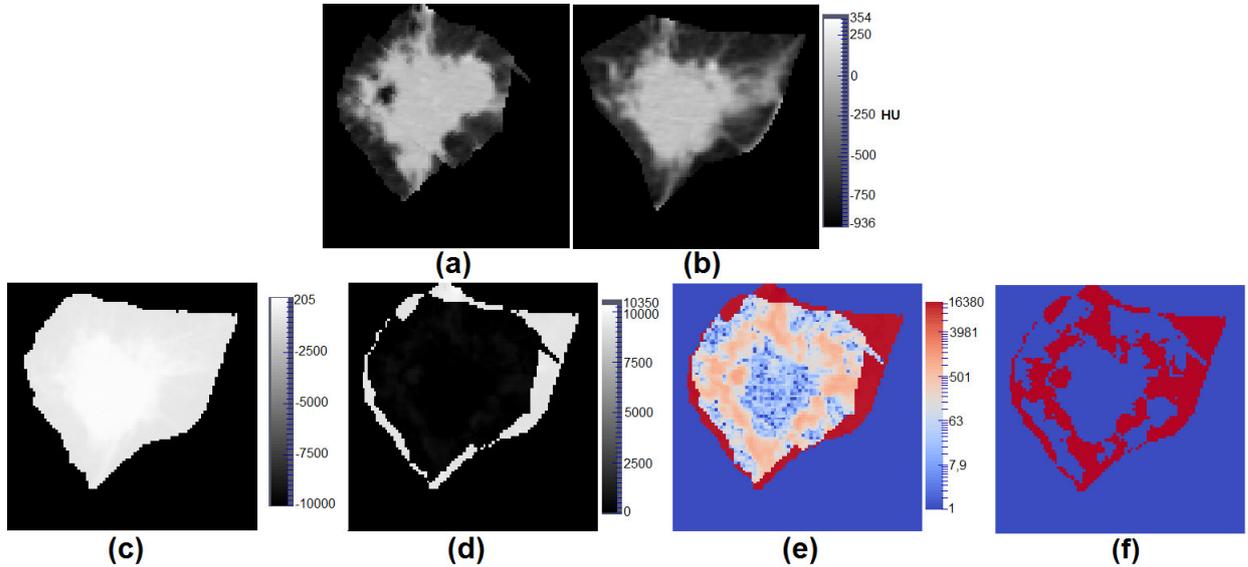
Fonte: Autor

4.7.5.3 Imagem Mediana

Os resultados para o método da imagem mediana foram semelhantes aos do método da diferença entre imagens. Uma vez que, os casos com maior percentual de mudança são os mesmos encontrados para o método de diferença entre imagens.

A Figura 4.21(c) mostra a mediana entre as imagens da Figura 4.21(a) e (b) que corresponde ao Caso 10, no qual o valor de P_m encontrado foi 15,34. A imagem mediana é então utilizada como uma das imagens no método de diferença entre imagens, com a imagem da Figura 4.21(b) como segunda imagem, resultando na imagem diferença da Figura 4.21(d), na qual a imagem colorida de realce pode ser vista na Figura 4.21(e). A imagem da Figura 4.21(f) realça onde ocorreram as mudanças para essa lesão pulmonar ao longo do tempo.

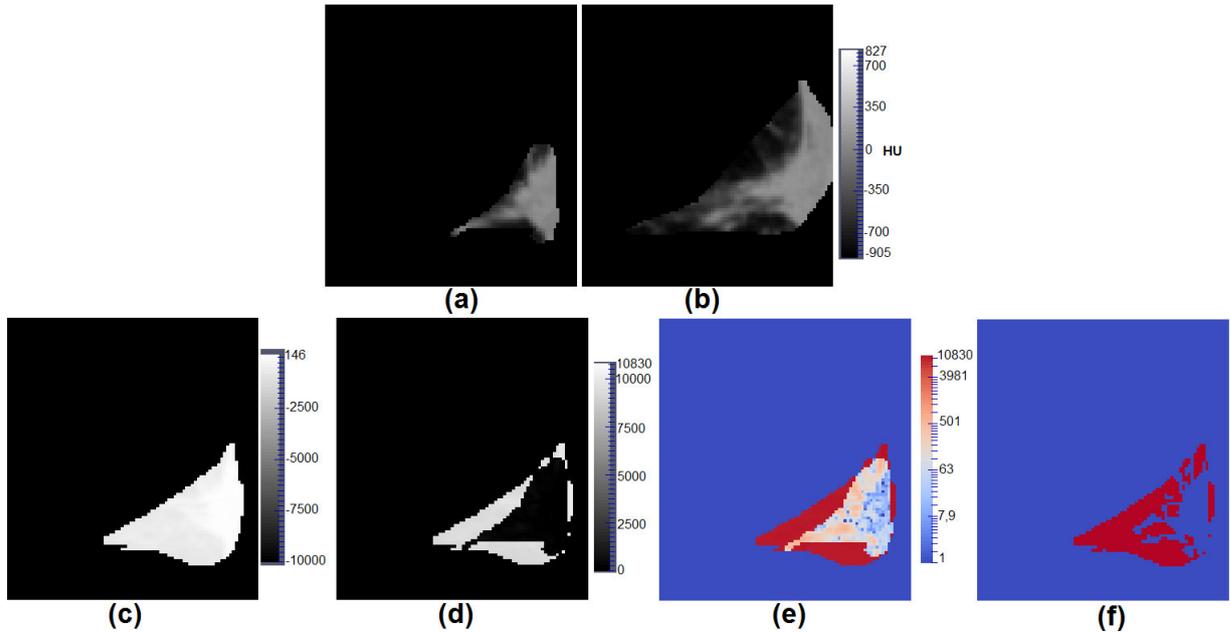
Figura 4.21 – Caso 10 apresentou menor valor de VVR . Em (a), tem-se a lesão pulmonar em um instante posterior e em (b), tem-se a lesão pulmonar em um instante anterior. A imagem em (c) é a imagem mediana de (a) e (b). A diferença entre a imagem mediana (c) com a imagem em (b) resulta na imagem em (d), na qual as regiões que sofreram mudanças estão em (f). Em (e), tem-se a imagem colorida da imagem diferença em (d).



Fonte: Autor

O segundo caso de menor VVR pode ser visto na Figura 4.22 e corresponde ao Caso 28, que produziu um valor de P_m igual a 11,21. Na Figura 4.22(a), tem-se a lesão pulmonar em um instante e a mesma lesão pulmonar em um outro instante, registrado com a lesão pulmonar anterior, conforme a Figura 4.22(b). Enquanto que na Figura 4.22(c), tem-se a imagem mediana e a imagem diferença resultante, na Figura 4.22(d), na qual a imagem colorida de realce pode ser vista na Figura 4.22(e). A Figura 4.22(f) mostra as regiões onde ocorreram as mudanças.

Figura 4.22 – Caso 28 apresentou o segundo menor valor de VVR . Em (a), tem-se a lesão pulmonar em um instante posterior e em (b), tem-se a lesão pulmonar em um instante anterior. A imagem em (c) é a imagem mediana de (a) e (b). A diferença entre a imagem mediana (c) com a imagem em (b) resulta na imagem em (d), na qual as regiões que sofreram mudanças estão em (f). Em (e), tem-se a imagem colorida da imagem diferença em (d).

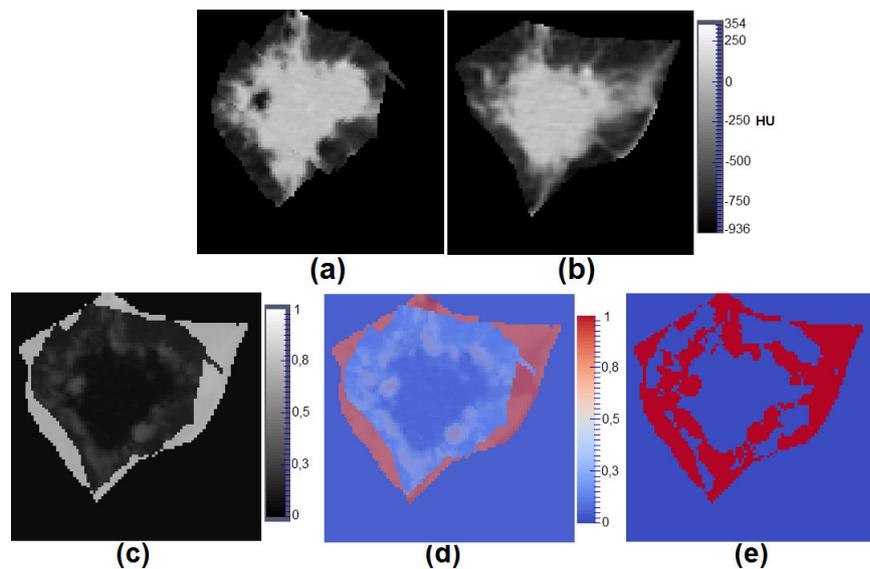


Fonte: Autor

4.7.5.4 Operador XOR *fuzzy*

O método consegue identificar mais as mudanças de densidade que ocorreram próximo aos limites da lesão do Caso 10. A partir da Figura 4.23, consegue-se visualizar esse resultado. Nela, visualizam-se as duas imagens da lesão pulmonar em tempos diferentes, Figura 4.23(a) e (b); a imagem resultante do operador XOR *fuzzy*, que está na Figura 4.23(c); e a imagem realçada com a aplicação de cores, na Figura 4.23(d). Da mesma forma, a imagem da Figura 4.23(e) realça onde ocorreram as mudanças para a lesão pulmonar ao longo do tempo. Apesar de ser o primeiro caso de menor valor de VVR , o valor de P_m encontrado é igual a 17,48%. Logo, seu valor relativamente pequeno, quando comparado aos demais casos, salienta que o método detectou poucas regiões internas à lesão.

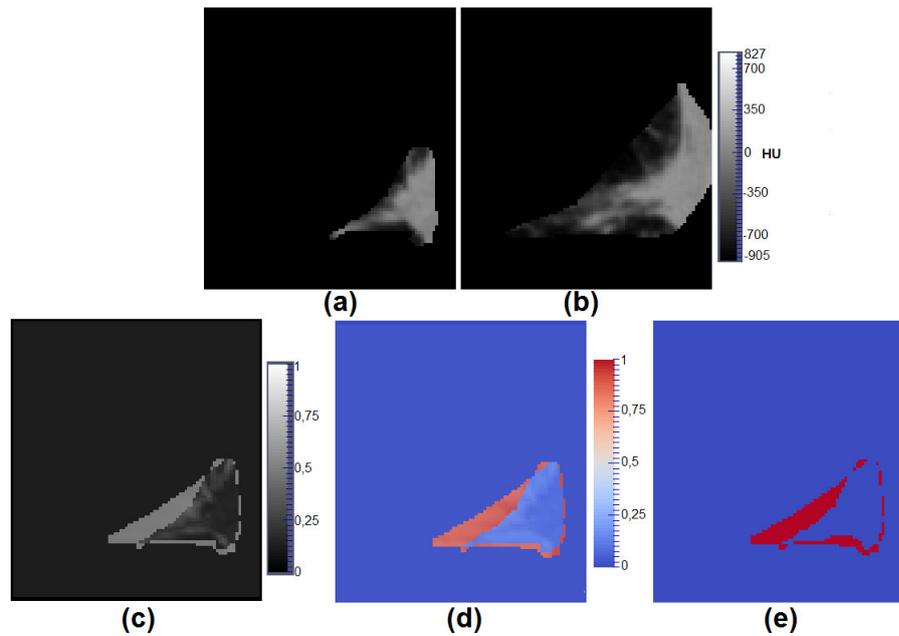
Figura 4.23 – Caso 10 apresentou menor valor de P_m . Em (a), tem-se a lesão pulmonar em um instante e em (b), tem-se a lesão pulmonar em um instante anterior. A imagem em (c) é a imagem resultante da operação XOR *fuzzy* das imagens em (a) e (b). A imagem em (d) é a imagem colorida da imagem em (c). Já a imagem em (e) é a imagem resultante da binarização com *k-means* da imagem em (c), que enfatiza as mudanças ocorridas para a lesão pulmonar ao longo do tempo.



Fonte: Autor

Para o par de lesões da Figura 4.24(a) e (b) pertencente ao Caso 28, o percentual de mudança encontrado foi de 13,69%. O baixo valor de P_m indica que o método novamente detectou mais as regiões nos limites da lesão. A imagem resultante da aplicação da operação XOR *fuzzy* pode ser vista na Figura 4.24(c) e realçada com a aplicação de cores na Figura 4.24(d). Verifica-se, na Figura 4.24(e), o realce das mudanças sofridas pela lesão pulmonar ao longo do tempo. Na figura, percebe-se que o método não consegue identificar mudanças internas, logo, as mudanças externas são capturadas em maior extensão.

Figura 4.24 – Caso 28 apresentou o segundo menor valor de P_m . Em (a), tem-se a lesão pulmonar em um instante e em (b), tem-se a lesão pulmonar em um instante anterior. A imagem em (c) é a imagem resultante da operação XOR *fuzzy* das imagens em (a) e (b). A imagem em (d) é a imagem colorida da imagem em (c). Já a imagem em (e) é a imagem resultante da binarização com *k-means* da imagem em (c), que enfatiza as mudanças ocorridas para a lesão pulmonar ao longo do tempo.

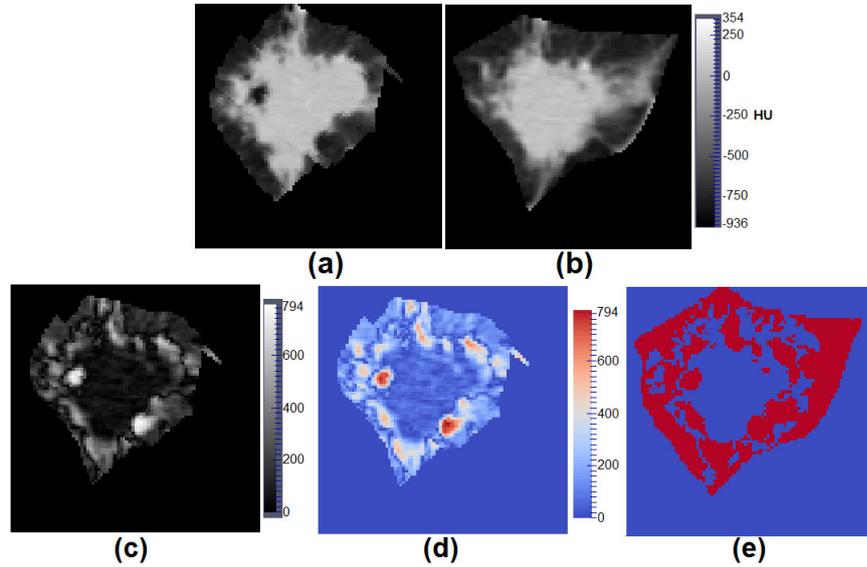


Fonte: Autor

4.7.5.5 Regressão de imagens

O Caso 10 de menor valor de VVR teve um percentual de mudança de 19,66. Percebe-se, pela Figura 4.25(c) e (d), que o método consegue detectar as mudanças internas à lesão pulmonar. A partir das imagens da Figura 4.25(a) e (b), uma imagem resultante da aplicação da Equação 49 é gerada na Figura 4.25(c), que é chamada de imagem regressão. A imagem regressão realçada com a aplicação de cores pode ser vista na Figura 4.25(d). Como resultado final, tem-se a imagem das modificações da lesão na Figura 4.25(e); e a cor branca, na figura, enfatiza essas mudanças.

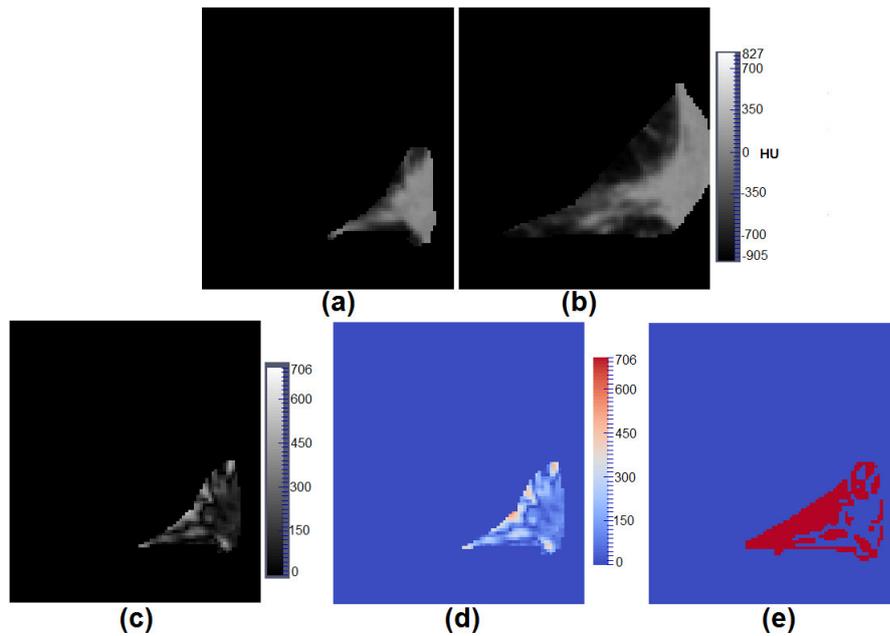
Figura 4.25 – Caso 10 de menor valor de VVR . Em (a), tem-se a lesão pulmonar em um instante e em (b), tem-se a lesão pulmonar em um instante anterior. A imagem em (c) é a imagem criada da regressão das imagens em (a) e (b). A imagem em (d) é a imagem colorida da imagem em (c). Já a imagem em (e) é a imagem resultante da aplicação do k -means da imagem em (c), que enfatiza as mudanças ocorridas para a lesão pulmonar ao longo do tempo.



Fonte: Autor

O mesmo ocorreu para o Caso 28, no qual as mudanças internas foram verificadas na imagem da Figura 4.26(e) resultante da aplicação do k -means da imagem regressão da Figura 4.26(c), que foi gerada a partir das imagens da Figura 4.26(a) e (b). Um realce colorido da imagem regressão da Figura 4.26(c) pode ser visto na Figura 4.26(d). As regiões em branco da Figura 4.26(e) realçam as mudanças que ocorreram na lesão. O valor de P_m encontrado foi igual a 13,06.

Figura 4.26 – Caso 28 com o segundo menor valor de VVR . Em (a), tem-se a lesão pulmonar em um instante e em (b), tem-se a lesão pulmonar em um instante anterior. A imagem em (c) é a imagem criada da regressão das imagens em (a) e (b). A imagem em (d) é a imagem colorida da imagem em (c). Já a imagem em (e) é a imagem resultante da aplicação do k -means da imagem em (c), que enfatiza as mudanças ocorridas para a lesão pulmonar ao longo do tempo.



Fonte: Autor

4.7.6 Estudo de casos com a base privada de lesões

A base privada, por possuir apenas nódulos pulmonares em acompanhamento, teve valores de VVR menores em comparação com a base pública de lesões, exceto o Caso 14, que teve uma redução de 60% do volume, possivelmente, causada pela diminuição do processo inflamatório.

Os valores de P_m para essa base de lesões são encontrados na Tabela 4.4.

Tabela 4.4 – Conjunto de valores de percentuais de mudança para os métodos baseados em *voxel*, aplicados a cada caso da base privada de lesões.

Caso	P_m (%)					
	Diferença	Razão	Mediana	XOR	Regressão	VVR
1	12,84	17,71	12,84	17,19	13,12	-2,86
2	15,65	21,94	14,63	17,22	19,82	5,71
3	11,73	19,43	11,30	14,86	11,50	8,82
4	9,84	15,95	10,59	13,78	9,53	1,35
5	16,05	22,89	15,57	18,20	15,19	7,14
6	13,70	21,51	13,42	16,25	13,38	10,29
7	19,18	30,57	20,71	26,63	19,53	7,41
8	16,08	26,87	15,09	24,36	16,22	20,37
9	19,23	28,90	19,63	25,79	19,87	9,26
10	20,82	31,88	20,04	23,98	20,16	12,07
11	21,64	33,24	20,00	25,00	20,59	-9,23
12	17,26	27,36	16,19	24,10	18,60	1,72
13	24,25	28,76	21,71	27,87	28,76	12,50
14	33,89	37,83	33,98	24,49	25,18	-60,00
15	12,25	19,04	12,76	10,94	13,29	-3,51
16	16,80	21,11	17,06	16,32	14,20	-23,64
17	18,82	25,14	19,18	19,14	15,31	-26,32

Fonte: Autor

4.7.6.1 Diferença entre imagens

Os Casos 4 e 12 apresentaram os menores valores de *VVR*, 1,35 e 1,72, respectivamente. Os percentuais de mudança para esses casos foram 9,84 e 17,26, respectivamente. Isso indica que, mesmo que esses nódulos pulmonares sejam considerados estáveis, o método da diferença foi capaz de detectar as mudanças nas densidades da lesão ao longo do tempo.

Na Figura 4.27, visualiza-se o Caso 4, de menor *VVR*. As imagens da lesão nos dois instantes analisados podem ser vistas na Figura 4.27(a) e (b). Conforme se visualiza na imagem da diferença, Figura 4.27(c), as mudanças detectadas foram maiores nos limites da lesão e as pequenas diferenças são mais uniformes no interior da mesma. Uma melhor visualização das mudanças da imagem diferença pode ser vista na Figura 4.27(d), na qual uma palheta de cores foi utilizada. A imagem binária gerada confirma que as mudanças se localizam mais nos limites da lesão e podem ser visualizadas na Figura 4.27(e). Na figura, as regiões em vermelho são as regiões onde ocorrem as mudanças.

O mesmo ocorreu para o Caso 12, em que as mudanças nos limites da lesão podem ser visualizadas na imagem da diferença na Figura 4.27(c). A imagem colorida através da

aplicação da mesma palheta pode ser vista na Figura 4.27(d). A imagem binária mostra que as mudanças ocorreram nos limites da lesão e podem ser vistas na Figura 4.27(e).

Figura 4.27 – Caso 4 apresentou o menor valor de VVR . Em (a), tem-se a lesão pulmonar em um instante e em (b), tem-se a lesão pulmonar em um instante anterior. A imagem em (c) é a diferença das imagens em (a) e (b). Em (d), tem-se a imagem colorida da imagem diferença em (c). Já a imagem em (e) é a imagem resultante da binarização com k -means da imagem em (c), que enfatiza as mudanças ocorridas para a lesão pulmonar ao longo do tempo.

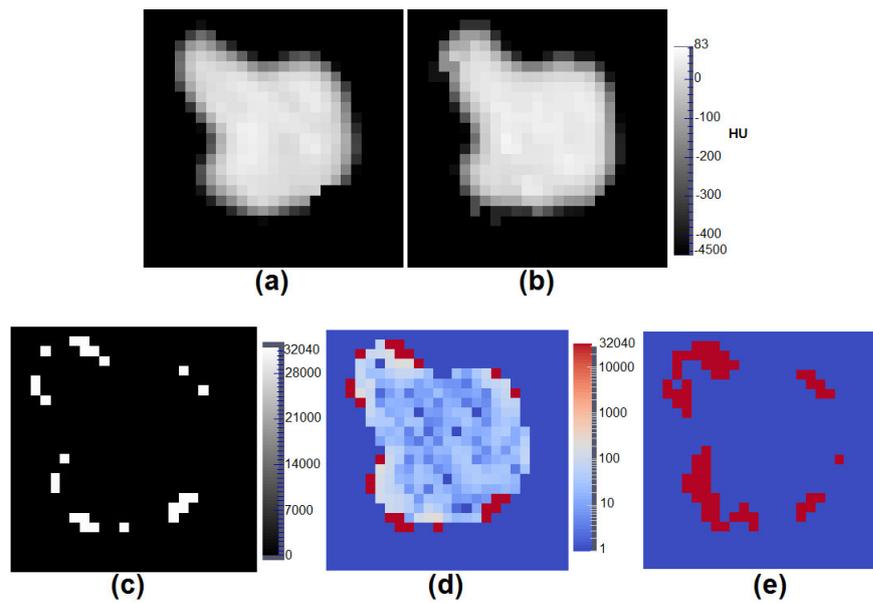
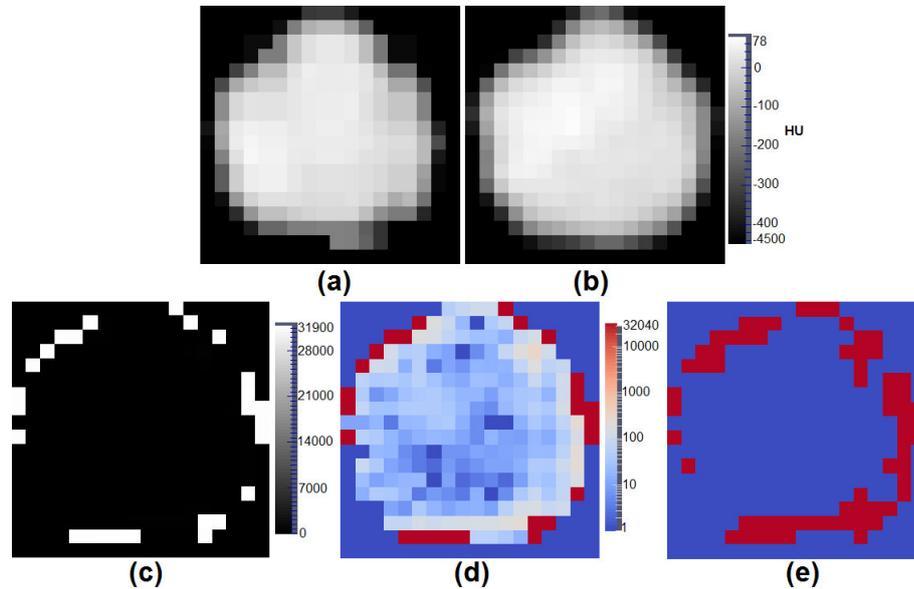


Figura 4.28 – Caso 12 apresentou o segundo menor valor de VVR . Em (a), tem-se a lesão pulmonar em um instante e em (b), tem-se a lesão pulmonar em um instante anterior. A imagem em (c) é a diferença das imagens em (a) e (b). Em (d), tem-se a imagem colorida da imagem diferença em (c). Já a imagem em (e) é a imagem resultante da binarização com k -means da imagem em (c), que enfatiza as mudanças ocorridas para a lesão pulmonar ao longo do tempo.



Fonte: Autor

4.7.6.2 Razão de imagens

Para os mesmos Casos 4 e 12, o método da razão produziu valores de P_m maiores que o método da diferença, 15,95 e 27,36, respectivamente.

Nas Figuras 4.29(c) e 4.30(c), visualiza-se a imagem da razão para os Casos 4 e 12, respectivamente. Nas imagens das figuras, existe a escala de variação da fase que informa o quanto os valores dos voxels são distintos. As Figuras 4.29(d) e 4.30(d) ilustram o comportamento da fase com uma palheta colorida, que facilita a visualização dos *voxels* cujas mudanças foram mais intensas. As Figuras 4.29(e) e 4.30(e) ilustram o resultado da binarização da imagem da razão, no qual a cor em vermelho representa os *voxels* que sofreram as maiores mudanças; e em azul, os *voxels* que sofreram as menores mudanças.

Figura 4.29 – Caso 4 apresentou o menor valor de VVR . Em (a), tem-se a lesão pulmonar em um instante e em (b), tem-se a lesão pulmonar em um instante anterior. A imagem em (c) é a razão das imagens em (a) e (b). Em (d), tem-se a imagem colorida da imagem razão em (c). Já a imagem em (e) é a imagem resultante da binarização com k -means da imagem em (c), que enfatiza as mudanças ocorridas para a lesão pulmonar ao longo do tempo.

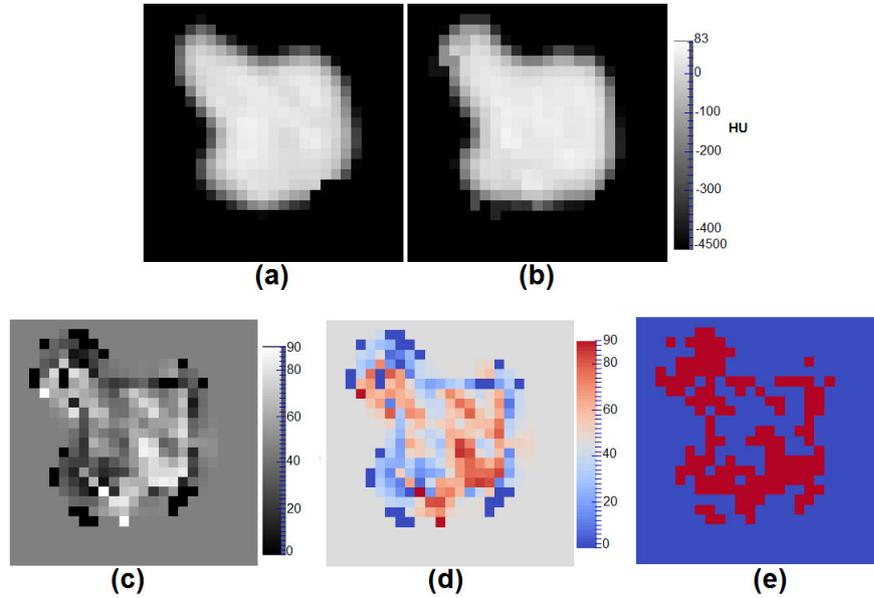
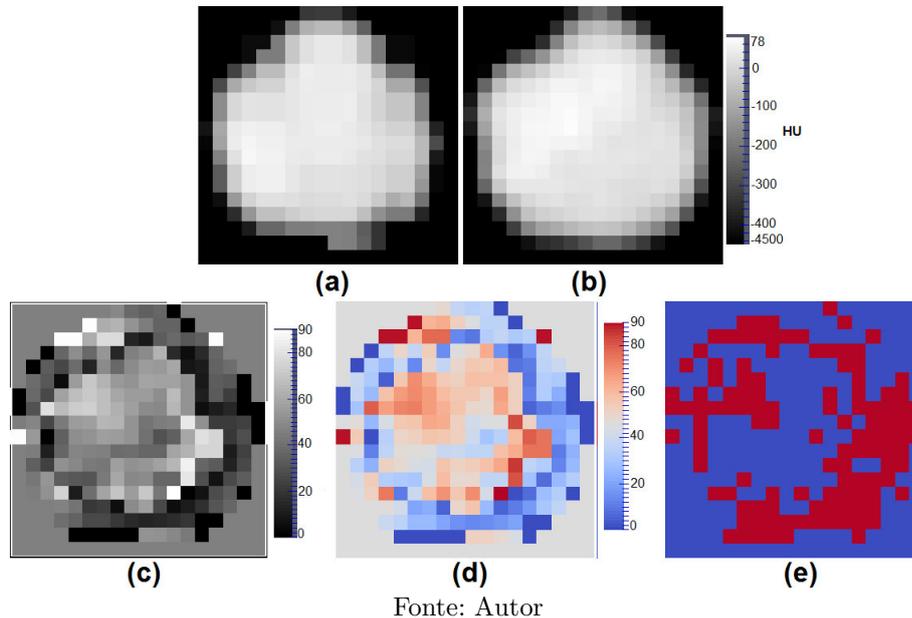


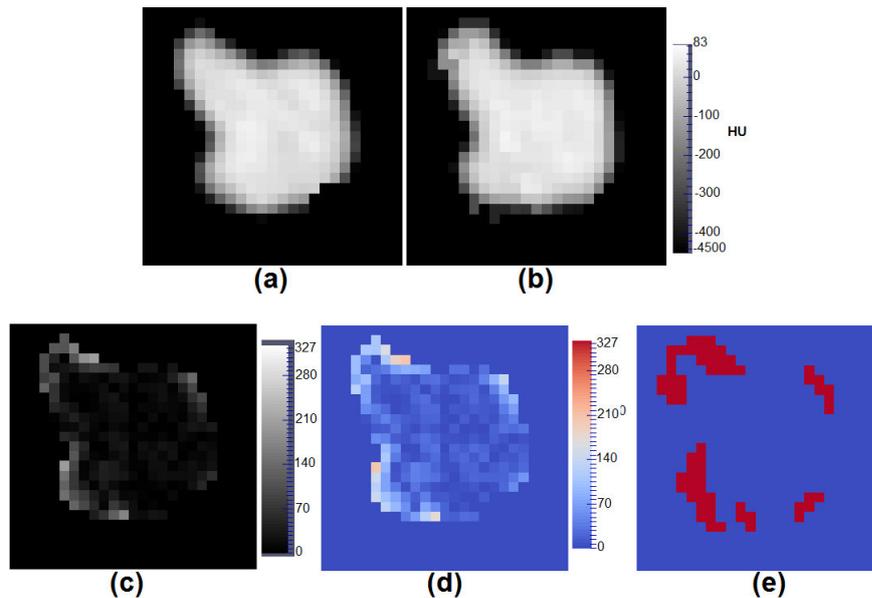
Figura 4.30 – Caso 12 apresentou o segundo menor valor de VVR . Em (a), tem-se a lesão pulmonar em um instante e em (b), tem-se a lesão pulmonar em um instante anterior. A imagem em (c) é a razão das imagens em (a) e (b). Em (d), tem-se a imagem colorida da imagem razão em (c). Já a imagem em (e) é a imagem resultante da binarização com k -means da imagem em (c), que enfatiza as mudanças ocorridas para a lesão pulmonar ao longo do tempo.



4.7.6.3 Regressão de imagem

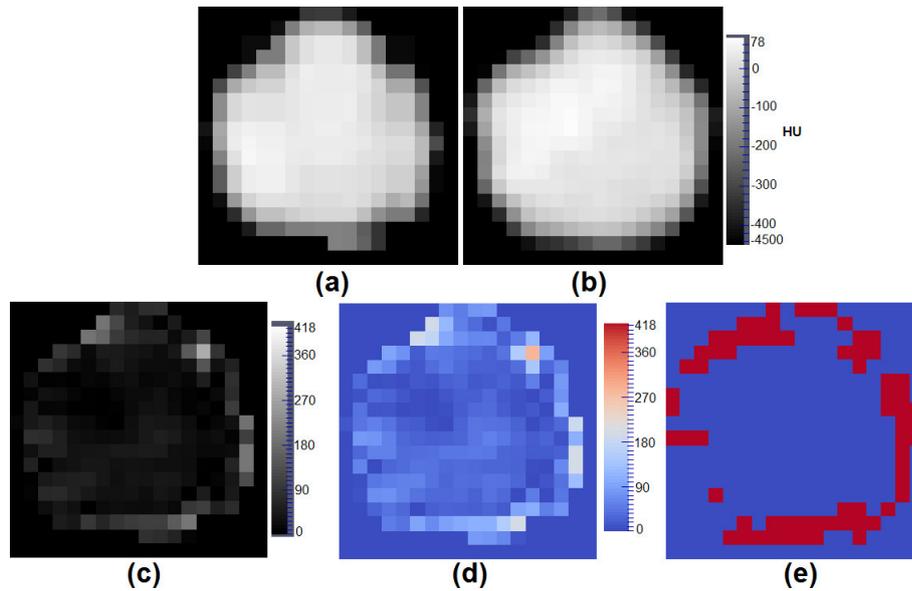
Os valores de P_m para o método foram 9,53 e 18,60 para os Casos 4 e 12, respectivamente. Nas Figuras 4.31(c) e 4.32(c), visualiza-se a imagem regressão, respectivamente, para os Casos 4 e 12. As Figuras 4.31(d) e 4.32(d) ilustram a imagem da regressão com uma palheta colorida, que facilita a visualização dos *voxels* cujas mudanças foram mais intensas. As Figuras 4.31(e) e 4.32(e) ilustram o resultado da binarização da imagem da regressão, onde o vermelho são *voxels* que sofreram as maiores mudanças; e em azul, os *voxels* que sofreram as menores mudanças.

Figura 4.31 – Caso 4 de menor valor de VVR . Em (a), tem-se a lesão pulmonar em um instante e em (b), tem-se a lesão pulmonar em um instante anterior. A imagem em (c) é a imagem criada da regressão das imagens em (a) e (b). A imagem em (d) é a imagem colorida da imagem em (c). Já a imagem em (e) é a imagem resultante da aplicação do *k-means* da imagem em (c), que enfatiza as mudanças ocorridas para a lesão pulmonar ao longo do tempo.



Fonte: Autor

Figura 4.32 – Caso 12 apresentou o segundo menor valor de VVR . Em (a), tem-se a lesão pulmonar em um instante e em (b), tem-se a lesão pulmonar em um instante anterior. A imagem em (c) é a imagem criada da regressão das imagens em (a) e (b). A imagem em (d) é a imagem colorida da imagem em (c). Já a imagem em (e) é a imagem resultante da aplicação do k -means da imagem em (c), que enfatiza as mudanças ocorridas para a lesão pulmonar ao longo do tempo.



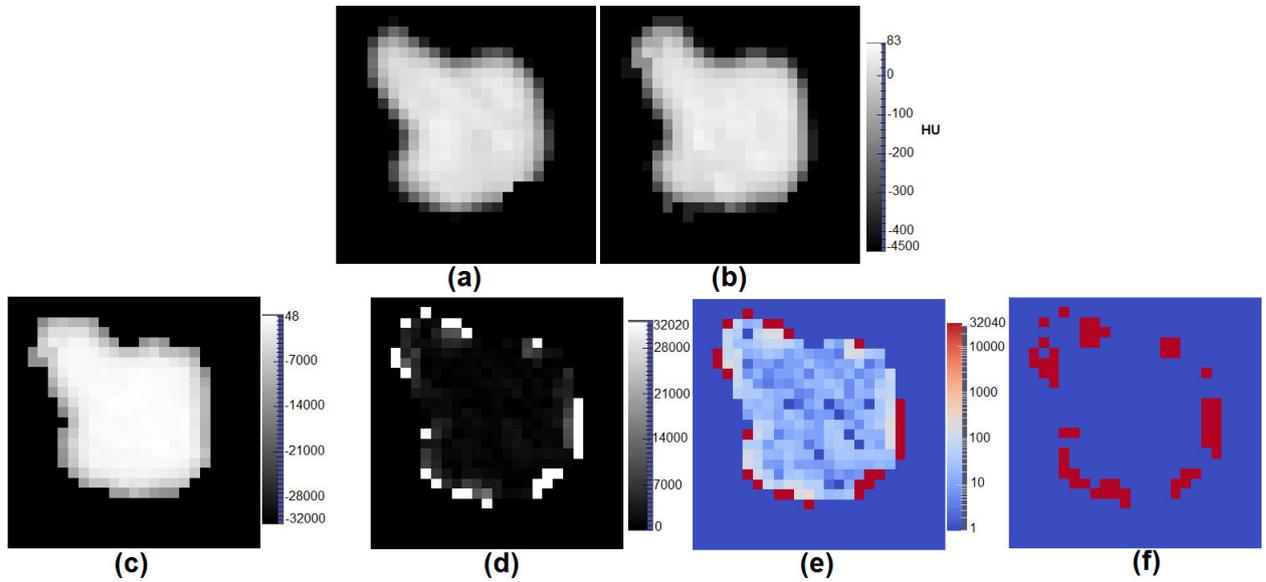
Fonte: Autor

4.7.6.4 Imagem mediana

O método produziu valores de P_m de 10,59 e 16,19 para os Casos 4 e 12, respectivamente. Esses valores são menores em comparação com o método da razão e próximos dos valores P_m obtidos pelo método da diferença.

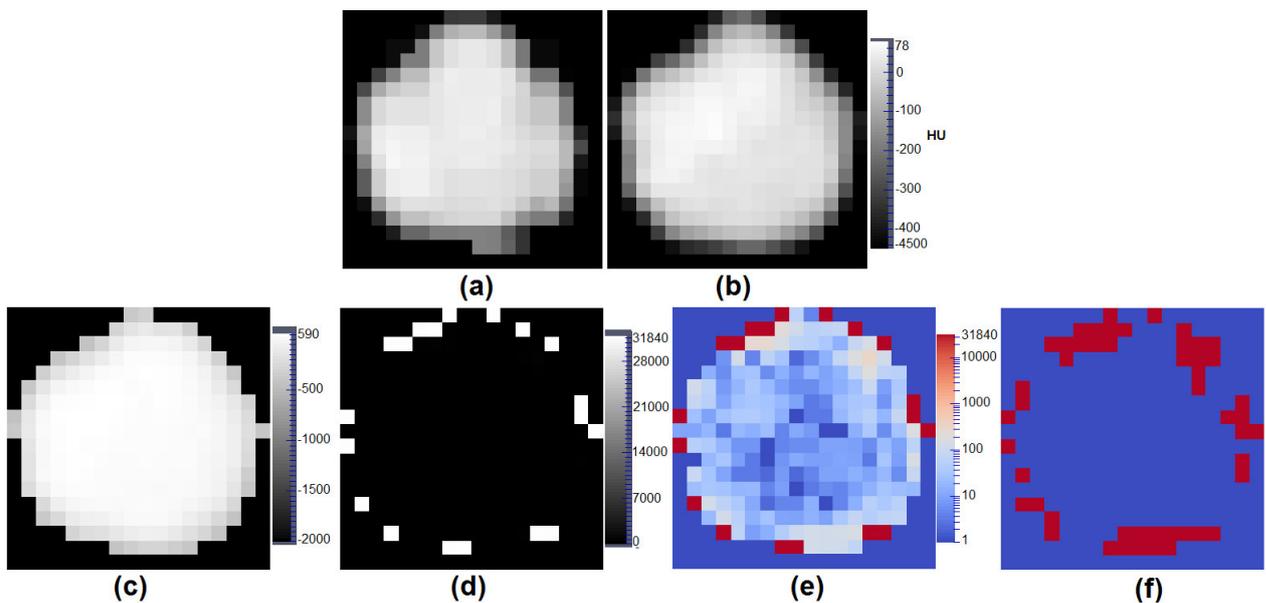
Nas Figuras 4.33(c) e 4.34(c), visualiza-se a imagem da mediana, respectivamente, para os Casos 4 e 12. As Figuras 4.33(d) e 4.34(d) ilustram a imagem da diferença com uma palheta colorida, que facilita a visualização dos *voxels* cujas mudanças foram mais intensas. As Figuras 4.33(e) e 4.34(e) ilustram o resultado da binarização da imagem da mediana, onde o vermelho são *voxels* que sofreram as maiores mudanças; e em azul, os *voxels* que sofreram as menores mudanças.

Figura 4.33 – Caso 4 de menor valor de VVR . Em (a), tem-se a lesão pulmonar em um instante posterior e em (b), tem-se a lesão pulmonar em um instante anterior. A imagem em (c) é a imagem mediana de (a) e (b). A diferença entre a imagem mediana (c) com a imagem em (b) resulta na imagem em (d), sendo que as regiões que sofreram mudanças estão em (f). Em (e), tem-se a imagem colorida da imagem diferença em (d).



Fonte: Autor

Figura 4.34 – Caso 12 apresentou o segundo menor valor de VVR . Em (a), tem-se a lesão pulmonar em um instante posterior e em (b), tem-se a lesão pulmonar em um instante anterior. A imagem em (c) é a imagem mediana de (a) e (b). A diferença entre a imagem mediana (c) com a imagem em (b) resulta na imagem em (d), sendo que as regiões que sofreram mudanças estão em (f). Em (e), tem-se a imagem colorida da imagem diferença em (d).

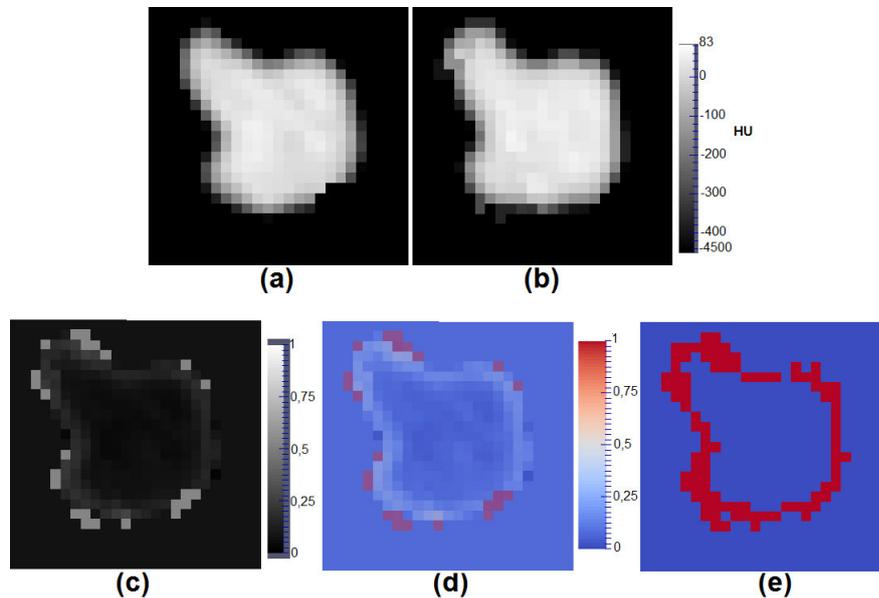


Fonte: Autor

4.7.6.5 Operador XOR *fuzzy*

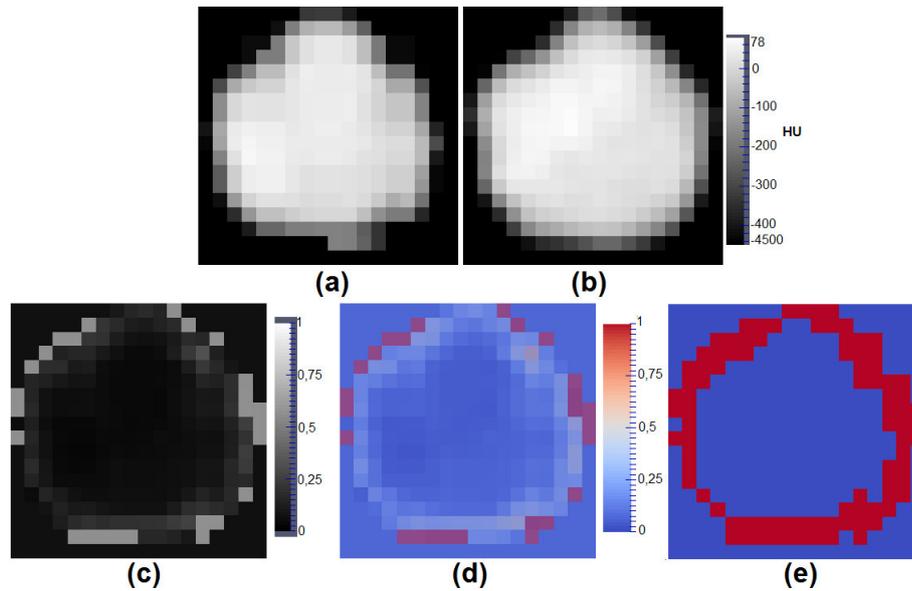
Os valores de P_m para o método foram 13,78% e 24,10%, respectivamente, para os Casos 4 e 12. Nas Figuras 4.35(c) e 4.36(c), visualizam-se as imagens resultantes da operação XOR *fuzzy*, respectivamente, para os Casos 4 e 12. As Figuras 4.35(d) e 4.36(d) ilustram esta mesma imagem com uma palheta colorida, que facilita a visualização dos *voxels* cujas mudanças foram mais intensas. As Figuras 4.35(e) e 4.36(e) possuem o resultado da binarização da imagem resultante da operação XOR *fuzzy*, onde o vermelho são os *voxels* que sofreram as maiores mudanças; e em azul, os *voxels* que sofreram as menores mudanças.

Figura 4.35 – Caso 4 de menor valor de VVR . Em (a), tem-se a lesão pulmonar em um instante e em (b), tem-se a lesão pulmonar em um instante anterior. A imagem em (c) é a imagem resultante da operação XOR *fuzzy* das imagens em (a) e (b). A imagem em (d) é a imagem colorida da imagem em (c). Já a imagem em (e) é a imagem resultante da binarização com *k-means* da imagem em (c), que enfatiza as mudanças ocorridas para a lesão pulmonar ao longo do tempo.



Fonte: Autor

Figura 4.36 – Caso 12 apresentou o segundo menor valor de VVR . Em (a), tem-se a lesão pulmonar em um instante e em (b), tem-se a lesão pulmonar em um instante anterior. A imagem em (c) é a imagem resultante da operação XOR *fuzzy* das imagens em (a) e (b). A imagem em (d) é a imagem colorida da imagem em (c). Já a imagem em (e) é a imagem resultante da binarização com *k-means* da imagem em (c), que enfatiza as mudanças ocorridas para a lesão pulmonar ao longo do tempo.

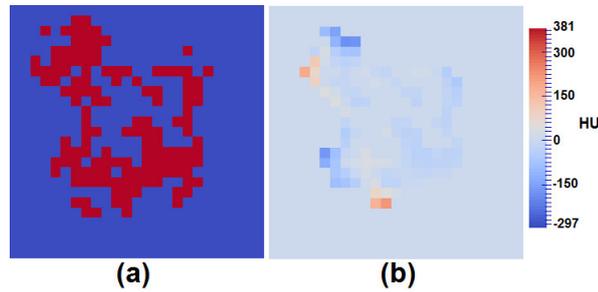


Fonte: Autor

4.7.7 Classificação de lesões quanto ao diagnóstico

O objetivo desta análise é verificar se as mudanças de densidade encontradas por cada um dos métodos baseados em *voxel* corroboram com o diagnóstico. Outro objetivo é verificar se o modelo de classificação encontrado com medidas de textura consegue prever o diagnóstico da lesão pulmonar. Para isso, as medidas de textura são extraídas apenas das regiões onde ocorreram as mudanças. Um exemplo das regiões em que são extraídas as medidas de textura pode ser visto na Figura 4.37, para o resultado da razão de imagens. Nessa figura, tem-se, em (a), a imagem em binário com as regiões em vermelho indicando as mudanças e a imagem em (b) é a diferença das imagens da lesão para essas regiões de mudança.

Figura 4.37 – Exemplo das regiões em que serão extraídas as características de textura para classificação de lesões em maligno e benigno.



Fonte: Autor

As características utilizadas na tarefa de classificação foram definidas na Seção 3.5. O total de características obtidas foram 36. Em razão do grande número de características, realizou-se a seleção de características com o método *Random Forest and RM-bound MVS* (CHEN; LIN, 2006). Da seleção, independente do método de detecção de mudança utilizado, todas as medidas estatísticas foram selecionadas. Logo, um total de 4 características foi utilizado para todos os métodos propostos.

Devido a pequena amostra para a tarefa de classificação, foi utilizada a técnica de amostragem *leave-one-group-out* (KOHAVI, 1995), como forma de avaliar o modelo gerado na classificação. O número de grupos utilizado foi 10.

O classificador utilizado foi o MVS com *kernel RBF* (Seção 3.6.1) e os parâmetros de classificação, C e g , para cada um dos métodos, estão na Tabela 4.5. A avaliação foi realizada em alguns casos com a estimação do C e do g e utilizando os valores padrões, isto é, $C = 1$ e $g = 1/4$, onde 4 é o número de características selecionadas. Nenhum ajuste para os pesos entre as classes foi utilizado.

O resultado da classificação quanto ao diagnóstico para a amostra contendo as 63 instâncias, das quais 17 instâncias são casos benignos (base privada) e 46 instâncias são casos malignos (base pública), estão na Tabela 4.5 para cada um dos métodos utilizados.

4.8 Discussão

Para a base pública de lesões, o valor de P_m encontrado para o Caso 10 foi maior que para o Caso 28 por todos os métodos baseados em *voxel*, mesmo o Caso 10 apresentando o menor valor de VVR . O método de regressão foi o menos sensível às variações de densidade em comparação com os outros métodos.

Tabela 4.5 – Métricas de avaliação da classificação utilizando MVS associadas a cada um dos métodos baseados em *voxel*.

Métodos	parâmetros do MVS		Métricas de avaliação						
	C	g	sens(%)	esp(%)	ac(%)	vp	vn	fp	fn
Diferença	1	1/4	97,83	100,00	98,41	45	17	0	1
Diferença	0,5	0,007813	97,83	100,00	98,41	45	17	0	1
Razão	1	1/4	95,65	94,12	95,24	44	16	1	2
Razão	2048	0,000122	100,00	94,12	98,41	46	16	1	0
Mediana	1	1/4	95,65	100,00	96,83	44	17	0	2
Mediana	8	0,000488	97,83	100,00	98,41	45	17	0	1
Regressão	1	1/4	97,83	100,00	98,41	45	17	0	1
Regressão	32	0,000488	97,83	100,00	98,41	45	17	0	1
XOR	1	1/4	97,83	64,71	88,89	44	11	6	1
XOR	2	8	97,83	88,24	95,24	45	15	2	1

Fonte: Autor

Na base privada de lesões, os pequenos valores de *VVR* podem ser resultantes do processo de segmentação da lesão. Logo, as mudanças externas ou mudanças lesão-fundo são facilmente detectadas pelos métodos.

Os métodos da razão e do operador *XOR-fuzzy* foram os mais sensíveis às variações de densidade da lesão ao longo do tempo, pois geralmente apresentam os maiores percentuais de mudança P_m . Para o método da razão, a fase facilmente se distancia de $\pi/4$ com a diferença entre os valores de *voxel*. Enquanto o operador *XOR-fuzzy* fica mais próximo de um, quanto maior for a diferença entre os *voxels*. Esse último é fortemente dependente da escala de variação dos valores dos *voxels*.

O desbalanceamento da base de lesões não teve forte influência na classificação, visto que o classificador MVS é bastante sensível ao desbalanceamento dos dados de treinamento.

O grande valor de C estimado demonstra que houve *overfitting* para a classificação com base nos resultados do método da razão de imagem. Também foi o método que apresentou o melhor resultado para a classificação, pois a sensibilidade quanto aos malignos foi 100% e a especificidade foi 94,12% (1 falso positivo). Entretanto, o *overfitting* não é aceitável em problemas de classificação, pois o classificador não generaliza para novas instâncias. Portanto, o resultado da classificação para o método da razão não é o mais apropriado para a tarefa.

Os métodos da diferença, mediana e regressão com estimação dos parâmetros tiveram os mesmos resultados para classificação. A maioria dos métodos resultaram em, no mínimo, um falso negativo, com exceção do método da razão com os parâmetros estimados.

4.9 Conclusão

O conjunto de métodos propostos mostra que é possível realçar valores de densidade para uma lesão pulmonar obtida em vários exames, de maneira a enfatizar as regiões onde ocorreram mudanças de densidade ao longo do tempo. Os métodos empregaram várias técnicas básicas de processamento de imagem, com o objetivo de fornecer informações de mudança de densidade ao especialista e facilitar as ações apropriadas em cada caso. Métodos de detecção de mudanças que originalmente eram utilizados nas análises de imagens de sensoriamento remoto foram utilizados neste trabalho para a tarefa de detectar mudanças nas lesões. Logo, após cada método detectar a localização dessas mudanças nas regiões da lesão, quantifica-se essas mudanças sobre as lesões, que podem ou não estar em tratamento.

A análise foi realizada sobre lesões malignas que estavam em tratamento e sobre nódulos pulmonares indeterminados, mas com alta probabilidade de serem benignas.

Com o intuito de explorar melhor os resultados encontrados, foram analisados em detalhes situações em que as lesões sofreram pequenas variações volumétricas. Os casos analisados estão entre as situações cotidianas mais difíceis, pois podem ser imperceptíveis durante a análise do especialista. Os métodos foram capazes de indicar nesses casos, que as lesões tiveram variações de densidade. Isso sugere que houve modificações em seus tecidos ao longo do tempo, independente da lesão pulmonar estar em tratamento ou não.

Os métodos baseados em *voxel* foram sensíveis às variações nos valores de densidade da lesão e foram capazes de realçar e mensurar as regiões que possuíam mudanças ao longo do tempo. Logo, com base nos resultados encontrados, pode-se dizer que são bons métodos para análise de mudanças em lesões em 3D. Entretanto, ainda existe a necessidade da realização de novos estudos com outras bases de imagens para confirmá-los como métodos complementares à medida de variação relativa do volume.

5 Métodos de estimação de incerteza

Os métodos e as técnicas presentes neste capítulo estão relacionados aos conceitos estatísticos de incerteza, base para o entendimento das técnicas propostas para detecção das mudanças de densidade. Inicialmente, o conceito de incerteza é definido. Em seguida, os métodos de amostragem são apresentados. A partir desses métodos, as distribuições populacionais são estimadas e os métodos de comparação das distribuições são utilizados para detecção das mudanças de densidade das imagens das lesões pulmonares. A detecção das mudanças é então realizada em duas abordagens, global e local. Portanto, as duas abordagens propostas determinam se as lesões pulmonares adquiridas em diversos instantes são diferentes ou semelhantes, com base nas suas distribuições das médias de densidade.

5.1 Incerteza

O grau de incerteza pode ser avaliado em termos das probabilidades das distribuições amostrais de alguma estatística estimada para uma população de dados ([MARTINS, 2005](#)). Portanto, mesmo que o valor da estatística estimada para a população varie com a amostra selecionada, a sua distribuição de probabilidade revelará o seu grau de incerteza. Nesse trabalho, a incerteza é extraída das inferências estatísticas realizadas através das amostras aleatórias selecionadas da população de estudo.

Essa estimação da estatística da população pode ser realizada por um intervalo provável, dado um nível de confiança ou probabilidade. Se o intervalo é pequeno, existe um elevado grau de precisão da inferência. Logo, a incerteza se reflete nos possíveis valores para a estatística estimada.

A análise de incerteza deste trabalho utiliza os métodos *bootstrap* e o *moving block bootstrap* (MBB). Esses métodos são uma forma não-paramétrica de estimar uma estatística de uma população através de uma amostra. Até onde se pesquisou, são poucos os recentes trabalhos nesta área de estudo. Apesar disso, encontrou-se o trabalho de [Gassiat, Cleynen e Robin \(2016\)](#), que analisa o uso de uma forma não-paramétrica da cadeia oculta de Markov para resolver problemas de classificação. Em adição, [Wang e Wertenleki \(2013\)](#) propõem um novo estimador de densidade de *kernel* não-paramétrico para inferência da circunferência da cabeça em regiões afetadas pela radiação em Chernobyl. Apesar desses

recentes trabalhos, com técnicas recentes de análise não-paramétrica, utilizou-se o *bootstrap* e o MBB pela simplicidade e robustez dos mesmos.

5.2 Método *Bootstrap*

O método *Bootstrap* é um método não-paramétrico para a estimação de uma estatística e definição do seu intervalo de confiança (WASSERMAN, 2010). É possível estimar alguma característica da população, por exemplo, a média ou desvio padrão, partindo apenas de uma amostra da mesma. A tarefa de estimação é definida pela obtenção de uma estatística em sucessivas pseudo amostras geradas aleatoriamente de uma amostra original com reposição (MOONEY; DUVAL; DUVAL, 1993). Na Figura 5.1, pode-se visualizar a repetição do processo de amostragem, com reposição, da amostra X_i , gerando as sucessivas amostras Y_i . A distribuição da estatística obtida em cada pseudo amostra, também chamada de amostra *bootstrap*, permite a inferência sobre a região mais provável da estatística θ a ser estimada e do seu intervalo de confiança.

De acordo com Wasserman (WASSERMAN, 2010), se $T_n = g(x_1, \dots, x_n)$ é uma estatística do conjunto de dados $\{x_1, \dots, x_n\}$, a variância é uma forma de encontrar o intervalo provável para a estatística T_n . A variância é indicada por $V_F(T_n)$ e obtida da distribuição F dos dados da população. Porém, essa distribuição é de difícil obtenção, visto que podem existir muitos dados ou não é possível coletá-los de maneira fácil.

O método *bootstrap* estima $V_F(T_n)$ através da distribuição aproximada \hat{F} . Fazendo isso, uma aproximação de $V_F(T_n)$ é calculada como $V_{\hat{F}}(T_n)$.

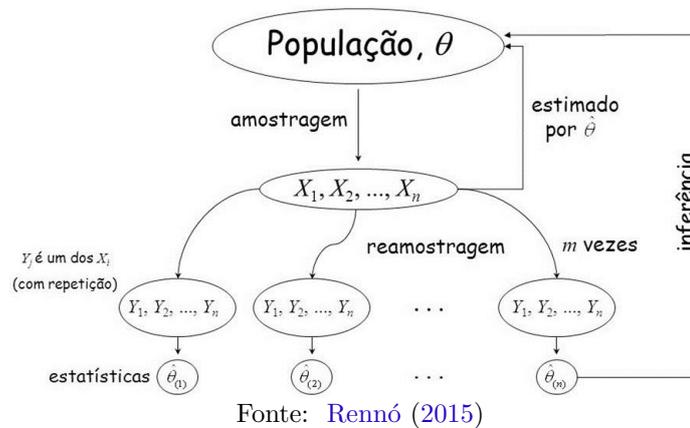
O método *bootstrap* é utilizado neste trabalho para estimar a distribuição da média populacional dos valores dos *voxels* da lesão pulmonar nos diversos instantes de aquisição. As distribuições encontradas são utilizadas para compará-las. A comparação é realizada através da medida de divergência e do teste de hipótese.

5.2.1 Método *moving block bootstrap*

Quando os dados da amostra não são independentes, um segundo e mais robusto método é usado, chamado *moving block bootstrap*.

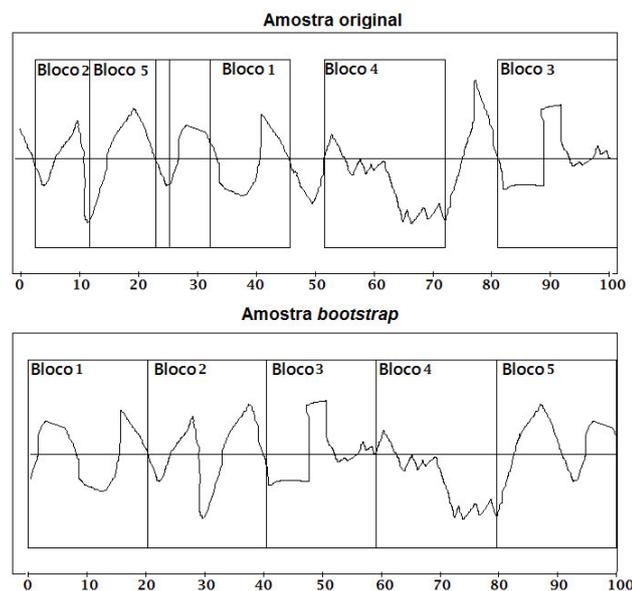
O *moving block bootstrap* (MBB) (EFRON; TIBSHIRANI, 1994), (KYBIC, 2010) é um método de reamostragem que amostra bloco de dados espacialmente consecutivos

Figura 5.1 – Processo de obtenção do parâmetro estatístico θ da população através de uma amostra, pelo método *bootstrap*.



de tamanho L da amostra original. A Figura 5.2 ilustra como uma amostra *bootstrap* é formada da amostra original. A amostra original é um sinal (por exemplo, corrente elétrica) cuja amplitude varia sobre o tempo e tem tamanho $n = 100$. Foram selecionados cinco blocos aleatórios de comprimento $L = 20$ da amostra original, formando a amostra *bootstrap*.

Figura 5.2 – Formação da amostra *bootstrap* pelo método MBB.



Usou-se o clássico *bootstrap* para a análise global e o MBB para a análise local. O MBB foi utilizado devido a existência de dependência do valor do *voxel* com sua vizinhança. Logo, ambos os métodos foram utilizados para estimar a distribuição da média populacional dos valores dos *voxels* da lesão pulmonar nos diferentes instantes de aquisição.

5.3 Testes de hipóteses

A partir das informações de uma amostra, pode-se tomar decisões sobre uma população. Para tomar essas decisões, precisa-se formular hipóteses ou conjecturas acerca de uma população. Essas suposições, que podem ser ou não verdadeiras, são denominadas hipóteses estatísticas (MARTINS, 2005).

Pode-se formular uma hipótese estatística, chamada de hipótese nula, e representá-la com H_0 , com o objetivo de rejeitá-la ou aceitá-la. A hipótese contrária a nula é denominada de alternativa e representa-se por H_1 . Um exemplo pode ser dado quando se deseja verificar se uma moeda está viciada. A hipótese principal é geralmente formulada contrária a pressuposta, isto é, que a mesma não é viciada e a probabilidade de cara é igual a 0,5, definindo a hipótese nula. Logo, essa hipótese determina que não existe uma face melhor que a outra e quaisquer diferenças observadas são meramente devido as flutuações das amostras provenientes da mesma população. Uma hipótese alternativa possível para esse exemplo é determinada quando a probabilidade de uma face da moeda é maior que 0,5.

Dada uma hipótese e admitindo-a como verdadeira, caso os resultados observados em uma amostra aleatória sejam divergentes para aquela hipótese com base na sua probabilidade, pode-se concluir que as diferenças observadas são significativas e isso motiva a rejeitá-la, ou ainda, não aceitá-las com base nas provas obtidas.

O processo de decidir se uma hipótese é verdadeira ou não é denominado de teste de hipótese. Portanto, durante um teste de hipótese, quando ocorre a rejeição de uma hipótese nula, isso implica na aceitação da hipótese alternativa. É geralmente especificada antes da extração da amostra de modo a não influenciar a sua escolha através dos resultados obtidos e essa probabilidade é representada por um α .

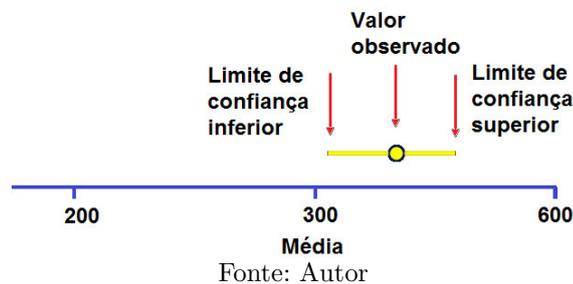
Para que o erro de decisão seja o menor possível durante o teste de hipótese, deve-se especificar a máxima probabilidade que se está dispostos a sofrer com o erro do tipo I, isto é, erro causado quando se rejeita a hipótese nula, mas deveria ser aceita, e é também denominado de nível de significância.

Os valores práticos mais usuais para o nível de significância são 0,01 ou 0,05, ou seja, 1% ou 5%, embora não haja restrição a outros valores. Dado um nível de significância α , temos $100\% \cdot (1 - \alpha)$ de confiança que uma decisão foi tomada corretamente.

5.4 Intervalo de Confiança

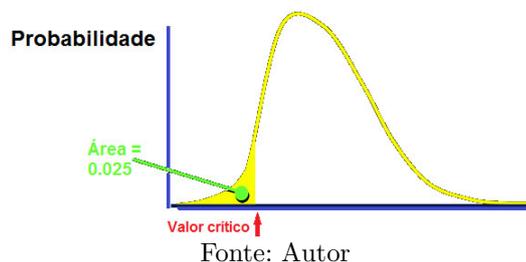
É uma forma de definir uma margem de erro ou incerteza para algum parâmetro da população, através de um intervalo. Também pode exprimir uma margem provável do valor verdadeiro, conforme a Figura 5.3. Na figura, pode-se observar o valor máximo e o mínimo dos valores obtidos para a média, além do valor observado e os limites de confiança inferior e superior para aquela medida. Logo, o intervalo de confiança é determinado pela diferença entre o limite de confiança superior e inferior.

Figura 5.3 – Intervalo de variação da média e seus limites de confiança.



A estimação do intervalo de confiança é realizada através da função de distribuição de probabilidade. Os limites são encontrados com base na área dos extremos da distribuição de probabilidade, conforme se pode visualizar na Figura 5.4. Portanto, o limite inferior foi encontrado no intervalo de variação, no ponto cuja área sob a curva é igual a 0,025 ou 2,5%.

Figura 5.4 – Determinação do limite de confiança inferior pela área da distribuição de probabilidade.



O valor da área para a determinação dos limites de confiança é frequentemente associado ao valor α do nível de significância pela Equação 55. Metade do valor α é utilizado para encontrar o limite de confiança inferior e a outra metade, o limite de

confiança superior. Portanto, o intervalo de confiança fica definido pela Equação 55, que está em função do valor β .

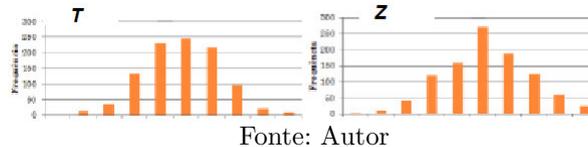
$$[\Theta^\beta, \Theta^{1-\beta}] \quad (55)$$

5.5 Divergência de Jensen

O estudo de medidas de incerteza é de grande utilidade no ramo da matemática chamado teoria da informação. É muito utilizado em problemas de transmissão de informações sobre canais de comunicação.

A divergência $D(T||Z)$ mede a diferença ou a distância entre duas distribuições de probabilidade ou dois conjuntos de pontos T e Z , com $D(T||Z) = 0$, se $T = Z$. A Figura 5.5 mostra essas duas distribuições e a medida de divergência informa o quanto a distribuição T se distancia da Z .

Figura 5.5 – Exemplo de duas distribuições T e Z .



A divergência de Jensen é definida pela Equação 56 e visa quantificar a diferença entre duas distribuições de probabilidade $T = \{t_1, \dots, t_n\}$ e $Z = \{z_1, \dots, z_n\}$ (TANEJA, 2001), resultando sempre em um valor positivo.

$$D(T||Z) = \sum_{i=1}^n \left[\frac{T_i \log T_i + Z_i \log Z_i}{2} - \left(\frac{T_i + Z_i}{2} \right) \log \left(\frac{T_i + Z_i}{2} \right) \right] \quad (56)$$

A medida de divergência foi utilizada para comparar a lesão em dois instantes diferentes nas análises global e local.

5.6 Identificação das mudanças de densidade

Os métodos baseados em incerteza deste trabalho são originados em alguns princípios de inferência estatística, como o teste de hipótese, amostragem *bootstrap* e intervalo de confiança. Os métodos foram utilizados para comparar a lesão em dois instantes diferentes

e duas abordagens para a análise foram propostas, uma global e outra local. A seguir, será mostrado como cada um desses conceitos foram utilizados na tarefa de detecção das mudanças de densidade para a lesão pulmonar.

5.6.1 Análise global de mudanças de lesão

Esta análise consiste em comparar o par das distribuições da média dos *voxels* e o par dos histogramas dos *voxels* da lesão pulmonar, em dois instantes diferentes.

5.6.1.1 Divergência Global de Jensen

Para uma lesão pulmonar em um instante inicial, cada *voxel* da mesma é considerado um elemento da amostra original. A partir dessa amostra original, são criadas amostras *bootstrap* com o método clássico, denominadas de lesões *bootstrap*, dos quais são obtidas as suas médias e gerada a sua distribuição. O mesmo ocorre para a lesão em um instante posterior. Com as duas distribuições da média, é possível compará-las através do cálculo da divergência, o qual é denominado de b-divergência. O processo completo de análise descrito é exemplificado na Figura 5.6.

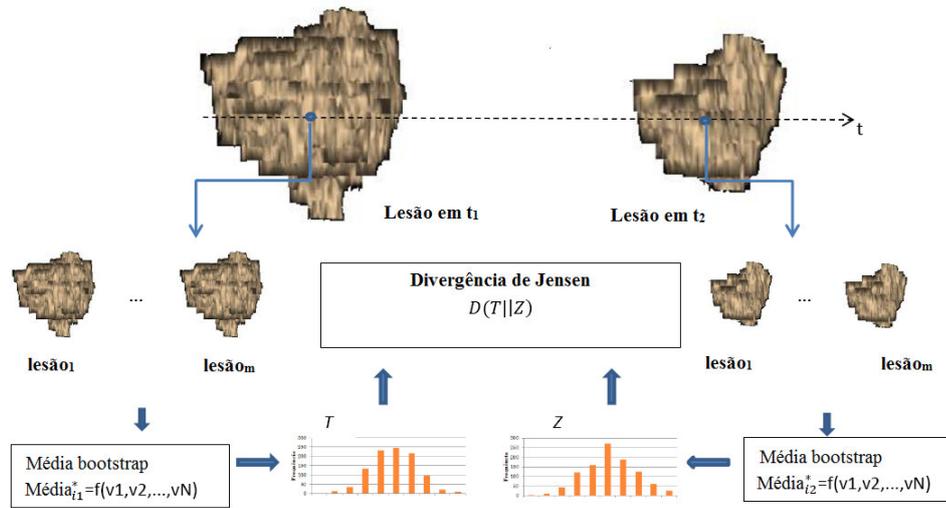
5.6.1.2 Divergência Global de Jensen do Histograma

Com o objetivo de avaliar as mudanças da lesão entre os instantes analisados, propõe-se uma análise baseada em histogramas dos valores dos *voxels* da lesão em cada instante, utilizando a divergência global de Jensen. Conseqüentemente, é possível verificar o quão diferentes são os histogramas do par de volumes da lesão.

5.6.1.3 Teste de hipótese global

Outra maneira de analisar globalmente as lesões pulmonares é utilizando o conceito de inferência estatística chamado teste de hipótese. A partir de uma hipótese, denominada de hipótese nula H_0 , contrária ao que se quer provar, verifica-se o quão provável é essa hipótese através do valor-p e da pré-especificação do nível de significância (DEGROOT; SCHERVISH, 2002; HOSHMAND, 1997).

Figura 5.6 – Fluxo das etapas da estimação global das incertezas associadas à média das intensidades, na comparação da lesão em instantes diferentes.



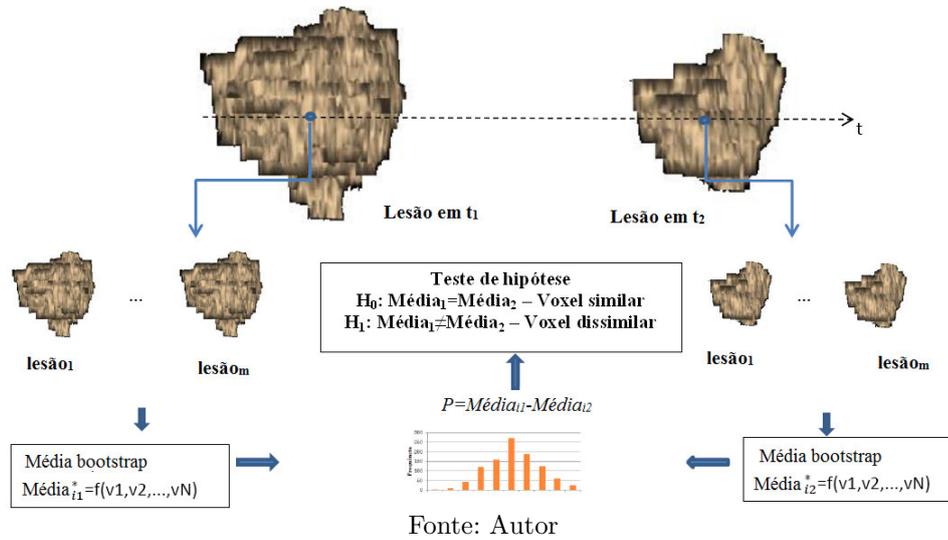
Fonte: Autor

A hipótese nula H_0 utilizada neste trabalho é a de não existência de diferença entre as médias da lesão em instantes diferentes. Essa forma de analisar as diferenças sofridas pela lesão ao longo do tempo usa o mesmo princípio estatístico de comparação de amostras, baseado em suas médias (WASSERMAN, 2010). Dessa forma, pode-se supor que uma lesão se tornou diferente em densidade ao longo do tempo e se a mesma aumentou ou diminuiu, com relação aos valores de intensidade dos seus *voxels*.

O objetivo da análise é verificar se duas amostras são da mesma população. A partir de duas amostras, chamadas de lesões *bootstrap*, encontra-se o intervalo de confiança da diferença entre as médias das suas respectivas amostras *bootstrap*. Cada amostra é formada pelos valores dos *voxels* da lesão em um determinado instante de aquisição. Então, encontrada a distribuição da diferença, pode-se determinar se a lesão se modificou ao longo do tempo.

A comparação da lesão nos dois instantes é realizada através da diferença entre as médias dos valores dos *voxels* das lesões *bootstrap* dos instantes diferentes. Dado um nível de significância, se a diferença nula estiver dentro do intervalo de confiança encontrado pela distribuição da diferença entre as médias, então a hipótese nula não pode ser rejeitada. Logo, a lesão não mudou entre os dois instantes analisados. Caso contrário, a hipótese nula é rejeitada e a lesão é considerada diferente entre os instantes analisados. A Figura 5.7 exemplifica cada um dos passos descritos para o teste de hipótese.

Figura 5.7 – Fluxo das etapas para comparação global da lesão em instantes diferentes através da verificação do teste de hipótese da semelhança entre as médias.



5.6.2 Análise local de mudanças de lesão

A análise consiste em comparar cada *voxel* da lesão em dois diferentes instantes e evidenciar aqueles que foram menos similares. Para a comparação local dos *voxels*, o método MBB foi utilizado. Seu uso é decorrente da existência de dependência do valor do *voxel* em análise com os valores da sua vizinhança.

O processo inicia-se com a determinação das amostras localmente em cada *voxel* nos diferentes instantes. Para cada *voxel*, é obtida uma vizinhança 3D contendo um percentual do volume da lesão. Essa estratégia foi tomada para que o tamanho da vizinhança tivesse informação suficiente para formar a amostra original. Para obter esse percentual, um raio R com centro em cada *voxel* em análise foi utilizado. Logo, cada *voxel* interior à esfera de raio R pertencente a vizinhança forma a amostra original, a partir do qual são geradas as amostras *bootstrap*.

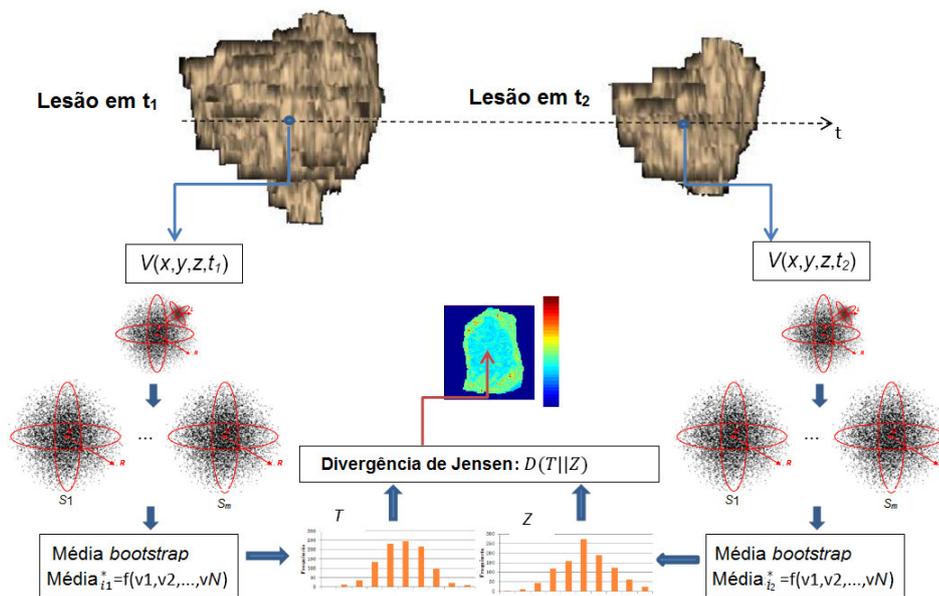
As amostras *bootstrap* foram formadas pelo método MBB e o raio L para formação dos blocos também utiliza um percentual do tamanho da amostra original.

5.6.2.1 Divergência Local de Jensen

Similarmente a análise global, foram encontradas duas distribuições da média das amostras *bootstrap*, uma para cada instante. As duas distribuições do *voxel* da lesão no par de instantes foram comparadas pelo cálculo da divergência de Jensen.

O processo de comparação é simplesmente um realce das diferenças das distribuições *bootstrap*. Quanto menor for a similaridade das distribuições da média da vizinhança local dos *voxels* dos diferentes instantes, maior será o valor da divergência e, portanto, esse *voxel* será evidenciado. A Figura 5.8 ilustra os passos em detalhes. Para cada *voxel* e sua vizinhança (amostra original) definida pelo raio R , blocos de raio L foram selecionados aleatoriamente para formar a i^{th} amostra *bootstrap* S_i .

Figura 5.8 – Fluxo das etapas da estimação local das incertezas associadas à média das intensidades, na comparação *voxel-a-voxel* da lesão em instantes diferentes.



Fonte: Autor

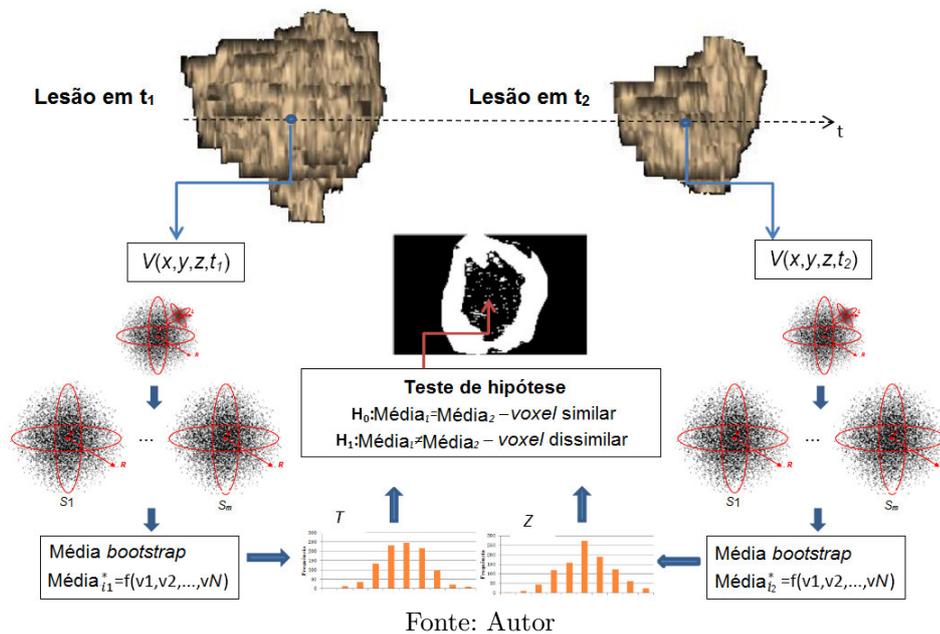
5.6.2.2 Teste de hipótese local

Esta análise local usou a hipótese nula H_0 de que não existe diferença entre as distribuições da média das amostras *bootstrap* dos diferentes instantes analisados. A partir de H_0 , gera-se uma imagem binária da lesão, em que os *voxels* com valor zero significa similaridade e, um, diferença no par de instantes da lesão. Para a formação da imagem, a mesma ideia do intervalo de confiança da diferença das médias *bootstrap* foi usada. Para cada *voxel* da lesão nos diferentes instantes t , a distribuição da média *bootstrap* foi criada pelo método MBB. Portanto, a partir da distribuição formada, pode-se verificar se os *voxels* dos diferentes instantes são diferentes, para um dado nível de significância.

Dado um nível de significância, se a diferença nula estiver no intervalo de confiança da distribuição da diferença, então a hipótese nula não pode ser rejeitada. Isto significa

que o valor do *voxel* não mudou entre os instantes analisados. Caso contrário, a hipótese nula é rejeitada e o valor do *voxel* é considerado diferente entre os dois instantes. A Figura 5.9 ilustra em detalhes estas etapas. Para cada *voxel* e sua vizinhança (amostra original) definida pelo raio R , blocos de raio L foram selecionados para formar a i^{th} amostra *bootstrap* S_i .

Figura 5.9 – Fluxo das etapas para comparação local da lesão em instantes diferentes através da verificação do teste de hipótese da semelhança entre as médias, *voxel-a-voxel*.



5.7 Resultados

Para atingir os resultados esperados na identificação das mudanças de densidade, as seguintes etapas iniciais foram definidas conforme as respectivas seções: base de lesões pulmonares (Seção 4.7.2), modelos de lesões (Seção 4.7.1) e registro das imagens da lesão (Seção 4.7.3). Em seguida, os resultados das configurações de funcionamento dos métodos *bootstrap* e *MBB* são apresentados. Além disso, um conjunto de tabelas é exposto com os resultados da análise global dos métodos propostos. Para verificação da sensibilidade dos métodos propostos às mudanças, outro conjunto de tabelas com os resultados da análise local dos métodos é fornecido. Com o mesmo objetivo, realiza-se um estudo dos casos que apresentam pequena variação volumétrica. Por fim, apresentam-se os resultados da classificação com base nas regiões identificadas como mudança.

Os resultados são apresentados para os métodos do teste de hipótese e divergência de Jensen em suas abordagens globais e locais. O número de amostras *bootstrap* foi 1000, independente da análise. Para a análise local, o percentual do volume utilizado para determinar o número de elementos na vizinhança de cada *voxel* foi 10%. O percentual usado para determinar o tamanho do bloco no método MBB foi 20%.

As Tabelas 5.2 e 5.1 resumem os resultados encontrados para os métodos propostos em sua abordagem global. A coluna denominada **Diferença** refere-se ao resultado da aplicação do teste de hipótese (Seção 5.3) e a letra D indica que houve mudanças nas densidades da lesão ao longo do tempo, enquanto a letra I indica que a lesão não se modificou de maneira significativa, isto é, ao nível de significância de 5%.

A Tabela 5.2 possui os resultados da aplicação dos métodos de teste de hipótese e divergência de Jensen para a base privada de lesões. As suas colunas possuem o mesmo significado das colunas da Tabela 5.1.

Na coluna denominada **b-divergência**, tem-se o resultado da aplicação da divergência de Jensen de forma global para a distribuição da média *bootstrap* e informa o quanto a lesão em instantes diferentes se modificou ao longo do tempo com relação ao valor médio dos valores de *voxel*. A coluna denominada **Divergência** é o resultado da divergência de Jensen para os histogramas dos valores dos *voxels* da lesão. A coluna denominada P_m possui os valores de percentuais de mudança e a coluna denominada VVR possui os valores de variação relativa do volume para cada lesão.

Os valores para b-divergência e divergência das Tabelas 5.2 e 5.1 são os valores escalados dos dados reais para o intervalo entre 0 e 1, em que valores próximos de 1 indicam máxima divergência entre os exames analisados. Os valores reais de b-divergência mínimo e máximo para a base pública de lesões foram, respectivamente, 0,262701 e 0,693147. Para a base privada de lesões, o mínimo e, o máximo encontrados para b-divergência foram, respectivamente, 0,076551 e 0,693147.

Analisando a Tabela 5.1, percebe-se que apenas 4 lesões foram consideradas semelhantes para o teste de hipótese da diferença entre as médias *voxel* das lesões (Casos: 3, 22, 29 e 42). Isto significa que não houve mudanças de densidades de maneira significativa. Os casos restantes foram considerados diferentes com o nível de significância de 5%. Logo, pode-se perceber, para esses casos, que as lesões sofreram mudanças expressivas em suas densidades.

Tabela 5.1 – Os resultados da aplicação do teste de hipótese global estão na Coluna **Diferença**. A Coluna **b-Divergência** possui a divergência global da lesão *bootstrap* de cada caso referente à base pública de lesões. A Coluna **Divergência** contém a divergência global para o histograma dos **voxels**. A Coluna P_m agrupa os valores percentuais de mudança para o teste de hipótese local e a Coluna **VVR**, a variação volumétrica relativa.

Caso	Diferença	b-Divergência	Divergência	P_m	VVR
1	D	1,0000	0,0621	40,0055	-48,30
2	D	1,0000	0,0639	39,9587	-70,10
3	I	0,4473	0,0806	37,4505	-42,17
4	D	1,0000	0,0391	13,2403	133,60
5	D	1,0000	0,2865	15,9144	-29,09
6	D	1,0000	0,4341	25,0543	257,66
7	D	1,0000	0,7822	21,1285	281,26
8	D	1,0000	0,2822	13,1888	6,60
9	D	1,0000	0,0738	25,1010	22,48
10	D	1,0000	0,0509	14,9333	-1,61
11	D	1,0000	0,1595	25,5742	-19,67
12	D	1,0000	0,4409	30,2703	121,24
13	D	1,0000	0,0383	26,5248	75,19
14	D	1,0000	0,3295	14,8581	185,52
15	D	1,0000	0,3124	23,2270	162,29
16	D	1,0000	0,0000	23,0482	8,86
17	D	1,0000	0,8342	3,83942	-17,94
18	D	1,0000	0,0811	23,8137	275,08
19	D	1,0000	0,0763	28,7922	-34,20
20	D	0,9012	0,4729	19,4891	-13,72
21	D	1,0000	0,0299	15,3646	-45,58
22	I	0,1129	0,3243	29,3587	55,84
23	D	1,0000	0,0565	26,7686	32,52
24	D	1,0000	0,8991	10,3490	-4,77
25	D	1,0000	0,9309	12,0703	11,68
26	D	1,0000	0,9311	16,7975	-7,74
27	D	1,0000	0,1188	14,6127	17,27
28	D	1,0000	0,0916	9,6350	-3,12
29	I	0,0000	0,1355	15,6617	-17,39
30	D	0,9638	0,2190	26,4369	17,29
31	D	1,0000	0,0287	15,2436	-5,29
32	D	1,0000	0,4579	6,9329	4,55
33	D	1,0000	0,6253	6,3058	10,38
34	D	1,0000	0,8974	11,7401	-11,22
35	D	1,0000	0,9158	16,3508	30,35
36	D	1,0000	0,9914	19,5111	102,14
37	D	1,0000	0,0483	16,1704	46,83
38	D	1,0000	0,1189	20,2158	127,69
39	D	0,9514	0,0816	20,1103	55,08
40	D	1,0000	0,1474	38,2426	26,74
41	D	1,0000	0,1343	22,4373	95,76
42	I	0,5508	1,0000	34,3873	521,34
43	D	1,0000	0,1156	14,5568	-17,12
44	D	1,0000	0,0232	21,3592	26,81
45	D	1,0000	0,2988	35,0810	-46,43
46	D	1,0000	0,3720	24,2171	-57,75

Fonte: Autor

Tabela 5.2 – Os resultados da aplicação do teste de hipótese global estão na Coluna **Diferença**. A letra I indica que as modificações na lesão não foram expressivas e a letra D o oposto. A Coluna **b-Divergência** possui a divergência global da lesão **bootstrap** de cada caso referente à base privada de nódulos pulmonares. A Coluna **Divergência** contém a divergência global para o histograma dos *voxels*. A Coluna P_m agrupa os valores percentuais de mudança para o teste de hipótese local e a Coluna **VVR**, a variação volumétrica relativa.

Caso	Diferença	b-Divergência	Divergência	P_m	VVR
1	D	1,0000	0,2382	7,49705	-2,86
2	D	1,0000	0,2627	10,9029	5,71
3	D	0,9846	0,0766	10,1257	8,82
4	I	0,1018	0,0774	5,7600	1,35
5	D	1,0000	0,2756	11,1262	7,14
6	D	0,8974	0,1355	19,4104	10,29
7	I	0,0000	0,3284	30,1282	7,41
8	D	0,9923	0,3687	24,8028	20,37
9	I	0,6496	0,3462	30,0296	9,26
10	D	0,9090	0,2955	25,8984	12,07
11	I	0,0314	0,2673	26,9922	-9,23
12	I	0,5240	0,2627	21,3524	1,72
13	D	1,0000	0,7837	27,3016	12,50
14	D	1,0000	1,0000	35,4340	-60,00
15	I	0,2932	0,0000	11,5039	-3,51
16	D	1,0000	0,3507	24,6072	-23,64
17	D	1,0000	0,5046	23,2955	-26,32

Fonte: Autor

As b-divergências relativas da Tabela 5.1 mostram que os Casos 3, 22, 29 e 42 tiveram valores próximos do valor zero. O contrário ocorreu com os casos restantes que sofreram mudanças, pois tiveram valores de b-divergência mais próximos de 1,0.

A divergência relativa dos histogramas foi próxima de zero para os Casos 3, 22 e 29. Esse resultado indica que os histogramas para as lesões não foram muito diferentes. Em oposição a esses casos, o Caso 42 apresentou a maior divergência relativa, visto que a mesma apresentou o maior aumento de volume, 521,34%, o que tem como consequência mudanças nas distribuições dos valores dos *voxels*.

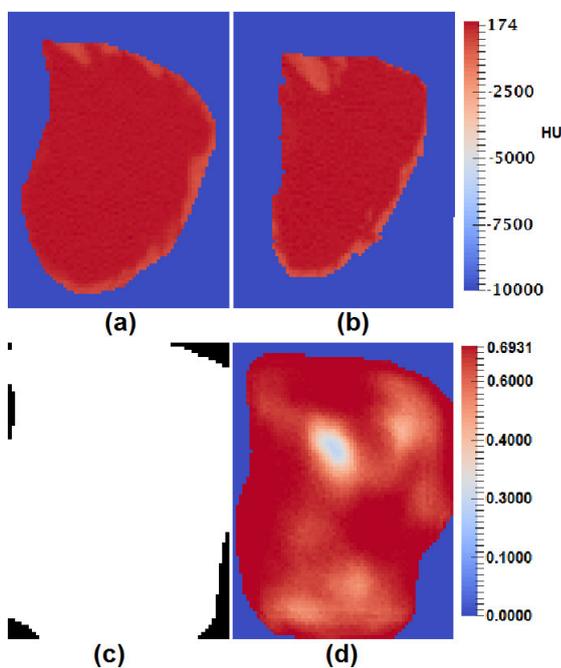
O percentual de mudança (P_m) para o Caso 42 na análise local do teste de hipótese mostrou que as mudanças regionais da densidade foram acentuadas, 45,91%. Um exemplo de quão acentuadas foram as mudanças pode ser visto na Figura 7.4. Entretanto, percebe-se que essas mudanças de densidade regionais não tiveram a presença de valores de densidade diferentes das que já existiam.

Da mesma forma, o Caso 3 apresentou um decréscimo do volume, mas as densidades perdidas foram em média semelhantes às densidades na lesão após a perda, pois a diferença

entre as médias *bootstrap* dos valores da lesão entre os instantes analisados não foi estatisticamente significativa pelo teste de hipótese.

O valor da b-divergência de Jensen relativa obtido para o Caso 3, 0,473 mostra que houve mudanças na densidade. Da mesma forma, o P_m também indica que quase 35% do volume da imagem inteira da lesão se modificou para a análise local do teste de hipótese. A Figura 5.10 mostra a lesão nos dois instantes analisados, o registro e as mudanças locais sofridas.

Figura 5.10 – Em (a), tem-se a lesão 1 e em (b), a lesão 2. Em (c), tem-se a imagem resultante da aplicação local da divergência. A imagem binária em (d) é o resultado da aplicação do teste de hipótese local.



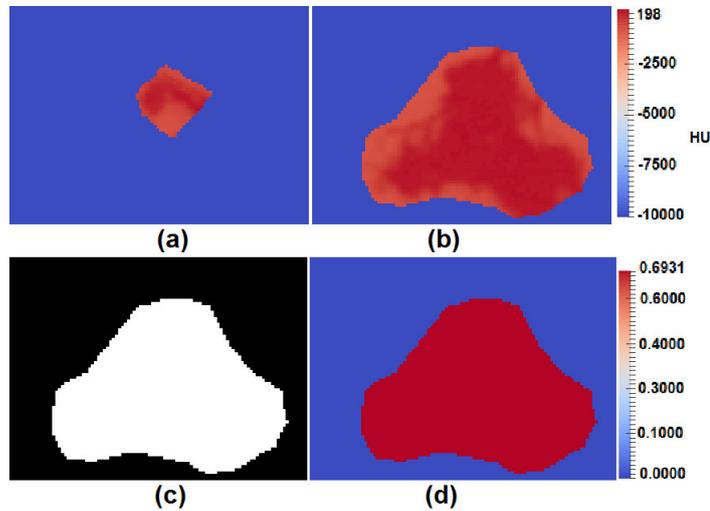
Fonte: Autor

Os resultados encontrados na Tabela 5.2 mostram que, apenas os Casos 4, 7, 9, 11, 12 e 15 foram estáveis em relação às suas densidades, segundo o teste de hipótese global com nível de significância de 5%. Desses, apenas os Casos 9 e 12 tiveram valores de b-divergência superiores aos mesmos casos considerados estáveis; e tiveram percentuais de mudança de 30,03% e 21,35%, respectivamente.

5.7.1 Estudo de casos

As seções seguintes exploram o uso de modelos que simulam uma lesão em transformação, nódulos benignos em acompanhamento e lesões em tratamento. O objetivo

Figura 5.11 – Em (a), tem-se a lesão 1 e em (b), a lesão 2. Em (c) tem-se a imagem resultante da aplicação local da divergência. A imagem binária em (d) é o resultado da aplicação do teste de hipótese local.



Fonte: Autor

é verificar a sensibilidade dos métodos propostos às mudanças de densidade. A avaliação será realizada de forma local e global; e ambos os resultados serão detalhados a seguir.

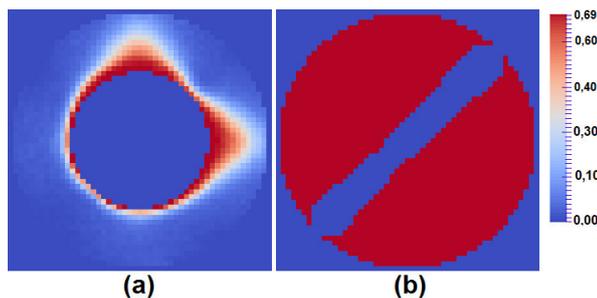
5.7.1.1 Estudo de casos com modelos de lesões

O método não teve o resultado esperado quando se analisa apenas o resultado do teste de hipótese. Ele não consegue detectar com precisão as esferas de mudança com a análise local. Essa imprecisão pode ser vista na Figura 5.12(b). Percebe-se, por essa imagem, que a esfera de maior raio foi dividida em duas regiões de mudança. Esse resultado está relacionado com o tamanho da vizinhança que define a amostra. Assim, todos os *voxels* nas proximidades das esferas de mudança foram detectados como mudança. Logo, extensas regiões foram identificadas como mudança pelo método do teste de hipótese, mas o método da divergência de Jensen local foi melhor no realce das mudanças para o estudo de caso com os modelos de lesões.

A divergência de Jensen local foi melhor na tarefa de realçar as mudanças de densidades. Na Figura 5.12(a), visualiza-se que as esferas de aumento e diminuição de 500 HU tiveram os maiores valores de divergência. As regiões em vermelho na figura indicam as regiões que mais se modificaram ao longo do tempo. Da mesma forma, as regiões próximas do branco tiveram divergência parcial e indicam que todas as esferas de mudanças foram detectadas. Portanto, se a imagem gerada pelo teste de hipótese não for precisa, ou seja, se

as regiões de mudança de densidade forem extensas, a divergência de Jensen pode auxiliar o entendimento das mudanças de densidade. Quando ambas as imagens geradas pelos métodos são analisadas em conjunto, as informações de mudança são melhor observadas e a análise global pode ser uma segunda avaliação para as mudanças.

Figura 5.12 – Resultado dos métodos baseados na incerteza para os modelos de lesões da Figura 4.11. Em (a), tem-se a imagem da divergência e em (b), a imagem binária de mudança gerada pelo teste de hipótese.



Fonte: Autor

A análise global dos modelos de lesões resultou na detecção de nenhuma diferença entre os mesmos. Na média, a divergência dos histogramas gerados pelo *bootstrap* foi igual a 0,0397, o que se pode considerar baixa. O mesmo ocorre para os histogramas dos modelos de lesões. Mesmo com as alterações e sem aplicação do *bootstrap*, a divergência teve um pequeno valor igual a 0,105882. Isso implica que o método não foi sensível às variações de densidade propostas para os modelos de lesões.

Nas seções seguintes, alguns casos serão apresentados para valores de VVR próximo de zero e valores significantes de P_m , para cada método.

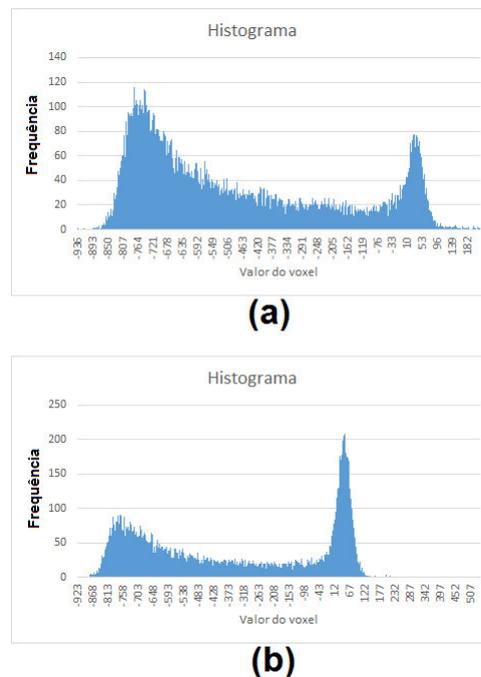
5.7.1.2 Estudo de casos com a base pública de lesões

Nesta seção, é apresentado o comportamento dos métodos na tarefa de detectar mudanças para os dois casos que apresentaram os menores valores de VVR . Esses casos mostram a importância dos métodos propostos na detecção de mudanças, pois tais mudanças podem não são perceptíveis pelo especialista durante a sua inspeção visual ou por meio de ferramentas computacionais, que determinam a variação volumétrica da lesão em diferentes instantes.

O Caso 10 apresentou o menor valor de VVR , -1,61%, o que sugere uma pequena redução do volume da lesão. O percentual de mudança de 14,93% encontrado pela aplicação do teste de hipótese local aponta que a lesão sofreu mudanças de densidade em parte da

sua extensão. A b-divergência global de Jensen apresentou seu valor máximo relativo, 1,0, confirmando que houve modificações de densidade, apesar da pequena variação volumétrica. Nas imagens da Figura 5.14, pode-se visualizar as modificações de densidade. Ao mesmo tempo, os histogramas dos instantes t_a e t_b mostram que a lesão ficou mais densa, conforme a Figura 5.13, pois houve um aumento na incidência de *voxels* com intensidades de -40 a 140 HU.

Figura 5.13 – Histogramas do Caso 10 para a lesão em t_a (a) e em t_b (b).

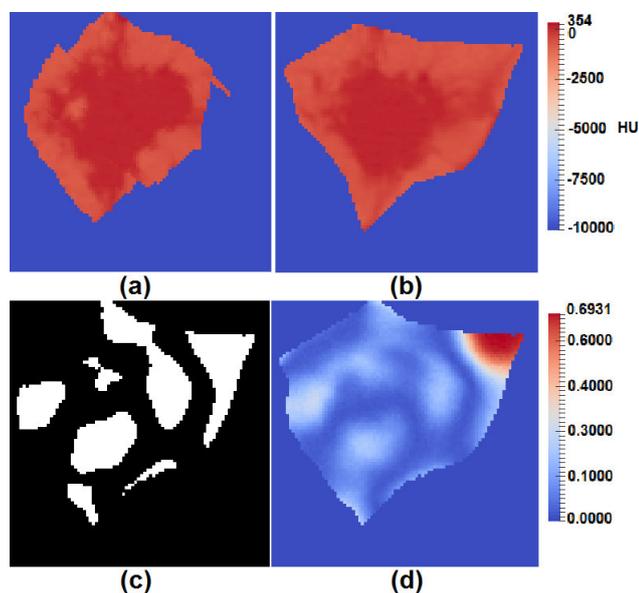


Fonte: Autor

A partir das imagens da Figura 5.14(c) para a local divergência e Figura 5.14(d) para o teste de hipótese local, pode-se visualizar que as modificações foram mais intensas nas bordas da lesão da Figura 5.14(a) e (b). As regiões em vermelho na Figura 5.14(c) indicam as regiões que apresentam maior divergência e são realçadas em branco na Figura 5.14(d).

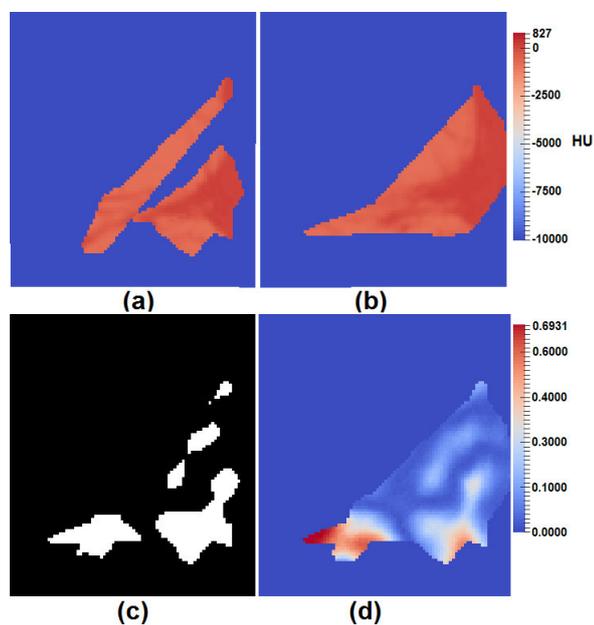
O segundo menor valor de VVR é visto no Caso 28 da Tabela 5.1, com VVR de -3,12%. Esta pequena redução é acompanhada de um percentual de mudança de 9,64% em seu volume, valor encontrado através do método local do teste de hipótese. Seu valor de b-divergência relativa foi máxima, isto é, 1,0, o que revela mudanças nas densidades da lesão entre os instantes analisados. A Figura 5.15 possui as imagens resultantes da aplicação dos métodos locais da divergência e teste de hipótese. É ilustrado na Figura 5.15(c) que a divergência é maior na região espicular da lesão; e através da imagem na Figura 5.15(d) do teste de hipótese local, visualiza-se que a lesão é mais semelhante internamente.

Figura 5.14 – Em (a), tem-se a lesão 1 e em (b), a lesão 2. Em (c), tem-se a imagem resultante da aplicação local da divergência. A imagem binária em (d) é o resultado da aplicação do teste de hipótese local.



Fonte: Autor

Figura 5.15 – Em (a), tem-se a lesão 1 e em (b), a lesão 2. Em (c), tem-se a imagem resultante da aplicação local da divergência. A imagem binária em (d) é o resultado da aplicação do teste de hipótese local.

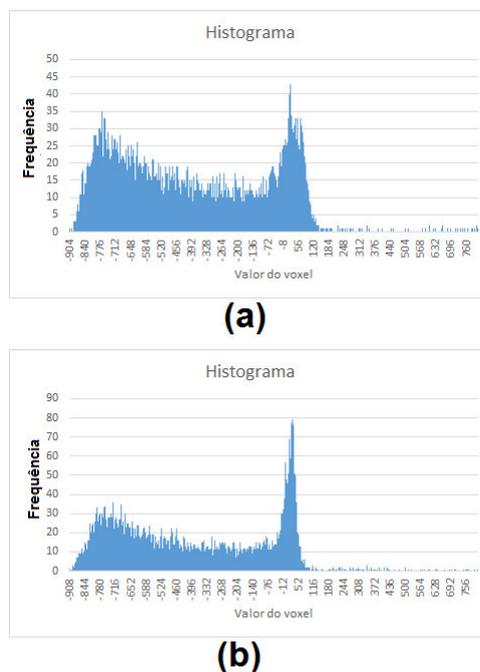


Fonte: Autor

As intensidades no intervalo de -70 a 120 HU tiveram um aumento conforme pode ser visto na Figura 5.16, para os histogramas da lesão nos dois instantes analisados.

Em contraste aos casos apresentados de menor variação volumétrica da lesão, o Caso 42 apresentou o maior valor VVR , o que poderia revelar a existência de muitas

Figura 5.16 – Histogramas do Caso 28 para a lesão em t_a (a) e em t_b (b).



Fonte: Autor

mudanças nos valores de densidade para a lesão. Entretanto, a partir dos histogramas da lesão nos dois instantes analisados na Figura 5.17, percebe-se que a lesão se tornou mais densa. Verifica-se, na figura, que em t_1 , o mínimo e o máximo de densidade foram -901 e 174 HU, respectivamente. Para a lesão em t_2 , o mínimo e o máximo foram -846 e 198 HU, respectivamente.

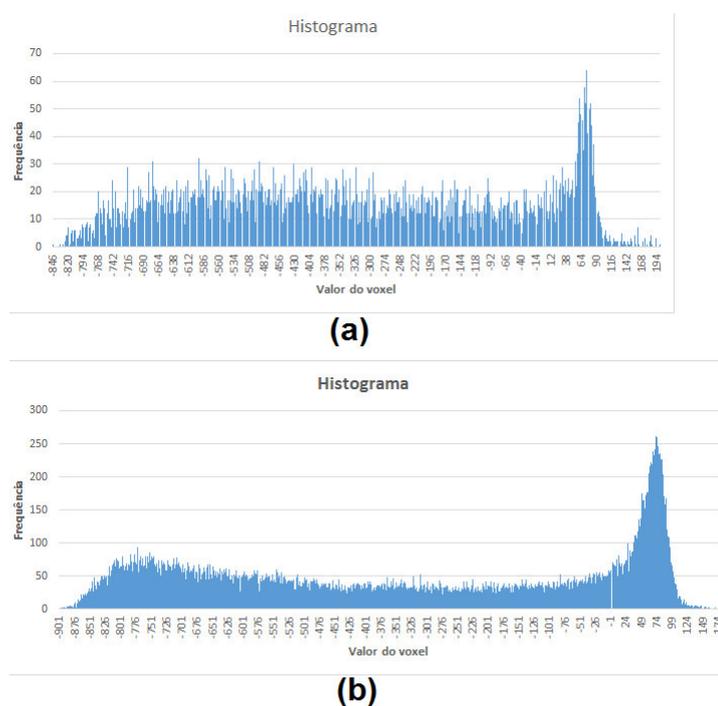
O aumento de densidade para o Caso 42 é pequeno comparado a grande variação volumétrica, VVR igual a 521,3%, logo, isso explica o grande valor da divergência dos histogramas dos valores dos *voxels*, mas a b-divergência não acentuada da média *bootstrap* concorda com o resultado global do teste de hipótese de que não há diferença significativa das densidades da lesão.

5.7.1.3 Estudo de casos com a base privada de lesões

Esta base contém apenas nódulos pulmonares benignos e foram avaliados em um período de 3 anos para verificação da sua estabilidade.

Os Casos 4 e 12 da Tabela 5.2 são os casos de menor VVR , 1,351% e 1,724%, respectivamente, para essa base. Essas lesões tiveram pequena variação volumétrica e foram consideradas estáveis pelos especialistas.

Figura 5.17 – Histogramas para a lesão do Caso 42 nos dois instantes analisados.



Fonte: Autor

A confirmação da estabilidade para os Casos 4 e 12 foi obtida através do teste de hipótese global. Entretanto, como o próprio valor VVR aponta, as pequenas variações do volume estão acompanhadas de mudanças percentuais das densidades de seus volumes, conforme os respectivos valores de 5,76% e 21,35%, obtidos para os Casos 4 e 12.

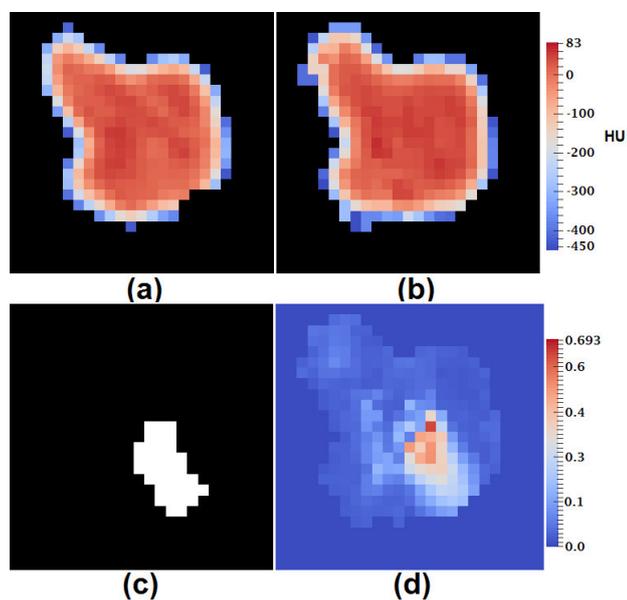
As respectivas imagens dos Casos 4 e 12 resultantes para a local b-divergência estão nas Figuras 5.18 e 5.19. Coincidentemente, ambos os casos apresentaram a imagem de divergência apontando modificações centrais às lesões, fato confirmado também pelas imagens do teste de hipótese local, Figuras 5.18(d) e 5.19(d).

Para o Caso 4, os histogramas da Figura 5.20, no par de instantes analisados, mostram que a lesão ficou mais densa, devido ao valor máximo ser maior. Ao mesmo tempo, houve uma redução da presença de valores de alta intensidade, pois ocorreu uma diminuição da frequência.

Os histogramas para o Caso 12 na Figura 5.21 revelam que ocorreu uma redução do valor máximo de densidade, entretanto, as frequências das intensidades mais altas foram maiores.

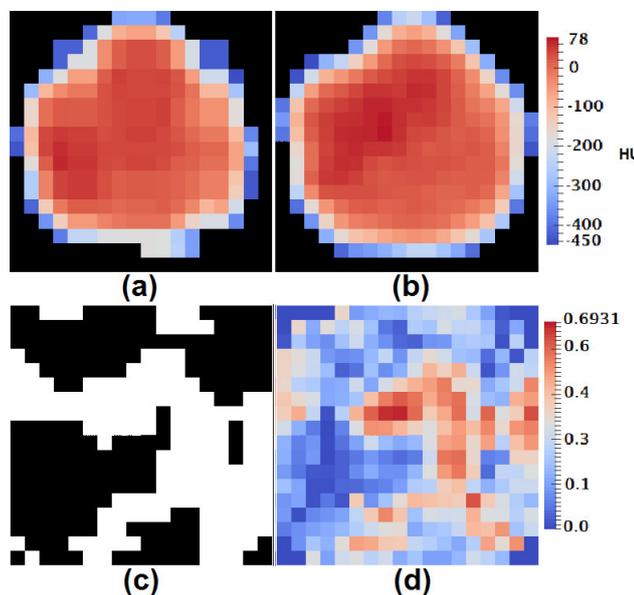
O caso de maior VVR foi o Caso 14, -60,00%, significando em uma redução volumétrica. Como consequência, 35,43% do seu volume se modificou, resultado encontrado através da aplicação local do teste de hipótese. A b-divergência relativa atingiu o seu

Figura 5.18 – Em (a), tem-se a lesão 1 e em (b), a lesão 2. Em (c), tem-se a imagem resultante da aplicação local da divergência. A imagem binária em (d) é o resultado da aplicação do teste de hipótese local.



Fonte: Autor

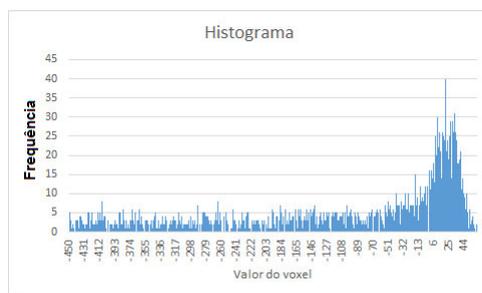
Figura 5.19 – Em (a), tem-se a lesão 1 e em (b), a lesão 2. Em (c), tem-se a imagem resultante da aplicação local da divergência. A imagem binária em (d) é o resultado da aplicação do teste de hipótese local.



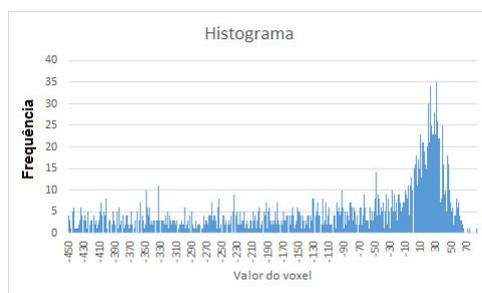
Fonte: Autor

valor máximo, 1,0, indicando que o nódulo teve grande variação em suas densidades. A Figura 5.22 possui as imagens analisadas para esse caso e as imagens da local divergência e do teste de hipótese. Devido a grande variação no volume para o Caso 14, os histogramas da Figura 5.23 mudaram em forma, escala e frequência.

Figura 5.20 – Histogramas para a lesão do Caso 4 nos dois instantes analisados.



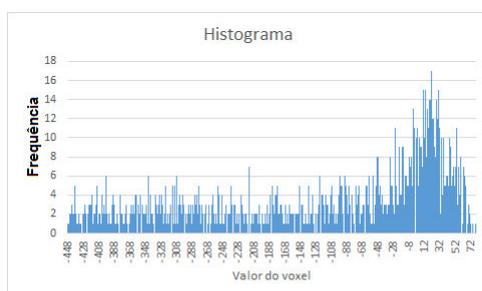
(a)



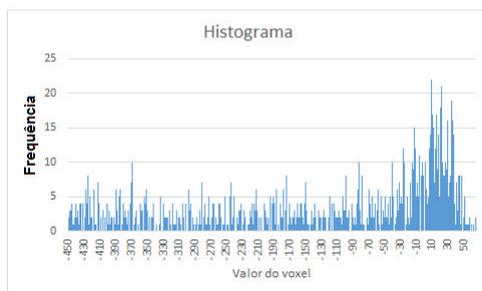
(b)

Fonte: Autor

Figura 5.21 – Histogramas para a lesão do Caso 12 nos dois instantes analisados.



(a)



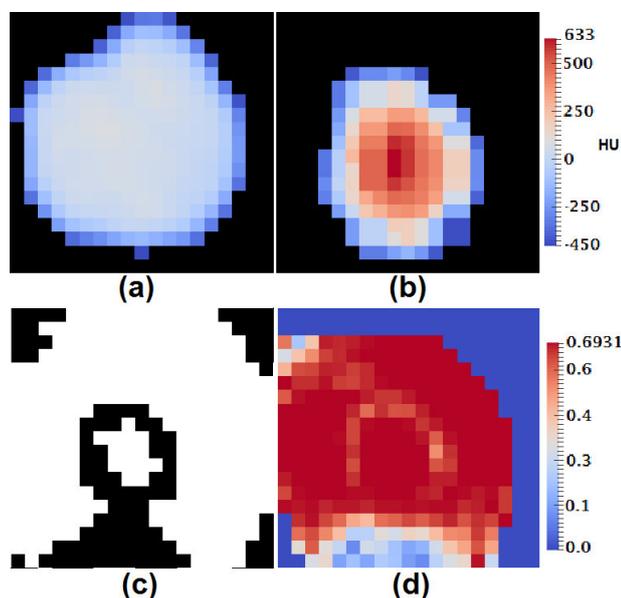
(b)

Fonte: Autor

5.7.2 Classificação de lesões quanto ao diagnóstico

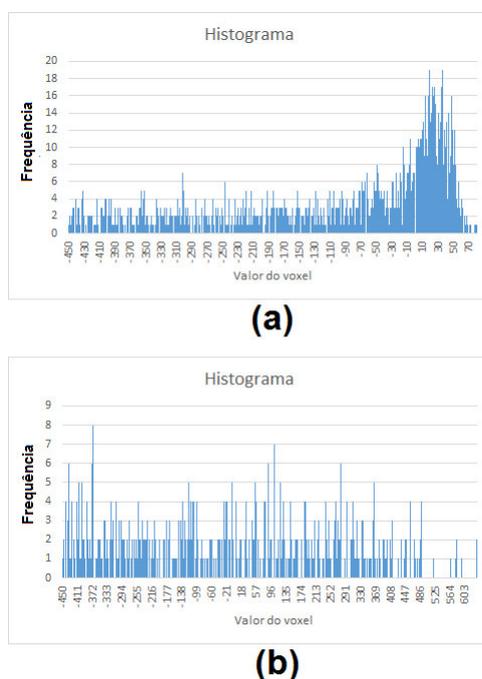
O objetivo desta análise é verificar se as mudanças de densidade encontradas pelo teste de hipótese corroboram com o diagnóstico. Da mesma forma, pode-se verificar

Figura 5.22 – Em (a), tem-se a lesão 1 e em (b), a lesão 2. Em (c), tem-se a imagem resultante da aplicação local da divergência. A imagem binária em (d) é o resultado da aplicação do teste de hipótese local.



Fonte: Autor

Figura 5.23 – Histogramas para a lesão do Caso 14 nos dois instantes analisados.

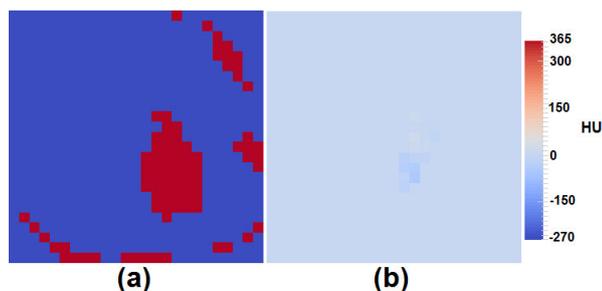


Fonte: Autor

se o modelo de classificação encontrado com as medidas de textura é suficiente para prever o diagnóstico da lesão pulmonar, independente da mesma se encontrar ou não em tratamento. O modelo é gerado das características de textura extraídas apenas das regiões onde ocorreram as mudanças. Um exemplo dessas regiões de mudanças é visto na

Figura 5.24, para o resultado do teste de hipótese. Na figura, tem-se em (a) a imagem em binário com as regiões em vermelho indicando as mudanças e a imagem em (b) é a diferença das imagens da lesão para essas regiões de mudança.

Figura 5.24 – Exemplo das regiões em que serão extraídas as características de textura para classificação de lesões em maligno e benigno.



Fonte: Autor

As características utilizadas na tarefa de classificação foram definidas na Seção 3.5. O total de características obtidas foram 36. Em razão do grande número de características, realizou-se a seleção de características com o método *Random Forest and RM-bound MVS* (CHEN; LIN, 2006). Da seleção, duas características do *RL* foram selecionadas: SRHGLE e LGRE. As outras duas restantes foram das medidas estatísticas: a máxima diferença (*MAXD*) e a média da diferença dos valores negativos (*MND*). Logo, um total de 4 características foram utilizadas para o método proposto.

Devido a pequena amostra para a tarefa de classificação, foi utilizada a técnica de amostragem *leave-one-group-out* (KOHAVI, 1995) como forma de avaliar o modelo gerado na classificação. O número de grupos utilizado foi 10.

O classificador utilizado foi o MVS com *kernel RBF* (Seção 3.6.1) e os parâmetros de classificação, C e g , estão na Tabela 5.3. A classificação também foi realizada com os parâmetros padrões, isto é, $C = 1$ e $g = 1/4$, onde 4 é o número de características.

O resultado da classificação quanto ao diagnóstico para amostra contendo as 63 instâncias, das quais 17 instâncias são casos benignos (base privada) e 46 instâncias são casos malignos (base pública), estão na Tabela 5.3.

5.8 Discussão

O Caso 10 para a base pública teve menor *VVR* e a divergência dos seus histogramas do par de instantes analisados foi pequena. Entretanto, a divergência das amostras *bootstrap*

Tabela 5.3 – Métricas de avaliação da classificação utilizando MVS associadas a cada um dos parâmetros do modelo.

Avaliação do teste de hipótese								
Parâmetros do MVS			Métricas de avaliação					
C	g	sens(%)	esp(%)	ac(%)	vp	vn	fp	fn
1	1/4	95,65	100,00	96,83	44	17	0	2
2	0,0078125	95,65	100,00	96,83	44	17	0	2

Fonte: Autor

para média foi máxima. Isso indica que a lesão foi detectada como proveniente de duas populações diferentes, para os dois instantes analisados. O mesmo ocorreu para o Caso 28.

As pequenas diferenças obtidas para os Casos 10 e 28 são mais facilmente visíveis pelas imagens das lesões e as imagens dos seus histogramas mostram que ocorreram mudanças nas suas densidades, mas essas mudanças não são facilmente perceptíveis a olho nu.

A base privada teve um comportamento diferente para a mesma situação de pequena variação volumétrica. O Caso 4 apresentou um pequeno valor de P_m , ocorrendo o mesmo para as medidas de divergência e b-divergência. Portanto, o referido caso foi considerada estável no intervalo de tempo analisado. O mesmo ocorreu para o Caso 12.

A classificação teve acurácia de 96,83%, com 2 falsos negativos, independentes da estimação dos parâmetros de classificação. Isso indica que as mudanças de densidade capturadas pela medida de textura dos casos malignos podem ter alguma semelhança com as dos casos benignos.

5.9 Conclusão

O presente conjunto de métodos mostra que é possível realçar valores de densidade para uma lesão pulmonar obtida em vários exames. Esse realce enfatiza as regiões da lesão que tiveram mudanças ao longo do tempo.

O conjunto de métodos propostos empregou técnicas e outros métodos de inferência estatística. O objetivo é fornecer informações de mudanças de densidade ao especialista e facilitar as ações apropriadas em cada caso. Como se pode observar dos resultados, além de mostrar a localização das mudanças da lesão, uma quantificação foi dada para essas mudanças.

A análise foi realizada sobre lesões malignas que estavam em tratamento e sobre nódulos pulmonares indeterminados. Como o objetivo é detectar mudanças de densidade, os resultados mostram que se pode utilizar os métodos propostos independente da lesão estar ou não em tratamento.

As técnicas e os métodos de incerteza propostos mostraram-se promissores para a tarefa de detecção de mudanças de densidade. Duas análises foram realizadas para as técnicas de divergência e teste de hipótese, uma local e outra global. A partir da análise local, foi possível gerar uma visualização das modificações da lesão ao longo do tempo. Além da visualização, a partir do teste de hipótese local, foi possível gerar valores quantitativos das mudanças de densidade.

A diferença encontrada pelos métodos propostos para avaliação da progressão da lesão ao longo do tempo pode sugerir ao especialista um indicativo quantitativo das mudanças de densidade. Além disso, em conjunto com as imagens das mudanças, o especialista pode ter uma segunda opinião sobre as mudanças da lesão pulmonar.

Com o intuito de explorar melhor os resultados encontrados, foram analisados em detalhes situações em que as lesões sofreram pequenas variações volumétricas. Os métodos propostos foram capazes de indicar, nesses casos, que as mesmas tiveram variações de densidade. Isso sugere que existem modificações em seus tecidos ao longo do tempo, independente da lesão pulmonar estar em tratamento ou não. Logo, exploram-se situações cotidianas de análise que podem ser imperceptíveis para o especialista. Dessa forma, os métodos propostos surgem como ferramentas para ampliar o leque de análise.

Outra situação importante analisada, foram os casos que tiveram grandes valores de *VVR*. Nesse casos, os resultados apontam que as modificações de densidade não apresentam desvios acentuados, pois em média, estas se encontram muito próximas do valor médio das densidades encontradas em um primeiro instante.

Os métodos propostos foram sensíveis às variações nos valores de densidade da lesão e foram capazes de realçar e mensurar as regiões que possuíram mudanças ao longo do tempo. Logo, com base nos resultados encontrados, pode-se dizer que são bons métodos para análise de mudanças em lesões em 3D. Entretanto, ainda existe a necessidade da realização de novos estudos com outras bases de imagens para confirmá-las como métodos complementares à medida de variação relativa do volume.

6 Método baseado em análise de componentes principais

No presente capítulo, serão descritos os detalhes matemáticos do funcionamento da análise de componentes principais e como esta é utilizada na detecção das diferenças das imagens da lesão ao longo do tempo. Logo, todas as etapas necessárias para a detecção das diferenças serão descritas.

6.1 Análise de Componentes Principais

A análise de componentes principais (ACP) é uma técnica que visa detectar as direções de maior variância de um conjunto de dados. Quando se aplica ACP aos dados no R^2 , como um par de valores de *pixel* ou *voxel* em uma mesma posição de duas imagens, a segunda componente principal identifica a segunda direção de maior variância dos dados. Essa direção é chamada de componente de erro, pois seus valores estão distantes da primeira componente principal.

A ACP é utilizada para identificar a componente de erro, com o objetivo de realçar as diferenças do par de imagens da lesão pulmonar em diferentes instantes de aquisição. Para esse propósito, deve-se compreender todos os passos da aplicação ds ACP.

A partir de dados multidimensionais, isto é, de pontos ou vetores no R^n , é possível obter as direções ou componentes principais dos dados. Cada ponto no R^n descreve um conjunto de n características colecionadas de um objeto. O conjunto de pontos, cada um originário dos vários objetos, forma uma matriz de observações (LAY, 2002).

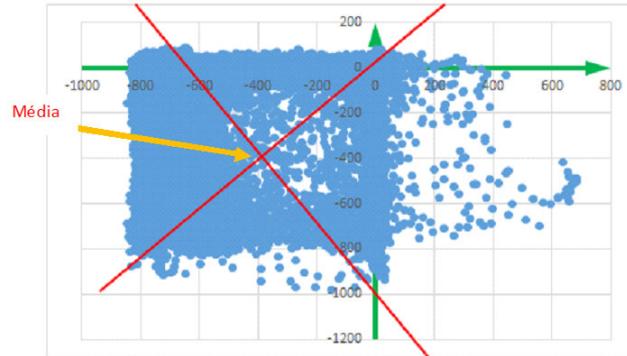
Para encontrar as componentes principais dos dados, é necessário encontrar a média e covariância desses pontos. Esses elementos podem ser vistos na Figura 6.1. Logo, dada uma matriz de observações X com f características e N pontos ou amostras, a média M dos pontos é dada por

$$M = \frac{1}{N} (X_1 + \dots + X_N) \quad (57)$$

A partir da média M dos pontos, pode-se centralizar (transladar) os N pontos dados para o ponto médio encontrado (nova origem), resultando em uma nova matriz de observações B com média nula. Utiliza-se, para isso, a seguinte transformação:

$$\hat{X}_k = X_k - M, k = 1, \dots, N \quad (58)$$

Figura 6.1 – Conjunto de pontos no R^2 onde a média é extraída.



Fonte: Autor

A matriz de covariância é a matriz S $f \times f$ definida por

$$S = \frac{1}{N-1} B B^T \quad (59)$$

A matriz S contém alguns elementos importantes. O primeiro deles é o elemento S_{jj} da diagonal principal. Cada elemento da diagonal é a variância para a característica (componente) j dos f valores componentes do ponto. Outro elemento importante é o traço de S , que é a variância total dos dados. Os elementos S_{ij} , com $i \neq j$, são as covariâncias entre as características i e j dos f valores componentes do ponto.

A partir da média e da matriz de covariância dos pontos, é possível obter uma matriz ortogonal P que transforme B em Y , $Y = P^T B$, com as seguintes propriedades: As componentes de Y não estão correlacionadas e estão ordenadas em ordem crescente de variância.

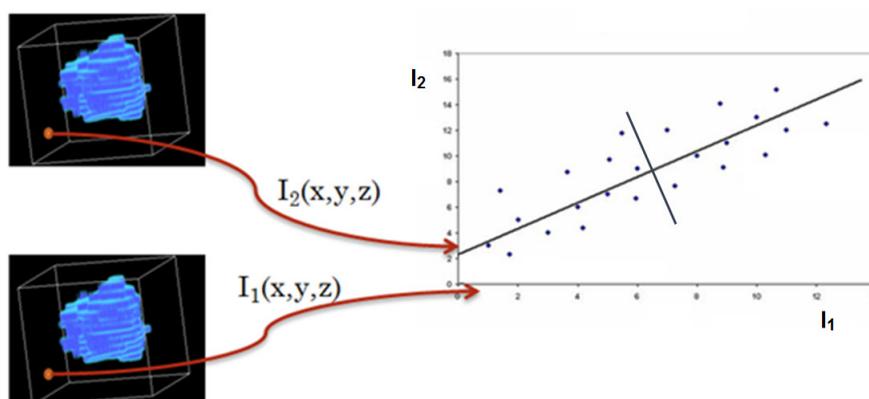
A matriz P se origina do processo de decomposição em valores singulares (LAY, 2002) da matriz $A = (1/\sqrt{N-1}) B^T$. O quadrado dos valores singulares de A são os autovalores da matriz S ; e os autovetores a direita de A são as componentes principais que formam a matriz P . Portanto, as colunas da matriz P contêm as direções de projeção de B , isto é, definem o novo sistema de coordenadas para B , onde a ordem de cada componente decresce em variância.

6.2 Identificação das mudanças de densidade

Para a detecção de mudanças de densidade da lesão pulmonar ao longo do tempo utilizando ACP, foi definido que cada ponto teria dimensão 2. A primeira componente

desse ponto no R^2 seria o valor do *voxel* em um instante inicial t_1 ; e a segunda componente, o valor do *voxel* em um instante futuro t_2 . A matriz de observações é constituída por cada um dos valores assumidos pelos *voxels* do volume. A Figura 6.2 ilustra o processo de formação dos pontos para o par de lesões pulmonares.

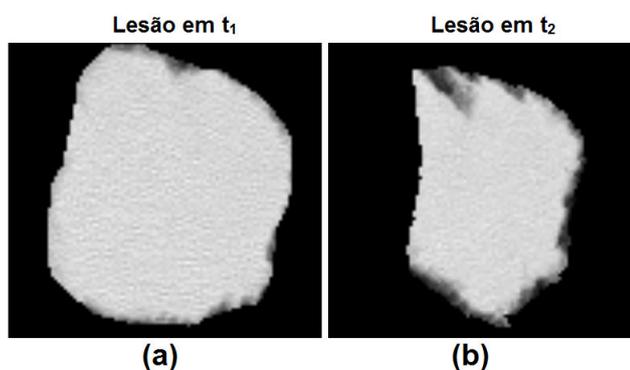
Figura 6.2 – Conjunto de pontos no R^2 a partir do qual serão extraídas as componentes principais. O eixo I_1 contém os valores de *voxel* da imagem no instante t_1 e o eixo I_2 os valores de *voxel* da imagem no instante t_2 .



Fonte: Autor

Com o objetivo de detectar pequenas mudanças, o processo de detecção das mudanças usando ACP ocorre com a dupla aplicação do método. Assim, a primeira aplicação detecta mudanças bruscas de densidade, como se pode visualizar na Figura 6.4 para o par de imagens da Figura 6.3. Na figura, pode-se verificar que as mudanças bruscas são as mudanças fundo-lesão ou lesão-fundo. Como consequência, pequenas mudanças de densidade não são realçadas nessa aplicação, necessitando de uma segunda.

Figura 6.3 – Exemplo de duas imagens da lesão em análise nos instantes t_1 e t_2 , respectivamente.



Fonte: Autor

A segunda aplicação detecta mudanças de menor escala de densidade. A diferença nessa etapa está no processo de formação da matriz de observações. Para a sua formação, a

Figura 6.4 – Imagem resultante da primeira aplicação da ACP para as imagens da Figura 6.3.

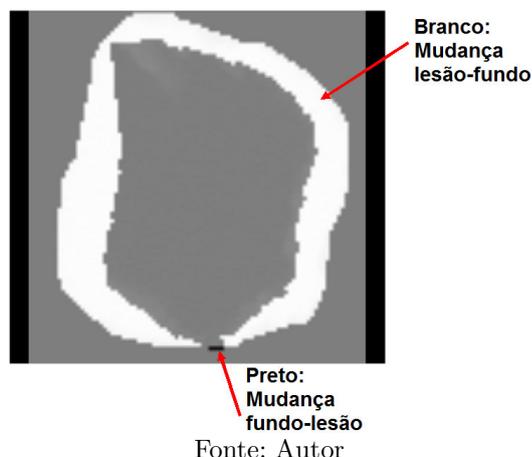
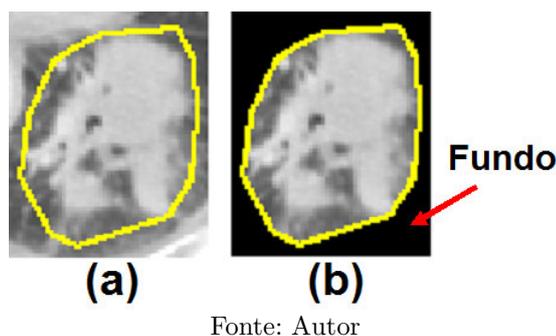


Figura 6.5 – (a) - Exemplo da lesão pulmonar e seu contorno e em (b), a imagem em (a) com fundo inserido.

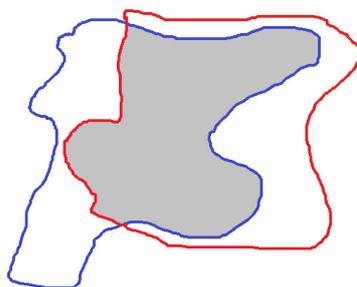


matriz de observações compreende apenas os *voxels* da interseção das imagens (irregulares) definidas pelos contornos da lesão nos pares de instantes analisados. Na Figura 6.6, pode-se visualizar essa região, onde são visíveis um contorno em azul do instante t_1 e em vermelho para um instante t_2 . Para essa região em comum, pode-se visualizar as mudanças de menor escala de densidade realçadas pela segunda aplicação da ACP na Figura 6.7. Depois do realce, o próximo passo é detectar as regiões com as maiores mudanças de densidade.

A detecção das mudanças de densidade da lesão pulmonar ocorre pela análise da distância ao longo da direção da segunda componente principal. Essa análise é realizada em cada uma das matrizes de observações obtidas na primeira e segunda aplicação. Quanto mais distante um ponto estiver da primeira componente ao longo da segunda, maior é a diferença observada para o par de valores de *voxel*.

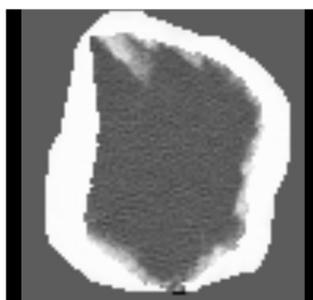
Após a aplicação da ACP, o realce das mudanças é realizado através da aplicação do algoritmo de agrupamento *k-means* com 4 grupos. O agrupamento separa os pontos do R^2 em duas regiões nas extremidades da segunda componente principal e nas regiões

Figura 6.6 – A lesão com seus contornos em diferentes instantes. A interseção dos contornos (região em cinza) é analisada pela ACP.



Fonte: Autor

Figura 6.7 – Resultado final da aplicação da ACP.

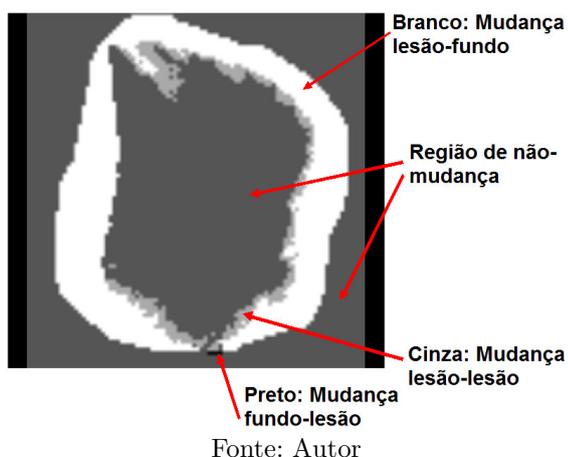


Fonte: Autor

de pequenas mudanças. Como pode ser percebido, as regiões de pequenas mudanças compreendem os pontos agrupados próximo ao centro da segunda componente principal. Como referência para análise, o grupo de não-mudança é aquele pertencente ao fundo da imagem, pois não varia ao longo do tempo. Portanto, todos os grupos restantes são rotulados como mudança. Os resultados são então compostos em uma única imagem, conforme a Figura 6.8. Nessa figura, é possível visualizar as regiões que sofreram mudança e não-mudança.

A partir das regiões classificadas em mudança e não-mudança, foram extraídas algumas informações de textura apenas da região de mudança de densidade. O objetivo é verificar se essas regiões possuem alguma influência sobre o diagnóstico da lesão quanto a sua natureza. Para isso, realizou-se a diferença algébrica das imagens apenas nas regiões classificadas como mudança e usaram-se descritores clássicos de textura, como a matriz de cumprimento de cinza RL e as características de Haralick 3D.

Figura 6.8 – Resultado final da ACP após o agrupamento.



6.3 Resultados

Para atingir os resultados esperados na identificação das mudanças de densidade, as seguintes etapas iniciais foram definidas conforme as respectivas seções: base de lesões pulmonares (Seção 4.7.2), modelos de lesões (Seção 4.7.1) e registro das imagens da lesão (Seção 4.7.3). Em seguida, apresentam-se todos os resultados quanto a sensibilidade às mudanças encontradas para as bases pública e privada de lesões pulmonares. Além disso, realiza-se um estudo de casos para as lesões que apresentaram pequena variação volumétrica. Na última etapa, mostra-se o quanto as características extraídas das regiões de mudança são indicadoras do diagnóstico da lesão pulmonar.

6.3.1 Estudo de casos

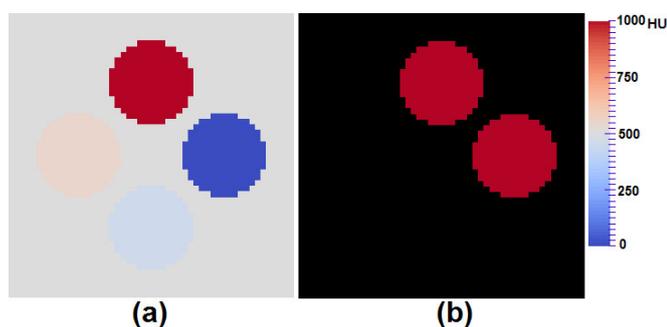
Os resultados são apresentados inicialmente para os modelos de lesões gerados. A partir dos modelos, onde mudanças pontuais foram determinadas, consegue-se verificar a sensibilidade às mudanças para o método proposto.

6.3.1.1 Estudo de casos com modelos de lesões

O método PCA conseguiu detectar as duas principais mudanças para os modelos de lesões. As principais mudanças são aquelas de maior variação de densidade ao longo do tempo. Para os modelos citados, as maiores mudanças correspondem ao aumento e a diminuição de 500 HU do valor original de 1000 HU. Como resultado da aplicação do PCA,

obteve-se a imagem PCA da Figura 6.9(a). Na figura, percebe-se que todas as quatro esferas de mudança foram destacadas do fundo. Entretanto, a última etapa conseguiu detectar apenas as esferas correspondentes às grandes mudanças de densidade, conforme a imagem da Figura 6.9(b), e produziu o P_m igual a 2,7%. Portanto, para os modelos de lesões propostos, o método PCA não foi sensível às pequenas variações de densidade, mas consegue realçar as mudanças de densidade de maior escala de variação.

Figura 6.9 – Resultado do método ACP para os modelos de lesões da Figura 4.11. Em (a), tem-se a imagem do ACP e em (b), a imagem binária de mudança.



Fonte: Autor

A seguir, mostram-se alguns casos da base pública e privada com as menores variações volumétricas relativas (VVR). É interessante observar o comportamento das mudanças da lesão pulmonar quando as variações volumétricas são pequenas. O objetivo é analisar se as mudanças de densidade ocorridas na lesão pulmonar são algum indicativo da natureza do seu diagnóstico. Da mesma forma, uma pequena variação volumétrica durante um intervalo de tempo de acompanhamento pode não ser suficiente para diagnosticar a lesão benigna em acompanhamento ou comprovar que a lesão maligna em tratamento continua com o caráter maligno ao longo do tempo.

6.3.1.2 Estudo de casos com a base pública de lesões

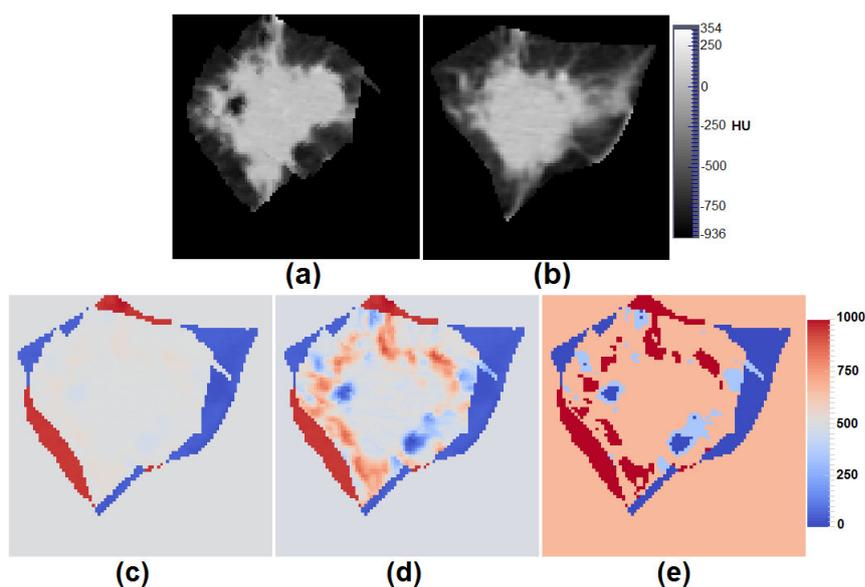
O método ACP para detecção de mudanças foi aplicado para a base pública e privada de lesões pulmonares. Para ambas as bases, as regiões onde ocorreram intensas mudanças de densidade foram evidenciadas e os percentuais de mudança dos *voxels* das lesões foram obtidos, conforme as Tabelas 6.1 e 6.2. Como já evidenciado na Seção 3.3.1, cada base de lesões pulmonares possui situações distintas de mudanças. A base pública é um conjunto de lesões pulmonares que estão em tratamento e a privada é composta apenas por nódulos pulmonares que estavam em acompanhamento. Portanto, independente da

lesão está em tratamento ou não, o método proposto realça as mudanças de densidades que ocorreram ao longo do tempo e pode ser utilizado em casos mais difíceis de monitoramento, por exemplo, em casos de pequena variação volumétrica.

Os Casos 10 e 28 apresentam as menores variações volumétricas relativas (VVR), -1,61 e -3,12, respectivamente. Na Tabela 6.1, pode-se visualizar tanto os valores de VVR quanto os valores de P_m encontrados sobre a base. Os valores de P_m foram 15,32 e 11,05, para os Casos 10 e 28, respectivamente. Os resultados indicam que a lesão do Caso 10 teve maior variação dos seus valores de densidade que o Caso 28, mesmo que ela tenha sofrido uma redução volumétrica menor em comparação com o Caso 28.

Os resultados de cada uma das etapas da metodologia de detecção das mudanças de densidade através da aplicação da ACP para o Caso 10 podem ser vistos na Figura 6.10. Na Figura 6.10(a) e (b), pode-se visualizar a lesão no par de instantes em análise. Na Figura 6.10(c), pode-se visualizar o resultado da primeira aplicação da ACP, onde as regiões de mudança lesão-fundo foram detectadas. A imagem da Figura 6.10(d) possui o resultado da segunda aplicação da ACP seguida da primeira. A imagem resultante do agrupamento pode ser vista na Figura 6.10(e), onde apenas o grupo que corresponde ao fundo é o grupo de não-mudança e os demais grupos são as componentes de mudança.

Figura 6.10 – Resultado final da aplicação da ACP para o Caso 10.



Os resultados para o Caso 28 podem ser vistos na Figura 6.11. Na Figura 6.11(a) e (b), pode-se visualizar a lesão no par de instantes em análise. Na Figura 6.11(c), pode-se visualizar o resultado da primeira aplicação do ACP, onde as regiões de mudança lesão-

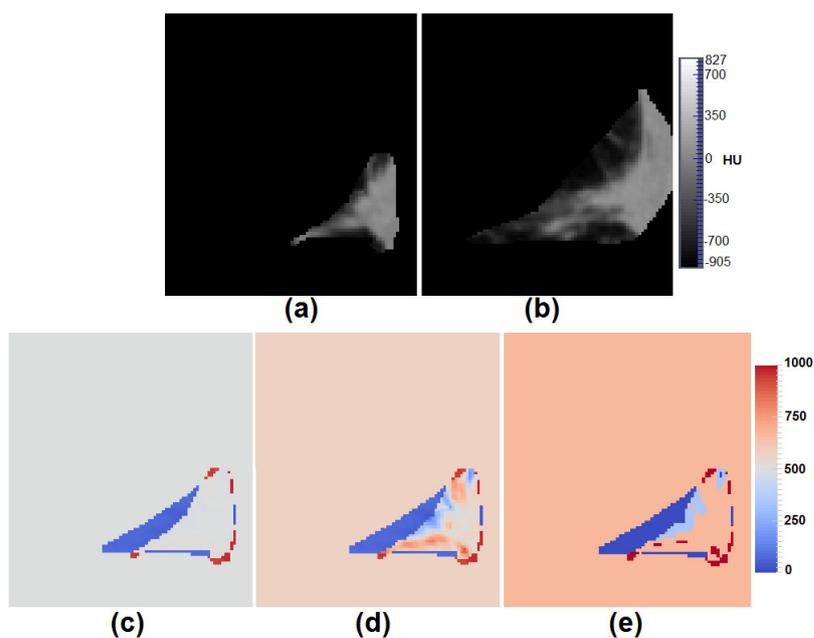
Tabela 6.1 – Conjunto de valores de percentuais de mudança obtidos do método ACP para cada caso da base pública de lesões.

Caso	P_m (%)	
	ACP	VVR
1	22,28	-48,30
2	31,90	-70,10
3	22,85	-42,17
4	8,26	133,60
5	9,28	-29,09
6	21,42	257,66
7	20,00	281,26
8	17,21	6,60
9	18,24	22,48
10	15,31	-1,61
11	18,93	-19,67
12	23,71	121,24
13	15,17	75,19
14	18,75	185,52
15	17,93	162,29
16	14,35	8,86
17	3,76	-17,94
18	18,69	275,08
19	21,30	-34,20
20	24,78	-13,72
21	9,28	-45,58
22	21,24	55,84
23	17,16	32,52
24	11,95	-4,77
25	12,88	11,68
26	12,51	-7,74
27	14,59	17,27
28	11,05	-3,12
29	10,94	-17,39
30	17,98	17,29
31	8,74	-5,29
32	5,65	4,55
33	6,70	10,38
34	10,50	-11,22
35	16,25	30,35
36	18,14	102,14
37	12,42	46,83
38	18,01	127,69
39	12,65	55,08
40	22,16	26,74
41	17,94	95,76
42	32,03	521,34
43	8,39	-17,12
44	27,08	26,81
45	19,72	-46,43
46	12,57	-57,75

Fonte: Autor

fundo foram detectadas. A imagem da Figura 6.11(d) mostra o resultado da segunda aplicação da ACP composta com a primeira. A imagem resultante do agrupamento pode ser vista na Figura 6.11(e), onde apenas o grupo que corresponde ao fundo é o grupo de não-mudança e os demais grupos são as componentes de mudança.

Figura 6.11 – Resultado final da aplicação da ACP para o Caso 28.



6.3.1.3 Estudo de casos com a base privada de lesões

Os casos que possuem os menores VVR foram 4 e 12, com 1,35 e 1,72, respectivamente. Os valores P_m para esses mesmos casos foram 8,75 e 22,59, conforme a Tabela 6.2. Esses resultados de P_m para estes casos confirmam que não houve modificações significativas de densidade para o Caso 4 e que foi menor quando comparado com o Caso 12.

Os resultados para os Casos 4 e 12 podem ser vistos nas Figuras 6.12 e 6.13. Nessas figuras, em 6.12(a),(b) e 6.13(a),(b), pode-se visualizar a lesão no par de instantes em análise. Em 6.12(c) e 6.13(c), pode-se visualizar o resultado da primeira aplicação do ACP, em que as regiões de mudança lesão-fundo foram detectadas. As imagens em 6.12(c) e 6.13(c) mostram os resultados da segunda aplicação da ACP combinada com a primeira. As imagens resultantes do agrupamento podem ser vistas em 6.12(d) e 6.13(d), onde

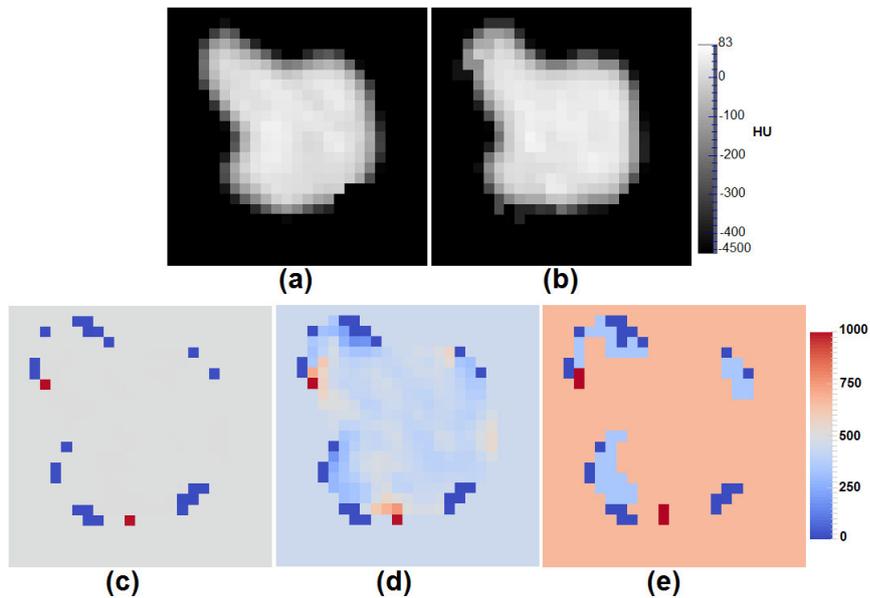
Tabela 6.2 – Conjunto de valores de percentuais de mudança obtidos do método ACP para cada caso da base privada de lesões.

Caso	P_m (%)	
	ACP	VVR
1	12,84	-2,86
2	15,65	5,71
3	11,73	8,82
4	9,84	1,35
5	16,05	7,14
6	13,70	10,29
7	19,18	7,41
8	16,08	20,37
9	19,23	9,26
10	20,82	12,07
11	21,64	-9,23
12	17,26	1,72
13	24,25	12,5
14	33,89	-60
15	12,25	-3,51
16	16,80	-23,64
17	18,82	-26,32

Fonte: Autor

apenas o grupo que corresponde ao fundo é o grupo de não-mudança e os demais grupos são as componentes de mudança.

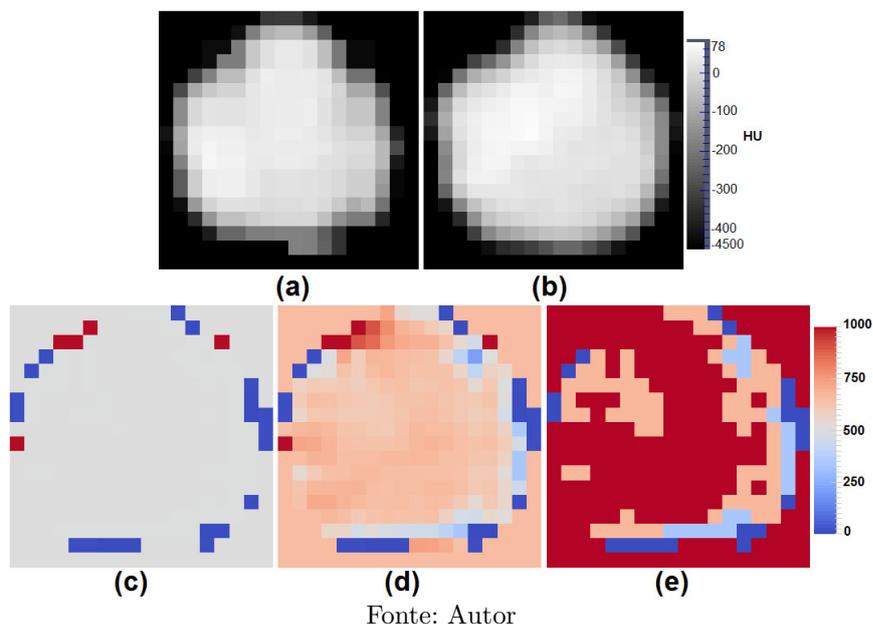
Figura 6.12 – Resultado final da aplicação da ACP para o Caso 4.



Fonte: Autor

O pequeno valor de P_m encontrado para o Caso 4 é resultado das poucas mudanças ocorridas internamente à lesão durante o intervalo de tempo do par de instantes analisados. Como se pode visualizar na Figura 6.12, que é apenas uma única imagem do volume da

Figura 6.13 – Resultado final da aplicação do ACP para o Caso 12.



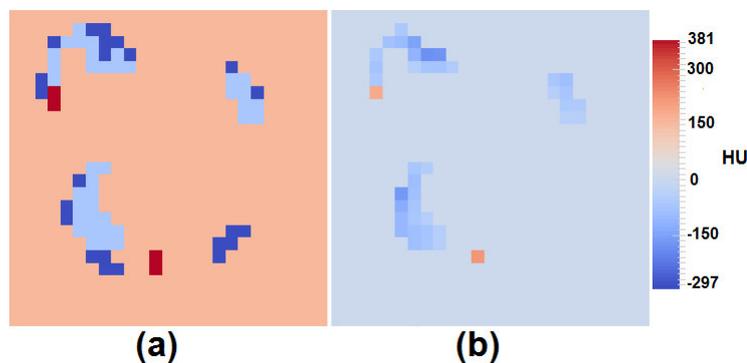
lesão, mas de mesmo comportamento ao longo do volume, as modificações detectadas foram mais extensas nas mudanças lesão-fundo quando comparadas às mudanças lesão-lesão.

A Figura 6.13 do Caso 12, permite observar que internamente à lesão pulmonar, mais especificamente nas mudanças lesão-lesão, as modificações não foram uniformes ao longo do volume. Parte da lesão internamente foi classificada como inalterada em relação à densidade. Esta observação fica facilitada quando se observa o grupo pertencente à componente que não muda ao longo do tempo, que são os *voxels* pertencentes ao fundo, isto é, *voxels* de mesma coordenada das imagens do volume da lesão no par de instantes em análise e que pertencem ao fundo.

6.3.2 Classificação de lesões quanto ao diagnóstico

O objetivo desta análise é verificar se as mudanças de densidade encontradas pela aplicação da ACP corroboram com o diagnóstico. Um segundo objetivo é verificar se o modelo de classificação encontrado com as medidas de textura é suficiente para prever o diagnóstico da lesão pulmonar, independente da mesma se encontrar ou não em tratamento. Pode-se visualizar um exemplo da textura analisada sobre o resultado da ACP, na Figura 6.14, onde em (a), tem-se a imagem resultante do agrupamento e as regiões com cores diferentes do vermelho fraco (fundo) indicam as mudanças; e a imagem em (b) é a diferença das imagens da lesão para essas regiões de mudança.

Figura 6.14 – Exemplo das regiões em que serão extraídas as características de textura para classificação de lesões em maligno e benigno.



Fonte: Autor

As características utilizadas na tarefa de classificação foram definidas na Seção 3.5. O total de características obtidas foram 36. Em razão do grande número de características, realizou-se a seleção de características com o método *Random Forest and RM-bound MVS* (CHEN; LIN, 2006). Da seleção, apenas a média da diferença (MD) das medidas estáticas não foi selecionada. Logo, um total de 3 características foram utilizadas para o método proposto.

Devido a pequena amostra para a tarefa de classificação, foi utilizada a técnica de amostragem *leave-one-group-out* (KOHAVI, 1995) como forma de avaliar o modelo gerado na classificação. O número de grupos utilizado foi 10.

O classificador utilizado foi o MVS com *kernel RBF* (Seção 3.6.1) e os parâmetros de classificação, C e g , estão na Tabela 6.3. A classificação também foi realizada com os parâmetros padrões, isto é, $C = 1$ e $g = 1/3$, em que 3 é o número de características.

O resultado da classificação quanto ao diagnóstico para amostra contendo as 63 instâncias, das quais 17 instâncias são casos benignos (base privada) e 46 instâncias são casos malignos (base pública), está na Tabela 6.3.

Tabela 6.3 – Métricas de avaliação da classificação utilizando MVS associadas a cada um dos parâmetros do modelo.

Avaliação do ACP								
Parâmetros do MVS		Métricas de avaliação						
C	g	sens	esp	ac	vp	vn	fp	fn
1	1/3	93,48	94,12	93,65	43	16	1	3
128	0,000122	95,65	94,12	95,24	44	16	1	2

Fonte: Autor

6.4 Discussão

Para a avaliação da pequena variação volumétrica para a base pública, mesmo observando somente através de uma única imagem do volume, é possível verificar que mudanças morfológicas foram mais acentuadas para o Caso 10. Tais modificações morfológicas foram visíveis através do resultado da ACP para o volume da lesão.

Analisando as imagens geradas para os nódulos pulmonares da base privada que apresentaram os menores valores de VVR , Casos 4 e 12, as mudanças ocorreram em menor intensidade para o Caso 4 que para o Caso 12. Da mesma forma, as regiões de mudança lesão-lesão foram mais frequentes nos limites de ambas as lesões.

Em uma observação com base apenas nos valores de P_m para ambas as bases, pública e privada, não parece revelar alguma distinção entre os casos observados. Entretanto, analisando a variação de densidade das regiões de mudança com simples descritores de textura, fazem revelar padrões discriminantes entre as duas bases de lesões com comportamentos completamente distintos. Esse padrão nas mudanças de densidade foi revelado através do uso do MVS. Logo, embora P_m sozinho não seja um bom indicador para a natureza da lesão, o mesmo é um novo quantitativo para a análise das mudanças da lesão.

Isso salienta que, mesmo que os P_m obtidos não revelem em uma primeira inspeção a natureza da lesão pulmonar ao longo do tempo, os padrões de textura extraídos pelas técnicas de Haralick e RL e que possuem funcionamento direcional, mostram que as variações de densidade obtidas nas regiões detectadas como mudança são promissoras na discriminação quanto ao diagnóstico da lesão pulmonar.

Mesmo com uma amostra pequena e da existência do desbalanceamento entre o número de casos malignos e benignos, proporção de 2,7 malignos para cada benigno, as métricas de avaliação da classificação foram satisfatórias. Afinal, as métricas de avaliação se encontram próximo dos 100% e o MVS não responde bem a um treinamento desbalanceado dos dados.

6.5 Conclusão

A análise de mudanças de densidade utilizando a análise de componentes principais provou ser uma valiosa ferramenta para a análise das modificações de densidade da lesão ao longo do tempo.

Os resultados mostram que, mesmo utilizando duas bases de lesões com comportamentos distintos, as densidades das lesões pulmonares sofreram alterações ao longo do tempo. A detecção das mudanças de densidade já era esperada para lesões em tratamento, mas não da mesma forma para nódulos benignos. Entretanto, tal comportamento é explicado pela característica intrínseca de detectar sempre duas classes, a de mudança e outra de não-mudança.

A análise sobre os valores obtidos de P_m mostrou que existem mudanças de densidade para lesões malignas e benignas em situações de pequena e grande variação volumétrica. Entretanto, as mudanças não ocorrem com a mesma intensidade, nem na mesma região e extensão para as lesões malignas e benignas. Como resultado, simples descritores de textura foram capazes de discriminar a natureza da lesão, baseados apenas nas regiões onde ocorreram as mudanças de densidade.

7 Método baseado em agrupamento

O capítulo possui os fundamentos necessários para compreensão do método de agrupamento proposto. Nele, há uma descrição geral do método de agrupamento *quality threshold* (QT) e como o mesmo foi utilizado para a tarefa de detectar mudanças de densidade. Os resultados obtidos da aplicação são apresentados e discutidos ao final do capítulo.

7.1 Método de agrupamento *quality threshold*

O QT (JIANG; TANG; ZHANG, 2004) é um método de agrupamento de fácil implementação e de alto custo computacional, quando comparado ao *k-means* e *c-means* (XU; WUNSCH, 2005). Sua principal característica está na geração do melhor número de grupos para um conjunto de dados de entrada.

Para a geração do melhor número de grupos, um único parâmetro precisa ser provido, chamado de *quality threshold*. Deste ponto em diante, esse parâmetro será denominado apenas de *threshold*. Este define o tamanho máximo do grupo. Portanto, devido a sua simplicidade de uso, por possuir um único parâmetro e da sua intrínseca capacidade de encontrar o número ideal de grupos em um conjunto de dados, o QT foi utilizado neste trabalho para verificar mudanças de densidade da lesão ao longo do tempo.

Os passos para o processo de agrupamento de dados utilizando o QT são (FILHO et al., 2014): 1- Dado um conjunto de dados, definir um *threshold* para a formação dos grupos; 2- Para cada um dos pontos do conjunto de dados, a distância euclidiana é obtida para todos os outros pontos, sem repetição do par. Assim, todo ponto tomado como referência é chamado de centro do grupo e todos os outros cujas distâncias são menores ou iguais ao valor de *threshold* pertencem a um grupo. 3- A etapa 2 é repetida para todos os outros pontos não pertencentes ao grupo recém-formado, até que não reste mais pontos no conjunto de dados.

O número de grupos é função do *threshold* definido e da quantidade de elementos no conjunto de dados. Portanto, apesar do *threshold* ser o único parâmetro para o QT, este é de grande importância na determinação do número de grupos. Para melhorar esse aspecto, modificou-se o QT para torná-lo o mais automático possível.

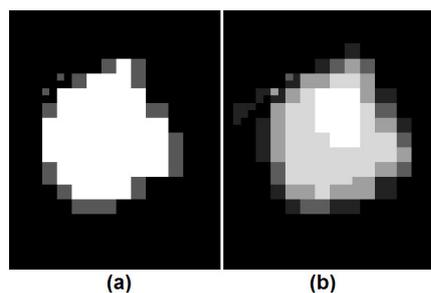
7.1.1 *Quality threshold modificado*

Do ponto de vista prático, foram duas alterações realizadas sobre o QT. Ambas alterações foram essenciais, pois tornaram possível capturar mudanças nas densidades que não eram possíveis sem as alterações. Cada uma delas será explicada a seguir.

A primeira modificação consiste na inclusão do método de avaliação da qualidade do grupo, o *jump method* (SUGAR; JAMES, 2003). Então, dado um intervalo provável do *threshold*, o *jump method* avalia o *threshold* que resulta no melhor número de grupos (k). O valor de K^* no *jump method* é dado pelo maior valor k com máxima distorção, conforme a linha 8 do algoritmo 1. Após a análise dos resultados do *jump method*, achou-se que esse não estava adequado ao nosso propósito, pois o valor de k era muito pequeno. Em adição ao pequeno número de grupos, os grupos gerados não representavam bem todo espectro de valores de densidade da lesão. O *jump method* não consegue formar grupos internamente à lesão, como se pode visualizar na Figura 7.1(a), onde apenas três grupos são formados e todos os *voxels* internos à lesão pertencem a um único grupo. Isso é um problema, visto que esse resultado não favorece a detecção das mudanças de densidade ao longo do tempo.

Para contornar esse problema, modificou-se a etapa da escolha do valor k na linha 8 do algoritmo 1 do *jump method*. Essa segunda modificação consiste em encontrar o valor de k que resulta na maior distância da reta linear que passa sobre os extremos da curva de distorção para a mesma. Um exemplo dessa curva está na Figura 7.2, que contém um gráfico do erro médio quadrático (EMQ) em função do valor k . Nele, está indicada em verde a maior diferença obtida e o correspondente melhor número de grupos K^* . Na Figura 7.1(b), tem-se a imagem resultante do *jump method* modificado. Nessa figura, alguns *voxels* internos à lesão foram identificados como um grupo, fato que não ocorreu com a imagem em Figura 7.1(a), resultante do *jump method* sem modificação.

Figura 7.1 – Resultado do agrupamento com o *jump method* sem modificação.



Fonte: Autor

Algoritmo 1 *jump method*. Método que determina o melhor número de grupos.

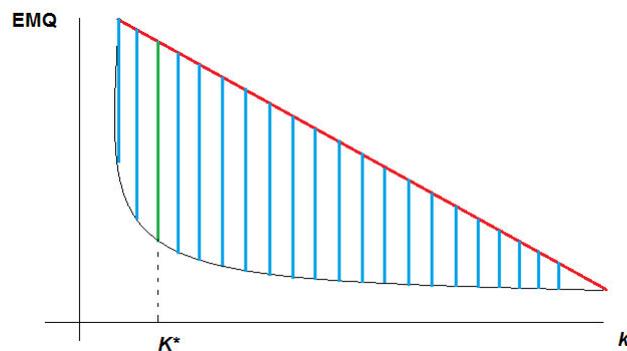
```

1: procedure BUSCAMELHORNUMGRUPOS
2:    $y = \dim(X)$ 
3:   for  $k = 1, \dots, n$  do
4:      $distortion = QT Agrup(k)$ 
5:      $D[k] = distortion^{(-y)}$ 
6:     if  $k > 1$  then
7:        $J[k] = D[k] - D[k - 1]$ 
8:    $K = argMax(J[k]) \quad - - - > \quad K = argMax(|D[k] - DEstimado[k]|)$ 

```

Fonte: Autor

Figura 7.2 – Curva EMQ x k . Em verde, está a maior diferença que resultou no melhor valor k .



Fonte: Autor

7.2 Identificação das mudanças de densidade

A análise temporal foi realizada através da aplicação do QT modificado. O objetivo é agrupar todos os *voxels* da lesão em diferentes instantes de aquisição. As mudanças de densidade são detectadas quando um *voxel* muda de um grupo para outro. Portanto, o método para detecção das mudanças de densidade com o QT modificado utiliza a ideia de mudança de grupo para esse propósito. De outra forma, dada uma lesão cujos tecidos estão se modificando ao longo do tempo, a variação temporal do *voxel*, quanto ao grupo o qual ele pertence, indica a referida mudança tecidual.

O método consiste em agrupar a lesão pulmonar em um instante (t_1) com o QT modificado. Com o resultado do agrupamento, os centros dos grupos encontrados pelo QT modificado são utilizados para agrupar a lesão em um instante futuro (t_2). Esse passo objetiva verificar as mudanças de densidade em regiões correspondentes no par de lesões em análise.

Notadamente, a detecção das mudanças de densidade ocorre quando um ou mais *voxels* de uma lesão em um instante não estão no mesmo grupo da mesma lesão em um diferente instante. Logo, é possível verificar quais grupos se modificaram nos diferentes instantes da lesão e através de medidas, tais como P_m , P_i e $R_i(t)$ (Seção 3.8), além de poder inferir o quanto cada grupo mudou e a sua importância para lesão.

As métricas de quantificação das mudanças descrevem as mudanças de densidade para cada grupo formado. Como o QT modificado determina o número de grupos em função da distribuição de densidade da lesão, essas medidas não são indicadas para prever o diagnóstico da lesão ao longo tempo. Para contornar esse problema, medidas estatísticas (Seção 3.5.3) serão utilizadas para descrever a lesão nas regiões onde ocorrem as mudanças de densidade ao longo tempo.

7.3 Resultados

Para atingir os resultados esperados na identificação das mudanças de densidade, as seguintes etapas iniciais foram definidas conforme as respectivas seções: base de lesões pulmonares (Seção 4.7.2), modelos de lesões (Seção 4.7.1) e registro das imagens da lesão (Seção 4.7.3). Em seguida, apresentam-se todos os resultados quanto a sensibilidade às mudanças encontradas para as bases pública e privada de lesões pulmonares. Além disso, realiza-se um estudo de casos para as lesões que apresentaram pequena variação volumétrica. Na última etapa, mostra-se o quanto as características extraídas das regiões de mudança são indicadoras do diagnóstico da lesão pulmonar.

7.3.1 Aplicação do método proposto

A detecção das mudanças de densidade com o QT modificado é realizada sobre as lesões registradas. Isso é realizado para que exista tanto a correspondência entre os grupos quanto a correspondência espacial dos *voxels*. Em seguida, devido a modificação do QT, define-se o intervalo de variação do *threshold* entre 1 e 1000. Este compreende ao mais provável intervalo de variação de densidade interno à lesão.

Os valores de *threshold* encontrados pelo QT modificado para a base pública estão no intervalo de 19 a 191 e média de 59,42. Para a base HPE, os valores encontrados de *threshold* estão no intervalo de 19 a 73 e média de 36,41.

Os maiores valores de *threshold* foram encontrados para a base pública. Em função do *threshold*, o número de grupos encontrados para essa base está no intervalo de 4 a 40 e média de 18,67. Da mesma forma, o número de grupos encontrados para base HPE está no intervalo de 3 a 20 e média de 12,47.

Na seção seguinte, analisa-se o comportamento dos grupos formados com o QT modificado para os modelos de lesões propostos.

7.3.1.1 Estudo de casos com modelos de lesões

O QT modificado não teve nenhum sucesso na detecção das mudanças de densidade dos modelos de lesões. O método não detectou nenhuma das mudanças. Entretanto, isso pode ser perfeitamente explicado. O método é baseado em agrupamento e foi modificado para automaticamente encontrar o número ideal de grupos durante a etapa de agrupamento. Assim, como a esfera inicial da Figura 4.11(a) é composta de densidade única igual a 1000 HU e com valor de fundo igual a -1024 HU, o QT modificado encontrou apenas dois grupos. Logo, um dos grupos da esfera é aquele correspondente a sua formação (1000 HU) e o outro correspondente ao fundo. Devido o processo de detecção das mudanças de densidade utilizar a lesão ou o modelo do instante inicial como referência para formação dos grupos, os mesmo dois grupos foram identificados na esfera no instante final da Figura 4.11(b). Portanto, todas esferas de mudança foram agrupadas no mesmo grupo do valor da esfera inicial (1000 HU) e nenhuma das mudanças de densidade foram detectadas.

Nas seções seguintes, analisa-se o comportamento dos grupos formados com o QT modificado, para os casos com *VVR* próximos do valor zero. Esses casos demonstram a importância do método proposto na identificação das mudanças de densidade, pois pode ser utilizado em situações em que o especialista não consegue percebê-las via inspeção visual. Da mesma forma, o método proposto pode auxiliar o especialista quando outras ferramentas computacionais não conseguem determinar com precisão pequenas variações volumétricas.

7.3.1.2 Estudo de casos com a base pública de lesões

O Caso 10 teve o menor *VVR*, 1,61%, mas muitas regiões da lesão sofreram redução dos valores de densidade. Uma das medidas indicativas dessa redução é o percentual de

mudança encontrado, igual a 27,06%. O *threshold* encontrado pelo QT modificado foi 37, com 22 grupos formados. Desses, 18 grupos possuem densidade negativa no instante inicial. Analisando os valores P_i , apenas 5 grupos de densidade negativa apresentaram aumento na quantidade de *voxels* ao longo do tempo. Entretanto, observando de forma complementar a relevância desses grupos no instante t_2 , conclui-se que esses são os mais relevantes, em especial o grupo de densidade -936 HU. A configuração dos grupos na lesão nos diferentes instantes pode ser vista na Figura 7.3(a-c) e os valores de densidade dos centros dos grupos obtidos pelo QT modificado estão na Tabela 7.1.

Tabela 7.1 – Conjunto de valores das medidas que quantificam as mudanças de densidade para valor do centro do grupo do Caso 10.

Caso 10 - PLD			
VVR	Threshold	Número de grupos	P_m
1,61	37	22	27,06
Valor do centro	P_i(%)	$R(t_1)$(%)	$R(t_2)$(%)
-936	101,51	73,13	74,23
-856	75,85	0,20	0,15
-818	85,59	1,69	1,45
-769	81,24	4,26	3,46
-708	82,28	3,28	2,70
-667	86,67	2,40	2,08
-602	83,76	1,86	1,55
-562	79,75	1,30	1,04
-515	86,96	1,12	0,97
-471	78,85	1,04	0,82
-418	78,60	1,05	0,82
-361	76,81	1,09	0,84
-286	82,94	0,99	0,82
-225	87,87	0,92	0,80
-156	114,75	0,65	0,75
-112	126,69	0,61	0,77
-49	132,64	0,73	0,96
-8	152,21	1,44	2,20
54	161,47	2,10	3,39
109	147,01	0,10	0,14
148	80,00	0,04	0,03
210	82,35	0,01	0,01

Fonte: Autor

O segundo menor valor de *VVR* (-3,12%) foi encontrado para o Caso 28 da Tabela 7.2. Novamente, a lesão teve uma redução nos valores de densidade e o valor de P_m encontrado foi igual a 15,50%. O QT modificado encontrou, para esse caso, um *threshold* igual a 50 e o número de grupos formados foi igual a 25. Esses valores são maiores que os obtidos no caso 10. A Figura 7.3(d-f) possui uma imagem do volume da lesão nos instantes analisados e as configurações dos grupos encontrados. A Tabela 7.2 possui os

valores dos centros dos grupos encontrados e os valores de P_i e $R_i(t)$ do i -ésimo grupo. A lesão teve mais grupos negativos e, de acordo com os valores $R_i(t_2)$, esses grupos são mais dominantes no instante t_2 . Observando esse resultado, as medidas de avaliação determinam uma diminuição do perfil de densidade da lesão com a diminuição volumétrica.

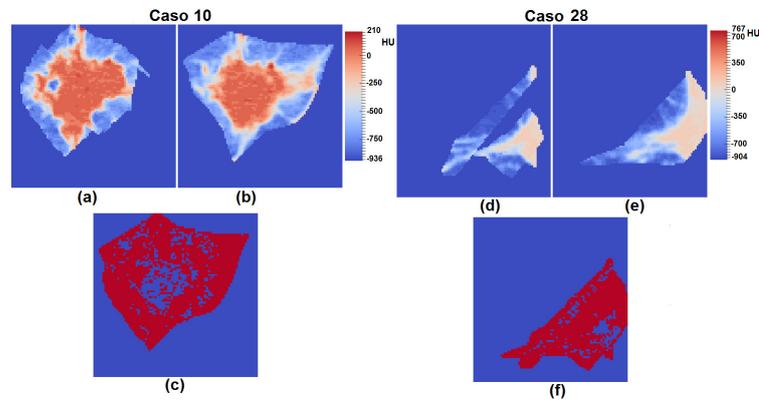
Tabela 7.2 – Conjunto de valores das medidas que quantificam as mudanças de densidade para valor do centro do grupo do Caso 28.

Caso 28 - PLD			
VVR	Threshold	Número de grupos	P_m
-3,12	50	25	15,5
Valor do centro	$P_i(\%)$	$R(t_1)(\%)$	$R(t_2)(\%)$
-904	100,98	86,20	87,04
-851	117,86	0,49	0,57
-800	111,84	1,16	1,29
-739	106,17	1,70	1,81
-638	97,81	1,46	1,43
-561	90,92	0,96	0,87
-505	93,92	0,71	0,66
-444	97,66	0,85	0,83
-354	97,67	0,69	0,67
-303	90,59	0,48	0,43
-250	98,96	0,48	0,48
-193	83,13	0,50	0,41
-136	84,55	0,58	0,49
-59	78,14	0,92	0,72
18	89,84	1,76	1,58
83	67,28	0,97	0,65
150	35,48	0,06	0,02
269	107,69	0,01	0,01
349	40,00	0,00	0,00
405	125,00	0,00	0,00
504	250,00	0,00	0,00
605	16,67	0,01	0,00
660	25,00	0,00	0,00
714	25,00	0,00	0,00
767	27,27	0,01	0,00

Fonte: Autor

Em contraste aos casos mostrados de pequena variação volumétrica, o Caso 42 teve o maior valor de VVR . Esse caso é utilizado para mostrar que uma grande variação no volume não implica em muitos grupos formados. A Figura 7.4 possui as imagens da lesão nos instantes analisados e como estão distribuídos os grupos gerados pelo QT modificado. Foram 20 grupos encontrados pelo QT modificado para o *threshold* igual a 38. A predominância é de grupos com densidade negativa e todos os 17 grupos negativos presentes representam 99,06% do volume no instante t_2 , de acordo com somatório das relevâncias encontradas. Somente o grupo de densidade -900 HU teve aumento ao longo do tempo. Logo, apesar do aumento do volume, esta se tornou menos densa. A Tabela 7.3

Figura 7.3 – (a),(d) Lesão 1 e (b),(e) Lesão 2, na qual as cores são resultado da aplicação da palheta adjacente associada com o valor em HU. A imagem binária em (c),(f) é o resultado da aplicação do QT modificado.



Fonte: Autor

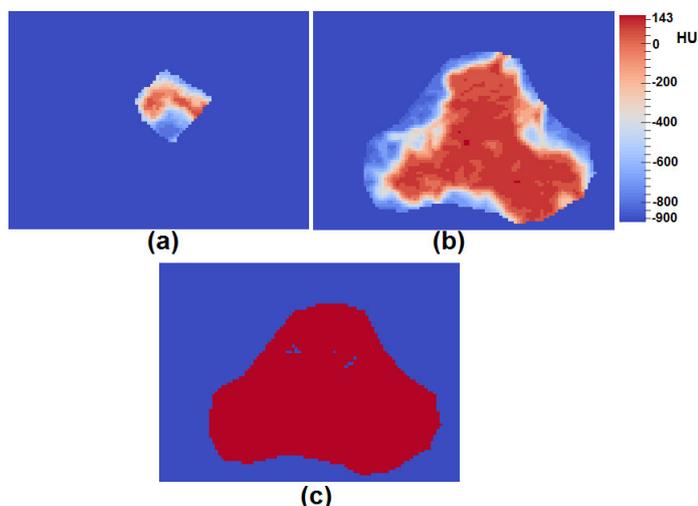
possui os valores dos centros dos grupos encontrados e os valores de P_i e $R_i(t)$ do i -ésimo grupo.

Tabela 7.3 – Conjunto de valores das medidas que quantificam as mudanças de densidade para valor do centro do grupo do Caso 42.

Caso 42 - PLD			
VVR	Threshold	Número de grupos	P_m
521,3	38	19	33,77
Valor do centro	$P_i(\%)$	$R(t_1)(\%)$	$R(t_2)(\%)$
-900	143,82	65,73	94,53
-847	0,36	0,95	0,00
-804	4,22	2,07	0,09
-751	8,71	2,59	0,23
-693	14,11	2,03	0,29
-654	15,31	1,42	0,22
-615	18,88	1,16	0,22
-582	22,47	1,11	0,25
-542	22,16	1,26	0,28
-498	26,01	1,26	0,33
-443	30,49	1,44	0,44
-367	27,44	1,66	0,46
-300	24,83	1,44	0,36
-232	27,18	1,41	0,38
-163	23,52	1,52	0,36
-94	19,31	1,71	0,33
-23	15,21	2,06	0,31
34	11,18	3,85	0,43
88	8,88	5,20	0,46
143	32,99	0,13	0,04

Fonte: Autor

Figura 7.4 – (a) e (b) são os grupos para o Caso 42 nos diferentes instantes. Em (c), têm-se as regiões de mudança.



Fonte: Autor

7.3.1.3 Estudo de casos com a base privada de lesões

A base contém apenas nódulos pulmonares benignos que foram considerados estáveis após um período de três anos de avaliação.

O Caso 4 teve o menor VVR , 1,35%. Embora essa lesão tenha tido um pequeno aumento volumétrico, essa apresentou mudança em 22,07% (P_m) dos seus *voxels*. Isso se deve ao grande número de grupos formados, 19 grupos, pois o *threshold* encontrado pelo QT modificado foi igual a 21. Dos 19 grupos encontrados, 16 deles possuem densidade negativa e 14 desses grupos negativos aumentaram ao longo do tempo. Esses mesmos grupos de densidade negativa somam 83,07% de relevância no instante t_2 . Isso implica que o pequeno aumento volumétrico não fez com que a lesão ficasse mais densa. Na Figura 7.5(a-c), têm-se as configurações de densidade encontradas para esse caso; e na Tabela 7.4, as densidades dos grupos e as medidas de mudança encontradas.

O Caso 12 teve o segundo menor VVR , 1,72%. Apesar da pequena variação volumétrica, o P_m encontrado foi de 33,95%. Essa lesão também sofreu redução de densidade em sua expansão volumétrica. De fato, dos 14 grupos encontrados com o *threshold* igual a 30, 5 dos 11 grupos de densidade negativa tiveram aumento ao longo do tempo, conforme os seus respectivos valores de P_i da Tabela 7.5. Outro fator que contribuiu para essa redução de densidade é a existência de um único grupo de densidade positiva aumentando no tempo, mas, de acordo com o seu valor de $R_i(t_2)$, este possui pouca relevância para o

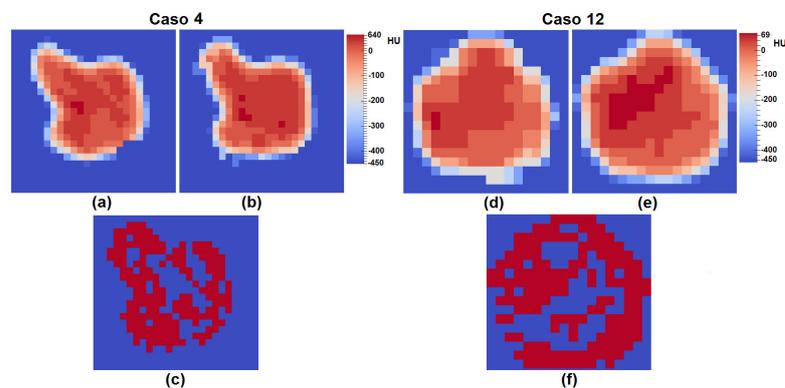
Tabela 7.4 – Conjunto de valores das medidas que quantificam as mudanças de densidade para valor do centro do grupo do Caso 4.

Caso 4 - HPE			
VVR	Threshold	Número de grupos	P_m
1,35	21	19	22,07
Valor do centro	$P_i(\%)$	$R(t_1)(\%)$	$R(t_2)(\%)$
-450	100,27	72,69	72,88
-428	115,87	0,72	0,83
-392	90,38	1,19	1,07
-349	77,88	1,29	1,01
-311	104,48	0,77	0,80
-288	138,64	0,50	0,70
-264	74,39	0,94	0,70
-232	88,51	0,99	0,88
-197	83,33	1,10	0,91
-166	117,11	0,87	1,02
-139	105,13	0,89	0,94
-115	107,69	0,89	0,96
-92	76,99	1,29	0,99
-65	110,42	1,10	1,21
-36	104,83	1,66	1,74
-14	102,34	1,95	2,00
8	118,30	4,43	5,25
35	98,65	5,92	5,84
64	33,80	0,81	0,27

Fonte: Autor

volume. A Figura 7.5(d-f) possui a lesão analisada no tempo, o resultado do agrupamento com o QT modificado e a imagem das mudanças da fatia.

Figura 7.5 – (a),(d) Lesão 1 agrupada e (b),(e) Lesão 2 agrupada, na qual as cores são resultado da aplicação da palheta adjacente associada com o valor em HU. A imagem binária em (c),(f) é a imagem binária de mudança.



Fonte: Autor

O Caso 14 teve o maior VVR em módulo, $-60,00\%$, e indica uma redução volumétrica. Essa lesão se tornou mais densa com a redução do volume. Esse efeito é resultado da determinação dos 11 grupos em t_1 , para o corresponde *threshold* encontrado igual a 70 e

Tabela 7.5 – Conjunto de valores das medidas que quantificam as mudanças de densidade para valor do centro do grupo do Caso 12.

Caso 12 - HPE			
VVR 1,72	Threshold 30	Número de grupos 14	P_m 33,95
Valor do centro	P_i (%)	$R(t_1)$ (%)	$R(t_2)$ (%)
-445	108,24	55,13	59,68
-414	102,94	1,91	1,96
-377	121,54	1,82	2,22
-340	75,00	2,81	2,10
-283	78,43	2,86	2,24
-247	90,80	2,44	2,22
-188	104,30	2,61	2,72
-152	111,29	1,74	1,94
-116	72,38	2,95	2,13
-73	88,99	3,06	2,72
-34	84,76	4,60	3,90
3	108,28	8,14	8,81
37	90,97	7,77	7,07
69	12,99	2,16	0,28

Fonte: Autor

somente 3 grupos em t_2 . Mas dos 3 grupos finais, dois deles apresentam densidade negativa e apenas um com densidade positiva. Pelos valores de $R_i(t_2)$ para esses 3 grupos finais, pode-se concluir que a lesão se tornou mais densa, pois o grupo de densidade positiva é o mais relevante e apresentou o maior aumento no tempo.

Na Tabela 7.6, podem-se encontrar as densidades dos grupos encontrados; e na Figura 7.6, como ficou a configuração dos grupos para este caso. Como se pode visualizar na Tabela 7.6, o nódulo pulmonar perdeu 5 grupos de densidade positiva e 3 grupos de densidade negativa após a sua redução. Esse resultado implica que a redução fez a lesão diminuir em densidade em algumas regiões e se calcificar em outras próximo da densidade de 32 HU.

7.3.2 Classificação de lesões quanto ao diagnóstico

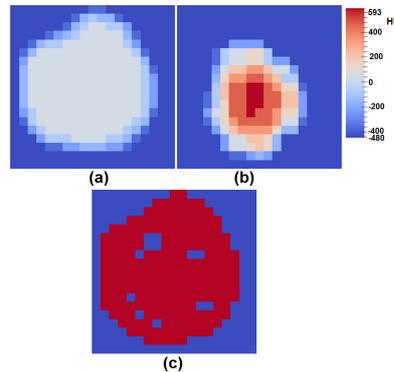
A análise deseja verificar se as mudanças de densidade identificadas pelo QT modificado confirma o diagnóstico ou verificar se o modelo de classificação encontrado com as medidas de textura é suficiente para prever o diagnóstico da lesão pulmonar. Isso é realizado independente da lesão estar ou não em tratamento e baseia-se apenas sobre as medidas de textura nas regiões das mudanças ocorridas. Um exemplo das texturas analisadas sobre o resultado do QT modificado é mostrado na Figura 7.7. A Figura 7.7(a)

Tabela 7.6 – Conjunto de valores das medidas que quantificam as mudanças de densidade para valor do centro do grupo do Caso 14.

Caso 14 - HPE			
VVR -60,00	Threshold 71	Número de grupos 11	P_m 36,06
Valor do centro	P_i (%)	$R(t_1)$ (%)	$R(t_2)$ (%)
-448	0,00	84,86	0,00
-371	185,71	2,26	4,20
-240	194,68	2,17	4,22
-162	0,00	1,15	0,00
-86	0,00	1,57	0,00
32	970,59	1,96	19,04
127	0,00	1,66	0,00
208	0,00	1,66	0,00
336	0,00	1,78	0,00
460	0,00	0,76	0,00
593	0,00	0,16	0,00

Fonte: Autor

Figura 7.6 – (a),(d) Lesão 1 agrupada e (b),(e) Lesão 2 agrupada, na qual as cores são resultado da aplicação da palheta adjacente associada com o valor em HU. A imagem binária em (c),(f) é a imagem binária de mudança.

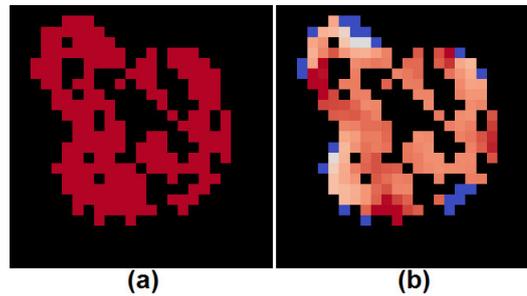


Fonte: Autor

mostra a imagem binária e as regiões em vermelho indicam as mudanças. A imagem na Figura 7.7(b) é a diferença das imagens da lesão para as mudanças nas regiões de mudança detectadas.

Embora tenha um pequeno número de características, total de cinco, nem todas contribuem para uma melhor acurácia de classificação. Portanto, para reduzir o número de características, realizou-se a seleção de características com o método *Random Forest and RM-bound* MVS (CHEN; LIN, 2006). O número de características selecionadas foi um, o máximo valor da diferença (MAXD).

Figura 7.7 – Exemplo das regiões a partir dos quais as características são extraídas para a classificação da lesão em benigno e maligno.



Fonte: Autor

Em razão do pequeno tamanho da amostra para a classificação, usou-se a técnica *leave-one-out* (KOHAVI, 1995) para avaliar o modelo gerado na classificação. O número de grupos utilizados para a técnica foi 10.

O classificador utilizado foi o MVS (VAPNIK, 1998). Para a classificação, o *kernel* utilizado foi a função de base radial e os parâmetros de classificação (C e g) estão na Tabela 7.7. Em seguida, a classificação foi realizada em duas configurações. A primeira utiliza os valores padrões, $C = 1$ e $g = 1$, tal que o valor de g é o número de características selecionadas. A segunda configuração foi realizada com a estimação dos mesmos parâmetros. Em cada uma das configurações citadas, os resultados da classificação com a técnica *leave-one-out* são obtidos.

Os resultados da classificação quanto ao diagnóstico foram obtidos sobre uma amostra de 63 casos, dos quais 17 são benignos (base privada) e 46 são malignos (base pública). Os resultados completos estão na Tabela 7.7 para todas configurações do classificador.

Tabela 7.7 – Métricas de avaliação da classificação utilizando MVS associadas a cada um dos parâmetros do modelo.

Avaliação do QT								
Parâmetros do MVS		Métricas de avaliação						
C	g	sens(%)	esp(%)	ac(%)	vp	vn	fp	fn
1	1	97,83	100,00	98,41	45	17	0	1
0,125	2	97,83	100,00	98,41	45	17	0	1

Fonte: Autor

Em ambos os parâmetros de classificação da Tabela 7.7, o Caso 20 da base pública (malignos) foi erroneamente classificado. O caso apresenta pequena variação

volumétrica. Portanto, nesse caso, as mudanças de densidade podem ter sido similares às aquelas encontradas nos casos benignos.

7.4 Discussão

Os resultados obtidos na determinação do *threshold* confirmam que o intervalo definido foi apropriado para ambas as bases de lesões pulmonares.

A variação do *threshold* encontrado para as lesões em tratamento foi maior, apesar do mesmo valor mínimo.

As lesões em tratamento possuíram os maiores *thresholds*, o que não era esperado. Mesmo que o *threshold* seja inversamente proporcional ao número de grupos, as lesões em tratamento deveriam ter os menores valores para o número de grupos formados. Entretanto, esse resultado é aceitável, pois são lesões malignas, logo, sua heterogeneidade é maior.

O QT modificado facilita a visualização das mudanças de densidade em ambas as bases de lesões pulmonares. Essa facilidade ocorre em razão do processo de agrupamento, que possibilita detectar mudanças pequenas no perfil de densidade das lesões. Logo, caso a lesão tenha mudanças de densidades ao longo do tempo, grupos diferentes serão formados e conseqüentemente ocorrerá a transição de um *voxel* de um grupo para outro.

7.5 Conclusão

O método utilizado para identificação das mudanças de densidade foi o QT modificado. O método foi modificado especialmente para a análise temporal das densidades da lesão. A aplicação é realizada sobre duas bases de imagens, uma maligna em tratamento (pública) e a outra de nódulos benignos (privada) em acompanhamento. A partir da aplicação do QT modificado sobre ambas as bases, a identificação das mudanças de densidade mostrou que nos casos de pequena variação volumétrica, o método é capaz de realçar as mudanças de densidade.

O realce produzido pelo método ocorre quando um *voxel* pertencente a um grupo muda para outro. Como resultado, é possível visualizar mais claramente quais *voxels* mudam de natureza. Esse resultado prova que mesmo que uma lesão possa ser considerada estável, mudanças de densidade ocorrem. Ao mesmo tempo, independente da base, os resultados indicam que muitas dessas mudanças podem ser realçadas através da aplicação

do método proposto. Fato que foi mostrado para lesões malignas e benignas, que neste último caso, inicialmente eram indeterminadas.

Na análise das mudanças das lesões da base pública, pode-se inferir que muitas delas sofrem efeitos consideráveis em seus *voxels* quando estão em tratamento. Essa percepção é possível quando se analisa os *voxels* através do percentual de mudança intra-grupo, em conjunto com a relevância. Essa última mostra como aquele grupo predomina em qualquer um dos instantes e se torna mais interessante quando é vista no instante final (t_2). Logo, quando o P_i indica crescimento ou diminuição do grupo no tempo, a relevância ajuda a mostrar o quanto essa mudança é importante para o volume. Portanto, quando as lesões malignas foram analisadas, as mesmas mostraram uma redução de suas densidade, mesmo quando estava se expandindo em volume.

Para as lesões benignas, os nódulos pulmonares da base privada mostraram comportamento bastante semelhante aos casos malignos em tratamento. Durante essa análise, quando as lesões estavam se expandindo volumetricamente, mesmo de forma imperceptível, suas densidades estavam decrescendo. Isso foi percebido quando se analisa as métricas de avaliação propostas para análise das mudanças. Entretanto, essa redução não é uma regra. Considerando a lesão do Caso 14, em que uma grande redução volumétrica ocorreu, ela se tornou mais densa ao longo do tempo, sendo essas características essenciais para a discriminação entre malignos e benignos.

Além das informações de mudança, o especialista pode visualizar de forma tridimensional como essas densidades estão sendo alteradas. Acredita-se que essa é uma importante contribuição para a análise temporal de lesões que estão em tratamento e de lesões que são consideradas indeterminadas inicialmente. Portanto, a análise tridimensional das mudanças de densidade e as medidas que avaliam essas mudanças podem contribuir para o auxílio do especialista.

8 Comparação dos métodos

Este capítulo expõe algumas das dificuldades encontradas na comparação com outros trabalhos e realiza uma comparação dos resultados obtidos entre todos os métodos propostos.

8.1 Dificuldades de comparação com trabalhos relacionados

Devido aos trabalhos relacionados da Seção 2 utilizarem bases de imagens privadas ou diferentes bases de imagens, a comparação dos resultados obtidos fica difícil. Outro fator que dificulta a comparação é o desconhecimento de trabalhos que analisam somente a densidade dos tecidos das lesões, pois a totalidade dos trabalhos encontrados está relacionada à análise volumétrica no tempo. O único trabalho encontrado que utiliza a mesma base foi o de [Reeves et al. \(2009\)](#). Entretanto, o referido trabalho também realiza a análise volumétrica e encontra os valores de VVR .

Ao se comparar os valores obtidos de P_m com os valores de VVR obtidos em [Reeves et al. \(2009\)](#), verificam-se poucos casos divergentes. A divergência quanto ao aumento ou diminuição destes valores é melhor verificada em casos com grande variação volumétrica relativa. Entretanto, essa discordância não foi total, visto que os métodos utilizados apenas não indicaram tais casos como os de maior variação volumétrica relativa. Portanto, os percentuais de mudanças obtidos confirmam que houve significantes mudanças nas densidades da lesão, relativamente a essa escala de medida de mudança.

8.2 Comparação dos resultados obtidos

Cada um dos métodos propostos para a detecção das mudanças de densidade da lesão pulmonar possui uma característica inerente à sua formulação matemática na busca de capturar essas mudanças. Os métodos desenvolvidos foram: os baseados em *voxel*, que compreendem os métodos: 1) diferença entre imagens, 2) razão de imagens, 3) regressão de imagens, 4) operador XOR fuzzy e 5) mediana das imagens; os que utilizam o conceito de incerteza, como o teste de hipótese e a divergência de Jensen, que foram utilizados para analisar a lesão de forma local e global; o método que utiliza a análise de componentes

principais; e o método baseado em agrupamento, que utiliza o QT modificado. Portanto, realiza-se a comparação dos métodos propostos com base nos seus resultados obtidos.

A comparação dos métodos será realizada com base nos valores de percentual de mudança P_m e através dos resultados obtidos da classificação. Para facilitar a análise, todos os valores de P_m de todos os métodos foram colocados nas Tabelas 8.1 e 8.2. No intuito de melhorar a visualização do comportamento do P_m em função da variação do volume, um gráfico foi colocado nas Figuras 8.1 e 8.2. Dessa forma, pode-se visualizar quais foram os métodos mais sensíveis à detecção das mudanças de densidade da lesão, em qualquer faixa de variação de volume.

Tabela 8.1 – Percentuais de mudança de todos os métodos para a base privada de lesões.

Caso	P_m (%)								
	Diferença	Razão	Mediana	XOR	Regressão	T. de hip.	ACP	QT	VVR
1	12,84	17,71	12,84	19,04	13,12	7,50	12,90	20,31	-2,86
2	15,65	21,94	14,63	19,97	19,82	10,90	16,11	22,71	5,71
3	11,73	19,43	11,30	20,26	11,50	10,13	12,23	20,26	8,82
4	9,84	15,95	10,59	20,82	9,53	5,76	8,75	22,07	1,35
5	16,05	22,89	15,57	21,44	15,19	11,13	16,66	27,83	7,14
6	13,70	21,51	13,42	22,15	13,38	19,41	16,72	25,96	10,29
7	19,18	30,57	20,71	34,17	19,53	30,13	21,15	38,81	7,41
8	16,08	26,87	15,09	34,52	16,22	24,80	17,60	29,98	20,37
9	19,23	28,90	19,63	33,63	19,87	30,03	22,34	38,26	9,26
10	20,82	31,88	20,04	32,73	20,16	25,90	21,64	33,63	12,07
11	21,64	33,24	20,00	35,70	20,59	26,99	19,92	35,59	-9,23
12	17,26	27,36	16,19	32,77	18,60	21,35	22,59	33,95	1,72
13	24,25	28,76	21,71	24,70	28,76	27,30	25,40	24,38	12,50
14	33,89	37,83	33,98	37,10	25,18	35,43	32,92	36,06	-60,00
15	12,25	19,04	12,76	20,49	13,29	11,50	12,16	21,79	-3,51
16	16,80	21,11	17,06	19,89	14,20	24,61	17,37	19,98	-23,64
17	18,82	25,14	19,18	24,11	15,31	23,30	20,28	24,01	-26,32

Fonte: Autor

8.2.1 Quanto ao percentual de mudança

Novamente, foca-se na análise dos casos que possuem maior dificuldade para o especialista, isto é, os que possuem pouca variação volumétrica da base pública de nódulos pulmonares e privada de lesões em tratamento (Seção 3.3.1). Em cada uma das tabelas, existe uma linha em vermelho para realçar estes casos de pequeno valor de VVR .

Tabela 8.2 – Percentuais de mudança de todos os métodos para a base pública de lesões.

Caso	Pm (%)									
	Diferença	Razão	Mediana	XOR	Regressão	ACP	T. de hip.	QT	VVR	
1	21,92	33,75	22,56	31,19	19,65	22,28	40,01	29,92	-48,30	
2	29,45	36,57	30,23	33,68	33,30	31,90	39,96	33,90	-70,10	
3	18,96	31,04	20,15	36,61	16,95	22,85	37,45	26,57	-42,17	
4	8,24	11,23	7,87	8,01	5,57	8,26	13,24	8,37	133,60	
5	9,01	12,97	11,25	17,81	8,76	9,28	15,91	10,57	-29,09	
6	21,56	24,94	19,71	18,81	9,99	21,42	25,05	25,34	257,66	
7	20,02	21,28	16,83	16,62	7,14	20,00	21,13	21,35	281,26	
8	17,45	19,86	16,06	17,25	21,28	17,21	13,19	22,67	6,60	
9	17,62	21,98	16,18	24,76	24,03	18,24	25,10	30,41	22,48	
10	16,36	18,19	15,34	23,07	19,66	15,31	14,93	27,06	-1,61	
11	19,89	23,22	19,40	26,36	23,86	18,93	25,57	30,22	-19,67	
12	23,47	27,85	23,01	25,65	12,34	23,71	30,27	28,68	121,24	
13	12,85	19,76	12,58	12,74	12,03	15,17	26,52	21,95	75,19	
14	18,87	20,74	18,52	19,84	7,95	18,75	14,86	17,33	185,52	
15	17,92	20,35	17,53	17,41	8,89	17,93	23,23	21,52	162,29	
16	15,77	18,86	14,66	19,30	17,68	14,35	23,05	23,43	8,86	
17	4,12	4,63	3,99	3,89	4,93	3,76	3,84	6,22	-17,94	
18	18,60	21,01	18,53	18,29	6,51	18,69	23,81	21,28	275,08	
19	22,27	27,74	22,19	31,37	22,18	21,30	28,79	30,76	-34,20	
20	22,32	19,31	23,96	23,70	26,28	24,78	19,49	32,87	-13,72	
21	9,97	11,81	10,06	13,51	10,36	9,28	15,36	12,49	-45,58	
22	21,19	25,24	20,14	25,33	27,73	21,24	29,36	29,48	55,84	
23	17,46	20,83	16,51	23,23	22,92	17,16	26,77	26,06	32,52	
24	13,10	12,90	12,40	15,39	14,07	11,95	10,35	13,17	-4,77	
25	14,04	14,31	13,43	16,60	16,64	12,88	12,07	19,82	11,68	
26	12,78	13,51	12,14	14,77	14,92	12,51	16,80	17,14	-7,74	
27	15,33	15,47	14,27	17,06	17,24	14,59	14,61	19,31	17,27	
28	11,58	11,86	11,21	12,99	13,06	11,05	9,63	15,50	-3,12	
29	11,94	12,76	11,35	11,23	13,42	10,94	15,66	16,18	-17,39	
30	18,46	23,89	17,13	26,05	19,04	17,98	26,44	31,87	17,29	
31	9,52	17,56	8,87	9,37	10,53	8,74	15,24	12,08	-5,29	
32	5,84	10,16	5,74	5,49	5,62	5,65	6,93	10,55	4,55	
33	6,81	11,02	6,66	6,66	6,78	6,70	6,31	11,32	10,38	
34	10,43	12,26	10,42	10,93	13,78	10,50	11,74	14,49	-11,22	
35	15,93	17,98	15,16	15,55	12,48	16,25	16,35	19,78	30,35	
36	17,81	19,85	17,35	19,85	9,82	18,14	19,51	21,51	102,14	
37	12,35	14,19	11,95	16,32	10,76	12,42	16,17	15,79	46,83	
38	17,80	19,91	17,16	20,57	10,07	18,01	20,22	20,80	127,69	
39	12,16	15,28	11,61	13,11	12,50	12,65	20,11	17,35	55,08	
40	21,80	34,50	20,69	19,32	21,12	22,16	38,24	36,93	26,74	
41	17,82	20,05	17,09	19,94	9,30	17,94	22,44	21,03	95,76	
42	31,82	32,62	30,36	31,48	5,47	32,03	34,39	33,77	521,34	
43	8,94	11,83	8,71	7,37	9,02	8,39	14,56	13,66	-17,12	
44	12,54	13,69	11,51	19,26	18,13	27,08	21,36	21,09	26,81	
45	27,47	29,08	27,47	31,57	33,52	19,72	35,08	35,13	-46,43	
46	20,28	21,32	20,79	21,39	21,98	12,57	24,22	23,98	-57,75	

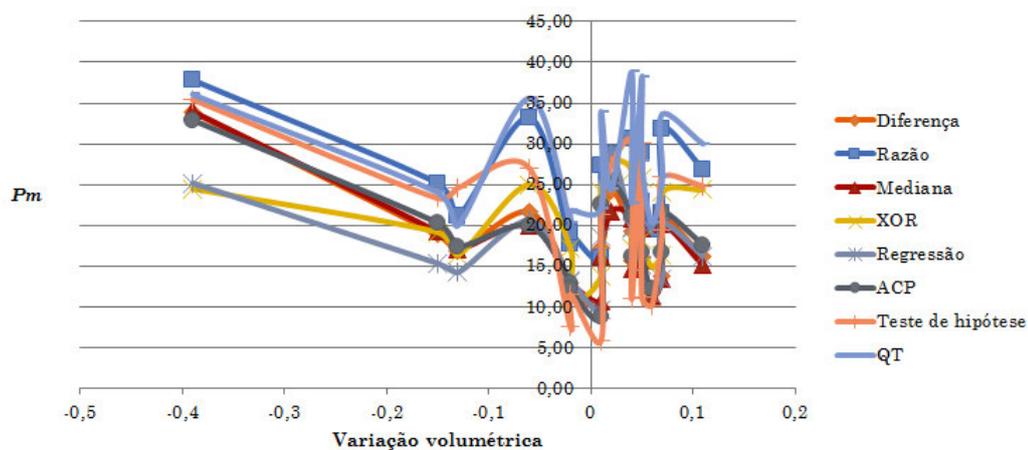
Fonte: Autor

8.2.1.1 Base privada

A Tabela 8.3 foi criada para melhor visualizar os métodos que foram mais e menos sensíveis às variações de densidade ao longo do tempo, com base nos valores de P_m da Tabela 8.1. Quanto maior for P_m , maior é a variação de densidade detectada ao longo do tempo. Os resultados apresentados na Tabela 8.1 foram resumidos na Tabela 8.4. A tabela possui os métodos que são mais e menos sensíveis em ordem de ocorrência sobre todos os casos. De acordo com as tabelas, o método QT modificado foi o mais sensível com relação a todos os casos analisados e o menos sensível foi o método do teste de hipótese. O segundo método mais sensível foi o método da razão de imagens e o segundo menos sensível foi o método da mediana. Logo, a partir das tabelas apresentadas, consegue-se visualizar os métodos mais e menos sensíveis aos casos da base.

Com relação aos casos de menor VVR , Casos 4 e 12, nos quais as linhas estão destacadas em vermelho, o mais sensível para ambos os casos foi novamente o QT modificado. O menos sensível para o Caso 4 foi o teste de hipótese. Da mesma forma, o método menos sensível para o Caso 12 foi a mediana. Em oposição a esses casos de pequeno VVR , o Caso 14 de maior VVR teve o método da razão como o mais sensível e o do operador XOR *fuzzy* como o menos sensível. Igualmente, a partir da curva da Figura 8.1, consegue-se visualizar a sensibilidade dos métodos em toda faixa de variação do volume.

A razão, o teste de hipótese e o QT modificado são os métodos mais sensíveis às mudanças para grandes variações do volume. Isso fica mais fácil de perceber a partir da curva do percentual de mudança em função da variação do volume na Figura 8.1. Portanto, os três métodos citados detectaram melhor as mudanças de densidade para os casos de maior variação volumétrica.

Figura 8.1 – Curva do P_m em função da variação volumétrica para a base privada.

Fonte: Autor

Tabela 8.3 – Métodos mais e menos sensíveis às variações de densidade para a base privada.

Caso	Método	
	Mais sensível	Menos sensível
1	QT	Teste de hipótese
2	QT	Teste de hipótese
3	QT	Teste de hipótese
4	QT	Teste de hipótese
5	QT	Teste de hipótese
6	QT	Regressão
7	QT	Diferença
8	QT	Mediana
9	QT	Diferença
10	QT	Mediana
11	QT	Mediana
12	QT	Mediana
13	Razão	Mediana
14	Razão	XOR
15	QT	XOR
16	Teste de hipótese	Regressão
17	Razão	Regressão

Fonte: Autor

Tabela 8.4 – Métodos ordenados pelo número de ocorrências dos casos da base privada.

Mais sensível			Menos sensível		
Posição	Metodologia	Frequência	Posição	Metodologia	Frequência
1º	QT	13	1º	Teste de hipótese	5
2º	Razão	3	1º	Mediana	5
3º	Teste de hipótese	1	2º	Regressão	3
4º	XOR	0	3º	XOR	2
4º	Mediana	0	3º	Diferença	2
4º	Diferença	0	4º	Razão	0
4º	ACP	0	4º	ACP	0
4º	Regressão	0	4º	QT	0

Fonte: Autor

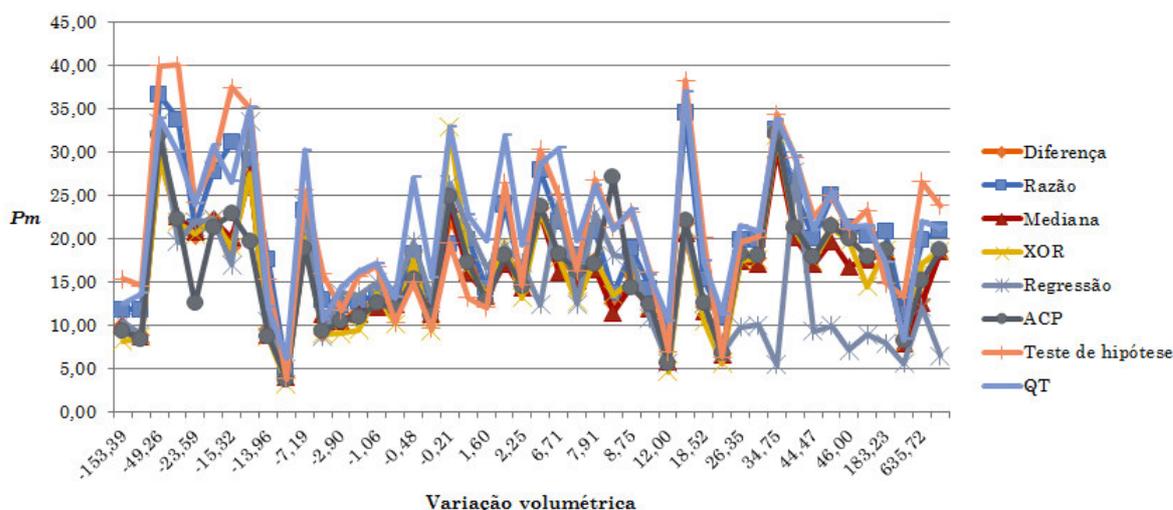
8.2.1.2 Base pública

Novamente, através da Tabela 8.5, derivada da Tabela 8.2, pode-se avaliar quais foram os métodos mais e menos sensíveis às variações de densidade, com base nos valores de P_m . Da tabela, percebe-se que o método QT modificado foi o mais sensível, seguido do método do teste de hipótese. Do mesmo modo, a Tabela 8.6 possui os mesmos resultados da Tabela 8.5, mas os mesmos estão em ordem de ocorrência. A partir dessa última tabela, visualiza-se que o menos sensível foi o método da regressão de imagens, seguido pelos métodos do operador XOR *fuzzy* e da mediana de imagem.

Os Casos 10 e 28 de menores valores de VVR tiveram o QT como o método mais sensível. Os métodos menos sensíveis foram diferentes para ambos os casos. O Caso 10 teve o teste de hipótese como o menos sensível e o Caso 28, o operador XOR-*fuzzy*. Em contraste aos casos apresentados, o mais sensível para o Caso 42 foi o teste de hipótese e o menos sensível, a regressão de imagens.

A sensibilidade dos métodos quanto a variação volumétrica pode ser vista na Figura 8.2. Da figura, percebe-se que os métodos da razão e do teste de hipótese também respondem bem às grandes variações do volume. Entretanto, o padrão de sensibilidade visto para os casos benignos é diferente para os casos malignos.

Figura 8.2 – Curva do P_m em função da variação volumétrica para a base pública.



Fonte: Autor

Tabela 8.5 – Métodos mais e menos sensíveis às variações de densidade para a base pública.

Caso	Método	
	Mais sensível	Menos sensível
1	Teste de hipótese	Regressão
2	Teste de hipótese	XOR
3	Teste de hipótese	Regressão
4	Teste de hipótese	Regressão
5	Teste de hipótese	Regressão
6	QT	Regressão
7	QT	Regressão
8	QT	Teste de hipótese
9	QT	Mediana
10	QT	Teste de hipótese
11	QT	ACP
12	Teste de hipótese	Regressão
13	Teste de hipótese	Regressão
14	Razão	Regressão
15	Teste de hipótese	Regressão
16	QT	ACP
17	QT	XOR
18	Teste de hipótese	Regressão
19	QT	ACP
20	XOR	Razão
21	Teste de hipótese	XOR
22	QT	Mediana
23	Teste de hipótese	Mediana
24	Regressão	XOR
25	QT	Teste de hipótese
26	QT	Mediana
27	QT	XOR
28	QT	XOR
29	QT	XOR
30	QT	Mediana
31	Razão	XOR
32	QT	XOR
33	QT	XOR
34	QT	XOR
35	QT	Regressão
36	QT	Regressão
37	Teste de hipótese	Regressão
38	QT	Regressão
39	Teste de hipótese	XOR
40	Teste de hipótese	Mediana
41	Teste de hipótese	Regressão
42	Teste de hipótese	Regressão
43	Teste de hipótese	ACP
44	ACP	Mediana
45	QT	ACP
46	Teste de hipótese	ACP

Tabela 8.6 – Métodos ordenados pelo número de ocorrências dos casos da base pública.

Mais sensível			Menos sensível		
Posição	Metodologia	Frequência	Posição	Metodologia	Frequência
1º	QT	23	1º	Regressão	17
2º	Teste de hipótese	18	2º	XOR	12
3º	Razão	2	3º	Mediana	7
4º	XOR	1	4º	ACP	6
4º	Regressão	1	5º	Teste de hipótese	3
4º	ACP	1	6º	Razão	1
5º	Diferença	0	7º	Diferença	0
5º	Mediana	0	7º	QT	0

Fonte: Autor

8.2.2 Quanto ao resultado da classificação

A Tabela 8.7 resume os resultados obtidos da classificação para todos os métodos desenvolvidos neste trabalho. A partir da tabela, considerando apenas o número de acertos ($vp + vn$), verifica-se que o método da diferença com e sem estimação dos parâmetros de classificação apresentou o melhor resultado. Semelhantemente, os métodos da mediana com estimação, regressão de imagens com e sem estimação e QT modificado com e sem estimação tiveram o mesmo desempenho na classificação.

O método da razão de imagens com a estimação dos parâmetros de classificação teve o segundo melhor resultado. O motivo do segundo melhor resultado da classificação está em possuir o maior número de verdadeiros positivos. Entretanto, o grande valor C indica que houve *overfitting* na estimação dos parâmetros do MVS. Portanto, conclui-se que o bom desempenho da classificação é devido ao *overfitting* do classificador e todos os outros abaixo do método da razão possuíram desempenho inferior ao mesmo.

Os métodos que tiveram os piores resultados de classificação com relação ao número de acertos ($vp + vn$) foram o teste de hipótese, a mediana, o operador XOR *fuzzy*, a razão sem estimação e o ACP.

Tabela 8.7 – Comparação dos resultados de classificação de todos os métodos desenvolvidos.

Método	Parâmetros do MVS			Métricas de avaliação					
	C	g	sens	esp	ac	vp	vn	fp	fn
Diferença	1	1/4	97,83	100,00	98,41	45	17	0	1
Diferença	0,5	0,0078125	97,83	100,00	98,41	45	17	0	1
Mediana	8	0,000488281	97,83	100,00	98,41	45	17	0	1
Regressão	1	1/4	97,83	100,00	98,41	45	17	0	1
Regressão	32	0,000488281	97,83	100,00	98,41	45	17	0	1
QT	1	1	97,83	100	98,41	45	17	0	1
QT	0,125	2	97,83	100	98,41	45	17	0	1
Razão	2048	0,00012207	100,00	94,12	98,41	46	16	1	0
Teste de hipótese	1	1/4	95,65	100,00	96,83	44	17	0	2
Teste de hipótese	2	0,0078125	95,65	100,00	96,83	44	17	0	2
Mediana	1	1/4	95,65	100,00	96,83	44	17	0	2
XOR	32	0,0078125	95,65	100,00	96,83	44	17	0	2
XOR	1	1/4	97,83	94,12	96,83	45	16	1	1
Razão	1	1/4	95,65	94,12	95,24	44	16	1	2
ACP	128	0,00012207	95,65	94,12	95,24	44	16	1	2
ACP	1	1/3	93,48	94,12	93,65	43	16	1	3
	Sem estimação			Com estimação					

Fonte: Autor

8.2.3 Conclusão

A partir da análise conjunta do P_m com a classificação, observa-se que o método mais sensível e com melhor resultado da classificação é o QT modificado. O método foi o mais sensível em ambas as bases de lesões e obteve a melhor acurácia de classificação, 97,83%. Na contramão desse resultado, estão os métodos da razão e do teste de hipótese. O método da razão foi o segundo mais sensível da base privada, mas possuiu o pior desempenho de classificação sem a estimação dos parâmetros e resultado intermediário com estimação. Da mesma forma, o método do teste de hipótese foi o segundo mais sensível e apresentou acurácia abaixo da maioria dos métodos propostos. Desse modo, o QT modificado é mais sensível às mudanças de densidade e essas mudanças indicadas pelo método são mais promissoras em determinar o diagnóstico da lesão em caso de indeterminação da lesão.

Os métodos menos sensíveis com maior frequência foram diferentes para ambas as bases de lesões. Para a base privada, os dois menos sensíveis foram o teste de hipótese e a mediana. Enquanto que para a base pública, os dois menos sensíveis em ordem de ocorrência foram os métodos da regressão e do operador XOR *fuzzy*. Logo, os resultados indicam que a baixa sensibilidade às mudanças detectadas pelos métodos é dependente da base de lesões utilizadas.

Outro fator que pode influenciar nos resultados encontrados é o fato de o registro ter sido realizado apenas com a lesão em um par de instantes analisados. Uma possível melhoria nos resultados pode ser encontrada com a aplicação do registro diretamente no volume de imagens completo do pulmão. Logo, um erro menor no registro pode ser obtido e a sensibilidade dos métodos às mudanças de densidade podem melhorar.

Mesmo que o desempenho dos métodos mais sensíveis não sejam os melhores quando comparados com os outros métodos desenvolvidos neste trabalho, a classificação realizada com as características de textura está muito próxima do 100% de acurácia. Portanto, o estudo de características de textura mais discriminativas podem ajudar a melhor classificar os casos malignos e benignos, auxiliando no diagnóstico e na progressão da doença ao longo do tempo.

9 Conclusão geral e atividades futuras

O presente capítulo possui a conclusão geral sobre o trabalho proposto e alguns dos trabalhos futuros que podem ser realizados nesta pesquisa. Em adição, alguns trabalhos em fase de submissão e aceitos são colocados.

9.1 Conclusão

O presente trabalho contém a proposta do desenvolvimento de um conjunto de métodos para auxiliar o especialista a identificar mudanças de densidade em lesões pulmonares ao longo do tempo. Verifica-se que os métodos propostos são mais úteis em casos de pequenas variações volumétricas, pois podem alertar o especialista às mudanças da lesão que antes eram imperceptíveis apenas visualizando os exames. Assim, para cada método proposto, foi encontrado um valor de P_m , que avalia quanto a lesão se modificou. Em seguida, uma análise da textura das regiões onde ocorreram as mudanças foi obtida, com o objetivo de validar o diagnóstico da lesão pulmonar. Essa última análise pode ser uma valiosa ferramenta no futuro, em casos onde a lesão é identificada inicialmente como indeterminada ou para verificar se uma lesão maligna em tratamento pode apresentar características benignas.

A partir da análise de P_m , conclui-se que, por mais que um método detecte extensas regiões da lesão como mudança, isso não é suficiente para um método obter boa acurácia de classificação. A exemplo disso, tem-se o método do teste de hipótese, que foi o segundo mais sensível às mudanças para a base pública de lesões. Da mesma forma, o método da razão é o mais sensível às mudanças para a base privada e apresentou um bom resultado de classificação. Em oposição a essa observação, tem-se o método QT modificado. O método foi o mais sensível às mudanças na maioria dos casos de ambas as bases de lesões. Logo, pode-se supor que as regiões identificadas pelo QT modificado produzem uma textura mais discriminante, quando comparado ao método do teste de hipótese ou da razão de imagens, que também são muito sensíveis às mudanças de densidade.

Devido ao pequeno número de casos malignos e benignos e por este trabalho estar em fase inicial, não se pode afirmar com convicção que algum método proposto por este trabalho seja o mais adequado para análise por um especialista, apesar de o QT modificado apresentar um bom resultado. Entretanto, percebe-se que as texturas capturadas pelas

técnicas clássicas podem ter forte influência no diagnóstico da lesão pulmonar e, portanto, clinicamente, essas regiões podem ser melhor avaliadas por um especialista. Logo, para que um dos métodos propostos se torne referência na avaliação da progressão de lesões pulmonares, necessita-se de um estudo mais amplo.

A partir dos resultados da classificação, conclui-se que todos os métodos possuem resultados promissores na predição do diagnóstico ao longo do tempo. Os resultados obtidos da classificação indicam que este trabalho pode ser uma valiosa ferramenta para auxiliar o especialista no prognóstico da lesão pulmonar.

9.2 Trabalhos futuros

Para trabalhos futuros, sugerem-se as seguintes atividades:

1. Ampliar a base de lesões pulmonares
2. Investigar o melhoramento do registro das lesões pulmonares. Os resultados obtidos por este trabalho têm forte influência do registro, pois se as imagens no par de instantes analisados não estiverem corretamente alinhadas, as diferenças encontradas serão resultado desse desalinhamento.
3. Detecção de mudanças:
 - a) Utilizar novos métodos de realce das mudanças nas densidades das lesões pulmonares, tal como a análise de componentes independentes.
 - b) Melhorar a última etapa de detecção das mudanças de densidade dos métodos propostos, para os seguintes métodos: método da razão, regressão, operador XOR *fuzzy*, ACP e teste de hipótese, que não detectam simultaneamente grandes e pequenas variações de densidade.
 - c) Melhorar o resultado da classificação para os métodos do teste de hipótese, do operador XOR *fuzzy*, da mediana, da razão e da ACP.
 - d) Utilizar um comitê de classificadores para avaliação do diagnóstico e das regiões de mudança.
 - e) Realizar um segundo realce das regiões detectadas como mudança pelos métodos propostos a partir de medidas com máxima variação de textura.

9.3 Artigos

- Aceitos

Autor Voxel-based comparative analysis of pulmonary lesions in CT for therapeutic purposes. *Jornal: Medical & Biological Engineering & Computing*. doi:10.1007/s11517-016-1510-0. 2016. p. 1-20.

Coautor Detecção de Anomalias em Termografia Dinâmica das Mamas. In: XVI Workshop de Informatica Medica, 2016, Porto Alegre. XXXVI Congresso da Sociedade Brasileira de Computação. Rio Grande do Sul: Sociedade Brasileira de Computação, 2016. p. 2609-2612.

Coautor Predição volumétrica de lesões pulmonares aplicando modelo oculto de Markov. *Journal of Health Informatics*. 2016.

- Submetidos

Autor Unsupervised Detection of Density Changes Through Principal Component Analysis for Lung Lesion Classification. Submetido ao *Jornal: Multimedia Tools and Applications*.

Autor Statistical tools for the Temporal Analysis of Pulmonary Lesions. Submetido ao *Jornal: Computers in Biology and Medicine*. Pequenas revisões.

Autor Modified Quality Threshold Clustering for Temporal Analysis and Classification of Lung Lesions. Submetido ao *Jornal: Computers in Biology and Medicine*.

Referências

- AERTS, H. J. W. L. et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*, v. 5, p. 4006, 2014. ISSN 2041-1723. Citado 2 vezes nas páginas 9 e 13.
- BAGLIO, P. T.; GAZZANA, M. B.; SILVA, D. R. Solitary pulmonary nodule. *Revista da Sociedade Brasileira de Clínica Médica*, 2009. Citado 2 vezes nas páginas 18 e 19.
- BAK, S. H. et al. Quantitative {CT} scanning analysis of pure ground-glass opacity nodules predicts further {CT} scanning change. *Chest*, v. 149, n. 1, p. 180 – 191, 2016. ISSN 0012-3692. Disponível em: <http://www.sciencedirect.com/science/article/pii/S001236921500080X>. Citado na página 10.
- BISHOP, C. *Pattern recognition and machine learning*. Secaucus, NJ: Springer-Verlag, 2006. Citado na página 37.
- BYRNE, M. J.; NOWAK, A. K. Modified recist criteria for assessment of response in malignant pleural mesothelioma. *Annals of Oncology*, v. 15, n. 2, p. 257–260, 2004. Disponível em: <http://annonc.oxfordjournals.org/content/15/2/257.abstract>. Citado na página 2.
- CHEN, B. et al. Automatic segmentation and identification of solitary pulmonary nodules on follow-up ct scans based on local intensity structure analysis and non-rigid image registration. In: GINNEKEN, R. M. S. B. van (Ed.). Lake Buena Vista, Florida: Proc. SPIE. Medical Imaging 2011: Computer-Aided Diagnosis, 2011. v. 7963, p. 79630B–79630B–10. Disponível em: <http://dx.doi.org/10.1117/12.878731>. Acesso em: 02 jan. 2015. Citado 2 vezes nas páginas 8 e 13.
- CHEN, Y.-W.; LIN, C.-J. Feature extraction: Foundations and applications. In: _____. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. cap. Combining SVMs with Various Feature Selection Strategies, p. 315–324. ISBN 978-3-540-35488-8. Disponível em: http://dx.doi.org/10.1007/978-3-540-35488-8_13. Citado 4 vezes nas páginas 84, 111, 126 e 140.
- CHU, A.; SEHGAL, C.; GREENLEAF, J. Use of gray value distribution of run lengths for texture analysis. *Pattern Recognition Letters*, v. 11, n. 6, p. 415 – 419, 1990. ISSN 0167-8655. Disponível em: <http://www.sciencedirect.com/science/article/pii/016786559090112F>. Acesso em: 01 jan. 2015. Citado na página 30.
- COROLLER, T. P. et al. Radiomic phenotype features predict pathological response in non-small cell lung cancer. *Radiotherapy and Oncology*, v. 119, n. 3, p. 480 – 486, 2016. ISSN 0167-8140. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0167814016310386>. Citado na página 10.
- CRISTIANINI, N.; SHAW-TAYLOR, J. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge: Cambridge University Press, 2000. Citado na página 36.
- DEGROOT, M.; SCHERVISH, M. *Probability and Statistics*. Pittsburgh, Pennsylvania: Addison-Wesley, 2002. (Addison-Wesley series in statistics). ISBN 9780201524888. Citado na página 93.

DUINDAM, T.; HOOP, B. de; GINNEKEN, B. van. Nodule volume change estimation in thoracic ct using sphere fitting, morphological segmentation and image registration. In: *2nd International Workshop on Pulmonary Image Analysis*. London, UK: The Pennsylvania State University, 2009. p. 389–400. Citado 3 vezes nas páginas 7, 11 e 12.

EDELIN, J. et al. Comparison of tumor response by response evaluation criteria in solid tumors (recist) and modified recist in patients treated with sorafenib for hepatocellular carcinoma. *Cancer*, Wiley Subscription Services, Inc., A Wiley Company, v. 118, n. 1, p. 147–156, 2012. ISSN 1097-0142. Disponível em: <http://dx.doi.org/10.1002/cncr.26255>. Citado na página 2.

EFRON, B.; TIBSHIRANI, R. *An Introduction to the Bootstrap*. New York: Taylor & Francis, 1994. (Chapman & Hall/CRC Monographs on Statistics & Applied Probability). ISBN 9780412042317. Citado na página 88.

EISENHAUER, E. et al. New response evaluation criteria in solid tumours: Revised {RECIST} guideline (version 1.1). *European Journal of Cancer*, v. 45, n. 2, p. 228 – 247, 2009. ISSN 0959-8049. Response assessment in solid tumours (RECIST): Version 1.1 and supporting papers. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0959804908008733>. Citado na página 2.

EL-BAZ, A. et al. Non-rigid registration techniques for automatic follow-up of lung nodules. *International Congress Series*, v. 1281, n. 0, p. 1115 – 1120, 2005. ISSN 0531-5131. CARS 2005: Computer Assisted Radiology and Surgery Proceedings of the 19th International Congress and Exhibition. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0531513105006011>. Acesso em: 01 jan. 2015. Citado 2 vezes nas páginas 5 e 12.

FIELD, J. K. et al. Prospects for population screening and diagnosis of lung cancer. *The Lancet*, v. 382, n. 9893, p. 732 – 741, 2013. ISSN 0140-6736. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0140673613616141>. Citado na página 1.

FILHO, A. O. de C. et al. Automatic detection of solitary lung nodules using quality threshold clustering, genetic algorithm and diversity index. *Artificial Intelligence in Medicine*, v. 60, n. 3, p. 165 – 177, 2014. ISSN 0933-3657. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0933365713001541>. Citado na página 129.

FOURNIER, L. et al. Imaging criteria for assessing tumour response: Recist, mrecist, cheson. *Diagnostic and Interventional Imaging*, v. 95, n. 7–8, p. 689 – 703, 2014. ISSN 2211-5684. Interventional radiology in oncology. Disponível em: <http://www.sciencedirect.com/science/article/pii/S2211568414001569>. Citado na página 2.

GASSIAT, E.; CLEYNEN, A.; ROBIN, S. Inference in finite state space non parametric hidden markov models and applications. *Statistics and Computing*, v. 26, n. 1, p. 61–71, 2016. ISSN 1573-1375. Disponível em: <http://dx.doi.org/10.1007/s11222-014-9523-8>. Citado na página 87.

GODOY, M. C. B.; NAIDICH, D. P. Subsolid pulmonary nodules and the spectrum of peripheral adenocarcinomas of the lung: Recommended interim guidelines for assessment

and management. *Radiology*, v. 253, n. 3, p. 606–622, 2009. PMID: 19952025. Disponível em: <http://dx.doi.org/10.1148/radiol.2533090179>. Acesso em: 01 jan. 2015. Citado na página 18.

GONÇALVES, A. R. *Otimização em ambientes dinâmicos com variáveis contínuas empregando algoritmos de estimação de distribuição*. Dissertação (Engenharia de Computação) — Faculdade de Engenharia Elétrica e de Computação, Campinas, SP, 2011. Citado na página 27.

HADJIISKI, L. M. et al. Computer-aided diagnosis for interval change analysis of lung nodule features in serial ct examinations. In: . Proc. SPIE. Medical Imaging 2007: Computer-Aided Diagnosis, 2007. v. 6514, p. 651411–651411–7. Disponível em: <http://dx.doi.org/10.1117/12.713770>. Acesso em: 01 jan. 2015. Citado 2 vezes nas páginas 6 e 12.

HARALICK, R.; SHANMUGAM, K.; DINSTEN, I. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, SMC-3, n. 6, p. 610–621, Nov 1973. ISSN 0018-9472. Citado na página 28.

HARRISON, L. C. et al. {MRI} texture analysis in multiple sclerosis: Toward a clinical analysis protocol. *Academic Radiology*, v. 17, n. 6, p. 696 – 707, 2010. ISSN 1076-6332. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1076633210000553>. Citado na página 2.

HAYKIN, S.; ENGEL, P. *Redes Neurais: Principios e Pratica*. Rio de Janeiro: Bookman, 2001. Citado na página 38.

HENSCHKE, C. I. et al. Lung cancers diagnosed at annual ct screening: Volume doubling times. *Radiology*, v. 263, n. 2, p. 578–583, 2012. PMID: 22454506. Disponível em: <http://pubs.rsna.org/doi/abs/10.1148/radiol.12102489>. Acesso em: 01 jan. 2015. Citado 2 vezes nas páginas 8 e 13.

HONDA, O. et al. Reproducibility of temporal volume change in ct of lung cancer: comparison of computer software and manual assessment. *The British Journal of Radiology*, v. 82, n. 981, p. 742–747, 2009. Disponível em: <http://dx.doi.org/10.1259/bjr/67746844>. Acesso em: 01 jan. 2015. Citado 3 vezes nas páginas 6, 11 e 12.

HOSHMAND, R. *Statistical Methods for Environmental and Agricultural Sciences*. New York: Taylor & Francis, 1997. (Applied statistics agriculture environmental science). ISBN 9780849331527. Citado na página 93.

IBANEZ, L. et al. *The ITK Software Guide*. First. New York, 2003. ISBN 1-930934-10-6. Disponível em: <http://www.itk.org/ItkSoftwareGuide.pdf>. Acesso em: 01 jan. 2015. Citado 2 vezes nas páginas 17 e 23.

İLSEVER, M.; ÜNSALAN, C. Pixel-based change detection methods. In: *Two-Dimensional Change Detection Methods*. Springer London, 2012, (SpringerBriefs in Computer Science). p. 7–21. ISBN 978-1-4471-4254-6. Disponível em: <http://dx.doi.org/10.1007/978-1-4471-4255-3-2>. Acesso em: 01 jan. 2015. Citado 3 vezes nas páginas 42, 47 e 50.

(INCA), I. N. de C. Câncer no Brasil: Dados dos registros de base populacional. 2011. Disponível em: <http://www1.inca.gov.br/cancernobrasil/2010/>. Acesso em: 01 jan. 2015. Citado na página 1.

(INCA), I. N. de C. Estimativa 2016: Incidência de câncer no Brasil. 2016. Disponível em: <http://www.inca.gov.br/estimativa/2016/index.asp?ID=2>. Acesso em: 10 agost. 2016. Citado na página 1.

JAFFE, C. C. Measures of Response: RECIST, WHO, and New Alternatives. *J Clin Oncol*, v. 24, n. 20, p. 3245–3251, jul. 2006. Disponível em: <http://dx.doi.org/10.1200/jco.2006.06.5599>. Citado na página 2.

JIANG, D.; TANG, C.; ZHANG, A. Cluster analysis for gene expression data: a survey. *IEEE Transactions on Knowledge and Data Engineering*, v. 16, n. 11, p. 1370–1386, Nov 2004. ISSN 1041-4347. Citado na página 129.

JUNIOR, G. B. et al. Classification of breast tissues using moran's index and geary's coefficient as texture signatures and svm. *Computers in Biology and Medicine*, Pergamon, v. 39, n. 12, p. 1063–1072, 2009. Citado na página 38.

KANUNGO, T. et al. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 24, p. 881–892, 2002. Citado na página 44.

KIM, T. et al. 3d texture analysis in renal cell carcinoma tissue image grading. *Comp. Math. Methods in Medicine*, v. 2014, p. 536217:1–536217:12, 2014. Disponível em: <http://dx.doi.org/10.1155/2014/536217>. Citado na página 2.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1995. (IJCAI'95), p. 1137–1143. ISBN 1-55860-363-8. Disponível em: <http://dl.acm.org/citation.cfm?id=1643031.1643047>. Citado 4 vezes nas páginas 84, 111, 126 e 141.

KYBIC, J. Bootstrap resampling for image registration uncertainty estimation without ground truth. *Image Processing, IEEE Transactions on*, v. 19, n. 1, p. 64–73, Jan 2010. ISSN 1057-7149. Citado na página 88.

LAI, D. et al. Independent component analysis applied to ultrasound speckle texture analysis and tissue characterization. In: HE", B. W. J. D. E. M. A. V. N. V. T. N. S. R. B. B. (Ed.). *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*. "Convention Center; Cité Internationale; Lyon, France": [s.n.], 2007. p. 6523–6526. ISSN 1557-170X. Citado na página 2.

LAY, D. *Linear Algebra and Its Applications*. Brasil: Pearson Education, 2002. ISBN 9788177583335. Citado 2 vezes nas páginas 114 e 115.

LENCIONI, R.; LLOVET, J. Modified recist (mrecist) assessment for hepatocellular carcinoma. *Seminars in Liver Disease*, Thieme Medical Publishers, v. 30, n. 1, p. 52–60, 2010. ISSN 0272-8087. Citado na página 2.

LI, Q. et al. Factors affecting uncertainty in lung nodule volume estimation with ct: comparisons of findings from two estimation methods in a phantom study. In: . Proc. SPIE. Medical Imaging 2015: Computer-Aided Diagnosis, 2015. v. 9414, p. 94140C–94140C–7. Disponível em: <http://dx.doi.org/10.1117/12.2081489>. Acesso em: 01 jan. 2015. Citado 2 vezes nas páginas 10 e 13.

LIBBY, D. M.; HENSCHKE, C. I.; YANKELEVITZ, D. F. The solitary pulmonary nodule: Update 1995. *The American Journal of Medicine*, v. 99, n. 5, p. 491 – 496, 1995. ISSN 0002-9343. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0002934399802253>. Citado 2 vezes nas páginas 18 e 19.

LINDELL, R. M. et al. Five-year lung cancer screening experience: Ct appearance, growth rate, location, and histologic features of 61 lung cancers. *Radiology*, v. 242, n. 2, p. 555–562, 2007. PMID: 17255425. Disponível em: <http://pubs.rsna.org/doi/abs/10.1148/radiol.2422052090>. Acesso em: 01 jan. 2015. Citado 2 vezes nas páginas 6 e 12.

LINNING, E. et al. Computed tomography quantitative analysis of components: a new method monitoring the growth of pulmonary nodule. *Acta Radiologica*, v. 54, n. 8, p. 904–908, 2013. Disponível em: <http://acr.sagepub.com/content/54/8/904.abstract>. Acesso em: 01 jan. 2015. Citado 4 vezes nas páginas 1, 8, 11 e 13.

LIU, Y. et al. The role of response evaluation criteria in solid tumour in anticancer treatment evaluation: Results of a survey in the oncology community. *European Journal of Cancer*, v. 50, n. 2, p. 260 – 266, 2014. ISSN 0959-8049. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0959804913009490>. Citado na página 2.

MARTINS, G. de A. *Estatística geral e aplicada*. São Paulo: Atlas, 2005. ISBN 9788522441723. Citado 2 vezes nas páginas 87 e 90.

MCNITT-GRAY, M. F. et al. Determining the variability of lesion size measurements from ct patient data sets acquired under “no change” conditions. *Translational Oncology*, v. 8, n. 1, p. 55 – 64, 2015. ISSN 1936-5233. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1936523315000029>. Acesso em: 01 jan. 2015. Citado 2 vezes nas páginas 9 e 13.

MOONEY, C.; DUVAL, R.; DUVAL, R. *Bootstrapping: A Nonparametric Approach to Statistical Inference*. United States: SAGE Publications, 1993. (Bootstrapping: A Nonparametric Approach to Statistical Inference, N° 94-95). ISBN 9780803953819. Citado na página 88.

NAGARAJAN, M. et al. Volumetric quantitative characterization of human patellar cartilage with topological and geometrical features on phase-contrast x-ray computed tomography. *Medical & Biological Engineering & Computing*, Springer Berlin Heidelberg, v. 53, n. 11, p. 1211–1220, 2015. ISSN 0140-0118. Disponível em: <http://dx.doi.org/10.1007/s11517-015-1340-5>. Acesso em: 01 jan. 2015. Citado 2 vezes nas páginas 9 e 13.

(NCI), N. C. I. *SEER Cancer Statistics Factsheets: Lung and Bronchus Cancer*. Bethesda, MD, 2011. Disponível em: <http://seer.cancer.gov/statfacts/html/lungb.html>. Acesso em: 01 jan. 2015. Citado na página 1.

OLIVEIRA, R. K. F. de et al. Nódulo pulmonar solitário parcialmente sólido. Casos Clínicos. 2010. Disponível em: http://www.sbpt.org.br/downloads/arquivos/Casos_Clinicos_PDF/CASO_31_NODULO_PULMONAR_SOLITARIO_PARCIALMENTE_SOLIDO.pdf.

Acesso em: 01 jan. 2015. Citado 2 vezes nas páginas 18 e 19.

OTSU, N. A Threshold Selection Method from Gray-level Histograms. *IEEE Transactions on Systems, Man and Cybernetics*, v. 9, n. 1, p. 62–66, 1979. Disponível em: <http://dx.doi.org/10.1109/TSMC.1979.4310076>. Acesso em: 01 jan. 2015. Citado na página 43.

PANDAV, R. S.; KHOT, S. R. Registration of ct and mr images in image fusion. *International Journal of Electrical, Electronics and Computer Systems*, v. 3, 2015. Citado na página 22.

PURI, M. et al. Texture analysis of foot sole soft tissue images in diabetic neuropathy using wavelet transform. *Medical and Biological Engineering and Computing*, Springer-Verlag, v. 43, n. 6, p. 756–763, 2005. ISSN 0140-0118. Disponível em: <http://dx.doi.org/10.1007/BF02430954>. Citado na página 2.

REEVES, A. et al. A public image database to support research in computer aided diagnosis. In: *Engineering in Medicine and Biology Society, 2009. EMBC 2009*. Minneapolis, MN: Annual International Conference of the IEEE, 2009. p. 3715–3718. ISSN 1557-170X. Citado 4 vezes nas páginas 19, 20, 61 e 144.

REEVES, A. et al. On measuring the change in size of pulmonary nodules. *Medical Imaging, IEEE Transactions on*, v. 25, n. 4, p. 435–450, April 2006. ISSN 0278-0062. Citado 3 vezes nas páginas 5, 11 e 12.

RENNÓ, C. D. *Estatística: Aplicação ao Sensoriamento Remoto*. 1º. ed. São Paulo, Brasil, 2015. Disponível em: ["http://www.dpi.inpe.br/~camilo/estatistica"](http://www.dpi.inpe.br/~camilo/estatistica). Acesso em: 01 jan. 2015. Citado na página 89.

RINGEL, E. *The Little Black Book of Pulmonary Medicine*. Ontario, Canada: Jones and Bartlett Publishers, 2008. (Jones and Bartlett little black book series). ISBN 9780763752446. Citado na página 19.

Sá, J. P. M. d. *Pattern recognition : concepts, methods and applications*. Berlin, Heidelberg, New York: Springer, 2001. ISBN 3-540-42297-8. Disponível em: <http://opac.inria.fr/record=b1102840>. Citado 2 vezes nas páginas 34 e 35.

SHEN, L. et al. Spatio-temporal analysis tool for modeling pulmonary nodules in mr images. In: . San Diego, CA: Proc. SPIE 6141, Medical Imaging 2006: Visualization, Image-Guided Procedures, and Display, 2006. v. 6141, p. 61412I–61412I–10. Disponível em: <http://dx.doi.org/10.1117/12.654469>. Acesso em: 01 jan. 2015. Citado 2 vezes nas páginas 5 e 12.

SILVA, A. C. *Algoritmos para Diagnostico Assistido de Nodulos Pulmonares Solitarios em Imagens de Tomografia Computadorizada*. Tese (Doutorado) — PUC-Rio, jan. 2004. Citado na página 18.

SOARES, F. A. P. *Produção de raios x em ampolas radiográficas: Estudo do tomógrafo computadorizado do Hospital regional de São José/SC*. Tese (Física Experimental) —

Universidade Federal de Santa Catarina, Florianópolis, 2006. Citado 3 vezes nas páginas 14, 15 e 16.

SOFKA, M.; STEWART, C. V. Location registration and recognition (lrr) for serial analysis of nodules in lung ct scans. *Medical Image Analysis*, v. 14, n. 3, p. 407 – 428, 2010. ISSN 1361-8415. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1361841510000216>. Acesso em: 01 jan. 2015. Citado 2 vezes nas páginas 7 e 12.

SONKA, M.; FITZPATRICK, M. *Handbook of medical imaging. Volume 2. , Medical image processing and analysis*. Bellingham, Washington, USA: SPIE Press, 2000. ISBN 0-8194-3622-4. Citado na página 24.

SONKA, M.; HLAVAC, V.; BOYLE, R. *Image Processing, Analysis and Machine Vision*. New York: PWS Publishing, 1998. Citado 2 vezes nas páginas 27 e 28.

STEWART, C.; IBANEZ, L. *Registration components*. [S.l.], 2004. Disponível em: www.cs.rpi.edu/courses/spring04/imagereg/lecture08.ppt. Acesso em: 01 jan. 2015. Citado na página 26.

SUGAR, C. A.; JAMES, G. M. Finding the number of clusters in a dataset. *Journal of the American Statistical Association*, v. 98, n. 463, p. 750–763, 2003. Disponível em: <http://dx.doi.org/10.1198/016214503000000666>. Citado na página 130.

TANEJA, I. J. *Generalized Information Measures and Their Applications*. second. www.mtm.ufsc.br/~taneja/book/book.html: online, 2001. v. 1. Acesso em: 01 jan. 2015. Citado na página 92.

TAO, C. et al. Automated Matching of Pulmonary Nodules: Evaluation in Serial Screening Chest CT. *American Journal of Roentgenology*, v. 192, p. 624–628, mar. 2009. Disponível em: <http://www.ajronline.org/cgi/reprint/192/3/624>. Acesso em: 02 jan. 2015. Citado 2 vezes nas páginas 7 e 12.

TEAM, T. N. L. S. T. R. Reduced lung-cancer mortality with low-dose computed tomographic screening. *The New England Journal of Medicine*, 2011. Disponível em: <http://dx.doi.org/10.1056/NEJMoa1102873>. Citado na página 1.

VAPNIK, V. *Statistical Learning Theory*. New York: Wiley New York, 1998. Citado 2 vezes nas páginas 36 e 141.

WANG, B.; WERTELECKI, W. Density estimation for data with rounding errors. *Computational Statistics Data Analysis*, v. 65, p. 4 – 12, 2013. ISSN 0167-9473. Special issue on Robust Analysis of Complex Data. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0167947312000989>. Citado na página 87.

WASSERMAN, L. *All of statistics : a concise course in statistical inference*. New York: Springer, 2010. ISBN 9781441923226 1441923225. Citado 2 vezes nas páginas 88 e 94.

XU, R.; WUNSCH, D. Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, v. 16, n. 3, p. 645–678, May 2005. ISSN 1045-9227. Citado na página 129.

XUE, Z.; WONG, K.; WONG, S. T. Joint registration and segmentation of serial lung ct images for image-guided lung cancer diagnosis and therapy. *Computerized Medical Imaging and Graphics*, v. 34, n. 1, p. 55 – 60, 2010. ISSN 0895-6111. Image-Guided Surgical Planning and Therapy. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0895611109000652>. Acesso em: 02 jan. 2015. Citado 2 vezes nas páginas 8 e 13.

ZHANG, L. et al. Internal growth of nonsolid lung nodules: radiologic-pathologic correlation. *Radiology*, v. 263, n. 1, p. 279–86, 2012. ISSN 1527-1315. Disponível em: <http://www.biomedsearch.com/nih/Internal-growth-nonsolid-lung-nodules/22438449.html>. Citado na página 1.

ZHANG, Y. Mri texture analysis in multiple sclerosis. *Int. J. Biomedical Imaging*, v. 2012, p. 762804:1–762804:7, 2012. Disponível em: <http://dblp.uni-trier.de/db/journals/ijbi/ijbi2012.html#Zhang12>. Citado na página 2.

ZHANG, Y. et al. A novel MRI texture analysis of demyelination and inflammation in relapsing-remitting experimental allergic encephalomyelitis. In: LARSEN, R.; NIELSEN, M.; SPORRING., J. (Ed.). *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2006, 9th International Conference, Copenhagen, Denmark, October 1-6, 2006, Proceedings, Part I*. [s.n.], 2006. p. 760–767. Disponível em: http://dx.doi.org/10.1007/11866565_93. Citado na página 2.

ZHENG, Y. et al. Lung nodule growth analysis from 3d ct data with a coupled segmentation and registration framework. In: *Computer Vision, 2007. ICCV 2007*. Rio de Janeiro, Brazil: IEEE 11th International Conference on, 2007. p. 1–8. ISSN 1550-5499. Citado 3 vezes nas páginas 6, 11 e 12.