

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE ENGENHARIA ELÉTRICA

ANA PAULA PASTRE FRONER

**CARACTERIZAÇÃO DE NÓDULOS PULMONARES EM IMAGENS DE TOMOGRAFIA
COMPUTADORIZADA PARA FINS DE AUXÍLIO AO DIAGNÓSTICO**

Porto Alegre
2015

ANA PAULA PASTRE FRONER

**CARACTERIZAÇÃO DE NÓDULOS PULMONARES EM IMAGENS DE TOMOGRAFIA
COMPUTADORIZADA PARA FINS DE AUXÍLIO AO DIAGNÓSTICO**

Dissertação apresentada como requisito para a obtenção do grau de Mestre pelo Programa de Pós-Graduação em Engenharia Elétrica da Pontifícia Universidade Católica do Rio Grande do Sul.

Área de Concentração: Sinais, Sistemas e Tecnologia da Informação

Linha de Pesquisa: Engenharia Biomédica

Orientadora: Profa. Dra. Ana Maria Marques da Silva

Porto Alegre
2015



CARACTERIZAÇÃO DE NÓDULOS PULMONARES EM IMAGENS DE TOMOGRAFIA COMPUTADORIZADA PARA FINS DE AUXÍLIO AO DIAGNÓSTICO


CANDIDATA: ANA PAULA PASTRE FRONER

Esta Dissertação de Mestrado foi julgada para obtenção do título de MESTRE EM ENGENHARIA ELÉTRICA e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia Elétrica da Pontifícia Universidade Católica do Rio Grande do Sul.



DRA. ANA MARIA MARQUES DA SILVA - ORIENTADORA

BANCA EXAMINADORA



DRA. ANA CLÁUDIA PATROCÍNIO - PÓS-GRADUAÇÃO EM ENGENHARIA
BIOMÉDICA - UFU



DR. ALEXANDRE ROSA FRANCO - DO PPGE/FENG - PUCRS

AGRADECIMENTOS

A professora Doutora Ana Maria Marques da Silva pela sua orientação e oportunidade de aprendizado.

Aos colegas de mestrado pelo companheirismo, paciência e disponibilidade.

A Jerusa Dalbosco Finatto, ao Mozart Guedes e a Caroline Dartora pelo auxílio e dedicação em minha dissertação.

Ao Professor Doutor João Feliz pelo auxílio em toda análise estatística que foi necessária para esta tese de dissertação.

RESUMO

Atributos morfológicos, de intensidade e textura são essenciais para o auxílio ao diagnóstico de nódulos em imagens de tomografia computadorizada de pulmão. Para melhorar a acurácia do diagnóstico, assim como a interpretação da imagem radiológica, são utilizados sistemas de diagnóstico auxiliado por computador, que ajudam a diminuir a variabilidade de interpretação. O objetivo geral deste estudo foi avaliar a utilização de dados de pacientes e atributos quantitativos de nódulos pulmonares em imagens de tomografia computadorizada de pulmão para a construção de um modelo de classificação em termos de malignidade. O estudo foi baseado em uma análise das imagens de tomografia computadorizada de pulmão de 51 pacientes, sendo 33 pacientes diagnosticados com lesões malignas e 18 pacientes com lesões benignas, todos confirmados por meio do laudo anatomopatológico do tecido pulmonar. O estudo contou com uma etapa de interpretação diagnóstica realizada a partir de um estudo cego com médicos radiologistas e a utilização da regressão logística para análise do poder preditivo de variáveis qualitativas e quantitativas extraídas das imagens de tomografia computadorizada de pulmão. Os atributos quantitativos dos nódulos foram constituídos de: doze atributos morfológicos (volume, área, perímetro, compacidade, irregularidade e os momentos invariantes de ordem 1 a 7), cinco atributos de intensidade (massa, densidade, média do número TC, índice de gordura e índice de calcificação) e três atributos de textura (homogeneidade, entropia e variância). Os resultados mostraram que na interpretação visual dos radiologistas, apenas o médico com maior experiência apresentou uma concordância com o laudo anatomopatológico (64,5%), quando se excluem os acertos devido ao acaso. Concluiu-se que o modelo que melhor classifica a malignidade dos nódulos incluiu um atributo qualitativo de localização, um atributo de intensidade relacionado ao índice de calcificação, desenvolvido neste trabalho, e um atributo morfológico relacionado à compacidade do nódulo. O valor preditivo do modelo (86,3%) foi muito maior do que o valor preditivo baseado na avaliação visual do médico mais experiente (65,3%), sem considerar o acaso, estando próximo da acurácia desse mesmo médico (85,1%), incluído o acerto devido ao acaso. Este fato revela a importância deste tipo de pesquisa na busca de um modelo quantitativo, baseado em imagens de tomografia computadorizada, que permita a classificação dos nódulos pulmonares.

Palavras-chave: Tomografia Computadorizada. Atributos quantitativos. Diagnóstico Auxiliado por Computador. Nódulos Pulmonares.

ABSTRACT

Morphological attributes, intensity and texture, are essential to aid the diagnosis of pulmonary nodules. To improve the accuracy of diagnosis, as well as the interpretation of radiological imaging, computer-aided diagnostic systems are used, which help to reduce the interpretation variability. The aim of this study was to evaluate the use of patient data and quantitative attributes of pulmonary nodules in lung computed tomography to build a classification model in terms of malignancy. The study was based on the analysis of 51 patients computed tomography images of lung, 33 patients diagnosed with malignant lesions and 18 patients with benign lesions, all confirmed through anatomic-pathological report of lung tissue. The study included a diagnostic interpretation stage made from a blind study with radiologists and the use of logistic regression to analyze the predictive power of qualitative and quantitative variables extracted from computed tomography lung images. The quantitative attributes of the nodules consisted of twelve morphological attributes (volume, area, perimeter, compactness, roughness and invariant moments of order 1 to 7), five intensity attributes (mass, density, CT number average, fat and calcification indexes) and three texture attributes (homogeneity, entropy and variance). The results showed that in the visual interpretation of radiologists, only the most experienced doctor showed a correlation with the pathology report (64.5%), when excluding the hits due to chance. The model which better predicts the nodules malignancy included a quantitative attribute of localization, the intensity attribute 'calcification index' developed in this work, and the compactness, a morphological attribute related to the nodule form. The predictive value of the classification model (86.3%) was much higher than the predictive value based on the visual assessment of the most experienced doctor (65.3%), regardless of the chance, being close to the accuracy of the same doctor (85.1%). This emphasizes the importance of this type of research in the search for a quantitative model that allows the classification of pulmonary nodules.

Keywords: Computed Tomography. Quantitative Features. Computer-Aided Diagnosis. Pulmonary nodules.

LISTA DE FIGURAS

Figura 1 - Exemplo de imagem de TC a qual identifica-se diagnóstico de nódulo em vidro fosco	6
Figura 2 - Exemplo de imagem de TC a qual identifica-se diagnóstico de nódulo benigno	11
Figura 3 - Exemplo de imagem de TC a qual identifica-se diagnóstico de hamartoma.....	11
Figura 4 - Sete tipos de bordas/margens de nódulos pulmonares.....	13
Figura 5 - Exemplo de imagem de TC, na qual identificam-se os valores de uma região de interesse em termos de unidades Hounsfield	16
Figura 6 - Exemplo de imagem de TC a qual identifica-se diagnóstico de hamartoma.....	17
Figura 7 - Padrões de calcificação benignos	18
Figura 8 - Exemplo de imagem de TC a qual identifica-se um nódulo de calcificação difusa.....	19
Figura 9 - Ângulos para o cálculo da matriz de concorrência	27
Figura 10 - (a) Exemplo de imagem na qual identifica-se uma ROI no centro de um nódulo. (b) Plugin <i>Segmenting Assistant</i> do ImageJ®. (c) Imagem segmentada.	43
Figura 11 - Diagrama de funcionamento do programa no MATLAB®.....	45
Figura 12 - Gráficos da média e desvio padrão das variáveis quantitativas: (a) volume; (b) massa; (c) compacidade; (d) IG; (e) IC; (f) irregularidade.	57
Figura 13 - Curva ROC do modelo de regressão logística proposto neste trabalho.....	61

LISTA DE QUADROS

Quadro 1. Características que diferenciam um Nódulo Pulmonar Solitário e Massa	7
Quadro 2. Características das margens de um nódulo pulmonar solitário	13
Quadro 3. Características Radiológicas sugestivas de benignidade ou malignidade de nódulos pulmonares solitários	20
Quadro 4. Atributos quantitativos em nódulos pulmonares, analisados por diferentes autores.....	35
Quadro 5. Atributos analisados com vistas à construção de um modelo de classificação dos nódulos pulmonares.	47
Quadro 6. Classificação dos atributos pelos radiologistas de acordo com o grau de importância para a classificação dos nódulos	51
Quadro 7. Número e percentual de casos válidos e não válidos para cada médico avaliador.	52
Quadro 8. Coeficientes Kappa e acurácia para os quatro médicos radiologistas.	53
Quadro 9. Número de pacientes por faixa etária, para os nódulos benignos e malignos.	54
Quadro 10. Número de pacientes por sexo, para os nódulos benignos e malignos.....	54
Quadro 11. Número de pacientes por localização do nódulo nos lóbulos pulmonares, para os nódulos benignos e malignos.	55
Quadro 12. Variáveis independentes bidimensionais analisadas pelo teste de Kruskal-Wallis para verificar a diferença significativa na inclusão de um mesmo atributo avaliado em diferentes cortes.	56
Quadro 13. Variáveis independentes analisadas no modelo de regressão logística. As variáveis com $p < 0,25$ foram incluídas no modelo de regressão logística multivariada (HOSNER & LEMERSHOW, 2000, p. 95).	58
Quadro 14 - Valores dos coeficientes estimados (b_i) nos modelos avaliados pela regressão logística multivariada neste trabalho, com os respectivos atributos e seus níveis de significância.....	60
Quadro 15. Estatística da área da curva ROC do modelo proposto neste trabalho, seu desvio padrão, nível de significância e intervalo de confiança.....	62

LISTA DE SIGLAS

CAD (*Computed Aided Diagnosis*) – Diagnóstico Auxiliado por Computador

HU – Unidades Hounsfield

NVF – Nódulo em Vidro Fosco

NPS – Nódulo Pulmonar Solitário

ROC – *Receiver-Operating Characteristics*

TC – Tomografia Computadorizada

SUMÁRIO

1	INTRODUÇÃO	1
2	OBJETIVOS	2
2.1	Objetivo Geral.....	2
2.2	Objetivos Específicos.....	3
3	FUNDAMENTAÇÃO TEÓRICA.....	4
3.1	Doenças Pulmonares	4
3.2	Tipos de Nódulos Pulmonares	5
3.2.1	Nódulo em Vidro Fosco	5
3.2.2	Nódulo Solitário	7
3.3	Nódulos Benignos e Malignos	9
3.4	Atributos da Imagem Radiológica	12
3.4.1	Atributos Morfológicos	12
3.4.2	Atributos de Intensidade.....	14
3.4.3	Atributos de Textura.....	20
3.5	Computed Aided Diagnosis – CAD.....	21
3.5.1	Detecção de Nódulos Pulmonares em Imagens Digitais.....	22
3.5.2	Quantificação de Atributos em Imagens Digitais	22
3.6	Revisão da Literatura: Análise dos Atributos em imagens de TC de pulmão	28
3.7	Classificadores: redes neurais, análise discriminante e regressão logística.....	36
4	METODOLOGIA.....	38
4.1	Seleção das Imagens	38
4.2	Atributos Qualitativos na avaliação do diagnóstico.....	39
4.3	Interpretação das Imagens pelos Radiologistas	39
4.4	Análise do Desempenho do Modelo na Classificação dos Nódulos	40
4.5	Segmentação dos Nódulos	41
4.6	Extração dos Atributos Quantitativos	42
4.6.1	Atributos Morfológicos	42
4.6.2	Atributos de Intensidade.....	45
4.6.3	Atributos de Textura.....	46
4.7	Construção de um Modelo para Classificação dos Nódulos.....	46
5	RESULTADOS E DISCUSSÃO.....	50
5.1	Atributos Relevantes para o Diagnóstico Médico.....	50
5.2	Interpretação das Imagens pelos Radiologistas	52
5.3	Perfil da Amostra	53
5.4	Análise dos Atributos Quantitativos das Imagens para o Modelo de Classificação....	55
6	CONCLUSÕES.....	63
7	REFERÊNCIAS.....	65
	ANEXO A – Análise de Regressão Logística Univariada de cada Atributo	71

1 INTRODUÇÃO

A detecção precoce do câncer de pulmão é importante para melhorar o prognóstico do paciente. O exame de tomografia computadorizada ou CT (do inglês, *Computed Tomography*) é utilizado no rastreamento do câncer, pois possui resolução de alto contraste em relação à radiografia de tórax e, portanto, fornece uma melhor detecção de lesões pulmonares (HAESEGAWA et al, 2000).

A ampliação do uso da técnica de CT na investigação clínica para o diagnóstico de câncer pulmonar vem resultando em achados usualmente não encontrados em radiografias convencionais, como pequenos nódulos (KANEKO et al,1996; MACMAHON et AL, 2005). Nódulos pulmonares são lesões bem definidas, discretas, menores do que 3 cm de diâmetro. Para lesões com diâmetro maior do que 3 cm, utiliza-se o termo massa (MULLER et al, 2008).

Um dos tipos de nódulos encontrados em pulmões é o nódulo em vidro fosco (NVF), uma área circunscrita de alta densidade, na qual ocorre um aumento da atenuação aos raios X, com preservação das margens brônquicas e vasculares (KANEKO et al,1996). Apesar dos NVF poderem ser associados a uma variedade de doenças, tais como inflamações, fibroses e hemorragias, sua presença persistente, sem mudanças de tamanho ou opacidade ao longo de um período de seguimento de 3 meses ou mais, sugere a possibilidade de um câncer pulmonar em estágio inicial (LEE et al, 2010).

A alta resolução e o crescente avanço tecnológico dos equipamentos de tomografia computadorizada permitem visualizar detalhes anatômicos dos pulmões e de pacientes com suspeita de doenças no parênquima pulmonar, contudo, é uma modalidade de difícil avaliação devido ao aumento da informação anatômica gerada pelo equipamento (SLUIMER et al, 2003).

Métodos de auxílio computadorizado ao diagnóstico ou CAD (do inglês *Computer-Aided Diagnosis*) vem sendo desenvolvidos para auxiliar radiologistas na detecção e classificação dos nódulos pulmonares (IWANO S., et al, 2005; WAY et al., 2009; EL-BAZ & SURI, 2012).

O CAD pode ser definido como um diagnóstico feito por um radiologista que utiliza o resultado de análises quantitativas automatizadas de imagens radiográficas como uma “segunda opinião”, para a tomada de decisões diagnósticas (AZEVEDO-MARQUES, 2001).

Para o desenvolvimento de métodos de auxílio ao diagnóstico, baseados na extração, cálculo de atributos e classificação, é necessário que inicialmente sejam definidos quais

atributos são clinicamente significativos para, a seguir, estabelecer os métodos para sua extração e quantificação (AZEVEDO-MARQUES, 2001).

No caso de nódulos pulmonares, os estudos utilizam usualmente atributos de forma, textura e dados demográficos, tais como a idade e se o paciente é ou foi fumante (IWANO S., et al, 2005; WAY et al., 2009; EL-BAZ & SURI, 2012).

No entanto, pesquisadores vêm investigando o diagnóstico computadorizado de lesões pulmonares em imagens de tomografia computadorizada, que reúnem diversas questões abertas e desafios computacionais, dentre os quais, o desenvolvimento de métodos de detecção de estruturas de interesse através de técnicas de segmentação (GIGER et al, 1994; TOSHIOKAET al,1997; KAWATA et al, 1997;), a análise dos atributos das imagens (KAWATA et al, 2000) e o seu uso nas técnicas de classificação (REEVES et al, 2000). O esforço de grupos de pesquisa vem resultando em sistemas completos de detecção e diagnóstico (ARIMRA et al, 2004; WAY et al, 2008).

No entanto, nos trabalhos desenvolvidos, observa-se que a escolha dos atributos que melhor classificam nódulos pulmonares com vistas à uma avaliação de sua malignidade, ainda é um problema aberto, não havendo consenso em relação àqueles que devem ser utilizados.

Este trabalho se propõe a investigar distintos atributos quantitativos de imagens de tomografia computadorizada de nódulos pulmonares benignos e malignos e dados dos pacientes, buscando sua relevância para o diagnóstico médico por imagem. Os atributos quantitativos, obtidos a partir de um banco de imagens com pacientes que foram para a biópsia, possuem o diagnóstico confirmado por meio da análise anatomopatológica do tecido pulmonar. Este conjunto de imagens, portanto, possui sua classificação de malignidade e benignidade considerada verdadeira (padrão ouro), permitindo que a pesquisa seja voltada para a investigação do modelo de classificação a partir dos atributos das imagens e suas limitações para uma classificação mais próxima àquela obtida pela biópsia.

2 OBJETIVOS

2.1 *Objetivo Geral*

O objetivo geral deste estudo é caracterizar nódulos pulmonares em imagens de tomografia computadorizada de pulmão, buscando a sua classificação em termos de sua malignidade.

2.2 *Objetivos Específicos*

Para tanto, pretende-se desenvolver o trabalho através da busca pelos seguintes objetivos específicos:

- Identificar em conjunto com uma equipe de radiologistas, as características qualitativas e quantitativas utilizadas para interpretação diagnóstica de nódulos pulmonares em imagens de tomografia computadorizada;
- Determinar os atributos quantitativos morfológicos, intensidade e textura, de um banco de imagens de nódulos pulmonares em imagens de tomografia computadorizada, relacionados aos atributos identificados pelos radiologistas;
- Desenvolver um modelo de classificação de nódulos pulmonares em imagens de tomografia computadorizada com base na caracterização dos atributos, comparando o seu desempenho com os resultados obtidos por médicos radiologistas e o padrão ouro.

3 FUNDAMENTAÇÃO TEÓRICA

3.1 Doenças Pulmonares

A partir da década de 40, com o surgimento contínuo de recursos terapêuticos específicos, o exercício do diagnóstico passou a ser indispensável para a prática clínica, pois as inúmeras opções de tratamento necessitam da exata identificação das doenças. A tuberculose e outras doenças infecciosas já podiam ser curadas, e fazia-se necessário um esquema terapêutico ajustado ao paciente. Já não bastava saber se o paciente apresentava um quadro infeccioso; era necessário identificar a causa para a escolha do antibiótico. Progressivamente, o diagnóstico tornou-se uma imposição ao exercício da medicina (CORREA DA SILVA, 2001).

Na evolução da medicina, o reconhecimento de que um paciente não está apenas enfermo, mas apresenta uma doença específica¹, surgiu à medida que se percebeu a existência de conjuntos constantes e repetitivos de fenômenos anormais aparecendo em grupos definidos de pacientes. Tal constatação de classes de doenças, ainda que inicialmente imprecisa, permitiu sua determinação e certa classificação, sendo possível maior entendimento das próprias doenças e melhor manejo terapêutico dos pacientes. (CORREA DA SILVA, 2001).

Com o aumento do uso da Tomografia Computadorizada, nódulos pulmonares são identificados com mais frequência por causa da alta resolução da tomografia em comparação com a radiografia (ALBERT et al, 2009).

Os achados nas imagens radiológicas são distribuídos em 3 grupos: neoplasia² maligna (primária ou metastática), granulomas infecciosos (tuberculose ou fungos), ou o mais comum, tumores benignos (neoplásico ou não-neoplásico) (LEE et al,2010).

O diagnóstico radiológico de neoplasias dos tratos gastrointestinal e geniturinário é bastante acurado, em parte devido à capacidade de recobrir tais neoplasias com agentes de alto contraste, permitindo dessa forma a avaliação detalhada da natureza da superfície da lesão. No caso de neoplasias pulmonares, o ar delinea as massas e cria um material de

¹ Doença específica é definida como um conjunto de fenômenos anormais apresentados por um grupo de indivíduos que ficam numa condição de inferioridade biológica em relação aos outros membros da mesma espécie (CORREA DA SILVA, 2001).

² Neoplasia -(*neo* = novo + *plasia* = formação) é o termo que designa alterações celulares que acarretam um crescimento exagerado destas células, ou seja, proliferação celular anormal, sem controle, autônoma, na qual reduzem ou perdem a capacidade de se diferenciar, em consequência de mudanças nos genes que regulam o crescimento e a diferenciação celulares. A neoplasia pode ser maligna ou benigna.

contraste natural contra o qual os tumores se destacam. Eles freqüentemente são tão acentuadamente contrastados pelo ar em relação às lesões em outros sistemas orgânicos, como o bário ou o iodo (GROSKIN, 1997). A pleura e a parede torácica ocasionalmente são locais de neoplasias primárias que exigem, em geral, drásticas soluções cirúrgicas, as quais podem ser associadas a outras modalidades de tratamento (CORREA DA SILVA, 2001).

Nos Estados Unidos, o câncer de pulmão é o tumor maligno mais comum em homens e mulheres. Cerca de 175 mil novos casos são diagnosticados a cada ano, dos quais 75% a 80% representam o câncer de pulmão de não-pequenas células (SWENSEN et al, 2003).

Os vários tumores primários que ocorrem no mediastino, em todas as idades e em ambos os sexos – especialmente os originários em tecidos tímicos, gânglios linfáticos e nervos – colocam muitas vezes o médico frente a dilemas diagnósticos, submetendo-o a verdadeiros exercícios anátomo-clínicos e freqüentemente exigindo atitudes terapêuticas envolvendo múltiplas áreas (CORREA DA SILVA, 2001).

Nódulos pulmonares são pequenas opacidades focais radiológicas que podem ser solitárias ou múltiplas (GOULD et al, 2007). Na subseção a seguir serão tratados os tipos de nódulos identificados em doenças pulmonares e suas características radiológicas.

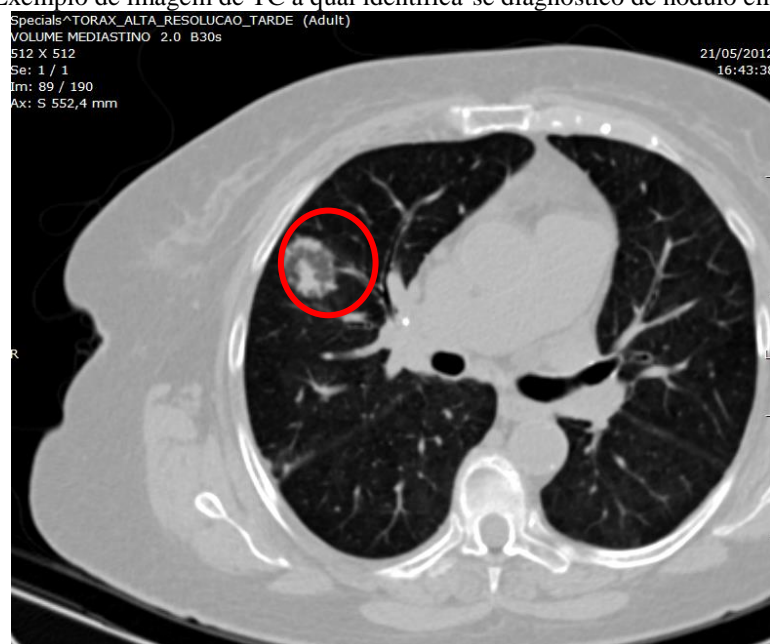
3.2 Tipos de Nódulos Pulmonares

Existem vários tipos de nódulos pulmonares, dentre eles, alguns serão discutidos nas subseções a seguir.

3.2.1 Nódulo em Vidro Fosco

Nódulo em vidro fosco (NVF) é uma área circunscrita do aumento da atenuação pulmonar com preservação das margens brônquicas e vasculares. Este nódulo pode ser parte sólida (parte da opacidade do nódulo em vidro fosco é visualizada completamente escura) ou não sólida (não há áreas completamente escuras). A Figura 1 mostra um exemplo de imagem de TC na qual identifica-se um nódulo em vidro fosco.

Figura 1 - Exemplo de imagem de TC a qual identifica-se diagnóstico de nódulo em vidro fosco



Fonte: Banco de Imagens do Hospital São Lucas da PUCRS, 2012.

Nódulos parcialmente sólidos são mais propensos a serem malignos do que nódulos não-sólidos. De acordo com Winer-Muram (2006), entre 40% e 50% dos nódulos sólidos menores do que 1,5 cm de diâmetro são cancerosos, e o risco de câncer aumenta com o tamanho do nódulo, particularmente se o componente sólido é no centro do nódulo. Doenças inflamatórias do pulmão, como tuberculose e micoses, geralmente produzem nódulos sólidos que podem calcificar e permitir a designação da doença benigna (WINER-MURAM, 2006).

Nódulos não-sólidos são frequentemente causados por condições benignas, tais como: doença inflamatória e lesões pré-malignas como: hiperplasia adenomatosa atípica ou hiperplasia bronco alveolar (WINER-MURAM, 2006).

O NVF representa um desafio para o clínico, porque ele cresce lentamente, e, ao mesmo tempo, tem uma taxa de malignidade de 63% (HOOP et all, 2010). Uma grande variedade de doenças benignas, como por exemplo, doença inflamatória ou fibrose, podem se manifestar como um nódulo em vidro fosco. No entanto, a maioria dos NVF que persistem ao longo do tempo está relacionada com hiperplasia adenomatosa atípica, carcinoma bronco alveolar, ou adenocarcinoma minimamente invasivo (HOOP et all, 2010). Estudos (PATEL et all, 2013) mostram que a presença de componentes sólidos em nódulos de vidro fosco está associado com características patológicas.

3.2.2 Nódulo Solitário

Um nódulo pulmonar solitário (NPS) é definido como uma discreta opacidade bem delimitada, arredondado inferior ou igual a 3 cm de diâmetro, completamente rodeado por parênquima pulmonar, e não está associado com adenopatias, atelectasia ou derrame pleural. Lesões maiores que 3 cm são consideradas “massas” e são tratadas como doenças malignas até que se prove o contrário (LEUNG et al, 2007, MELO et. al, 2012).

A justificativa para o acompanhamento de uma lesão incidentalmente encontrada é que a detecção e o tratamento do câncer de pulmão precoce podem levar à melhores resultados (ALBERT and RUSSELL, 2009).

Um resumo das diferenças entre as características do diagnóstico de um nódulo solitário ou massa com diferentes etiologias, detectada na radiografia ou na TC é visualizada no Quadro 1. (MULLER et al, 2008).

Quadro 1. Características que diferenciam um Nódulo Pulmonar Solitário e Massa

Etiologia	Comentários
Congênito	
Cisto Broncogenico	Morfologia: normalmente são ovalados / arredondados Número TC: 0 HU (água) a 20 HU Localização: subcarinal / paratraqueal direita Outras características: Todos os achados possuem alta intensidade de sinal em ponderação T2 (MRI).
Malformação arteriovenosa pulmonar	Morfologia: circular, oval, ou ligeiramente lobulada Número TC: Localização: - Outras características: As imagens de TC mostram a alimentação pela artéria e drenagem pela veia, demonstrado após injeção de contraste endovenoso.
Infecção	
Tuberculose	Morfologia: Frequentemente apresentam calcificações Número TC: achados tomográficos são variados Localização: lobos superiores Outras características:
Histoplasmoze	Morfologia: Frequentemente apresentam calcificações. Número TC: achados tomográficos são variados Localização: Frequentemente em áreas endêmicas Outras características:
Paracoccidiodomicosis	Morfologia: Frequentemente em cavidades. Número TC: achados tomográficos são variados Localização: Comum em áreas endêmicas Outras características: -
Echinococcus	Morfologia: normalmente são ovalados Número TC: 0HU (água) a 20 HU Localização: intratoracico Outras características:

Abcesso	Morfologia: normalmente são nodulares Número TC: próximo aos valores de 0HU Localização: variável Outras características:-
Pneumonia Focal	Morfologia: margens mal definidas Número TC: densidade de partes moles (20 a 60 HU) Localização: variável Outras características: Menos comum em adultos do que em crianças.
Inflamatório	
Granulomatose Wegener's	Morfologia: nódulos possivelmente escavados Número TC: densidade interna do nódulo com ar Localização: variável Outras características: Nódulo solitário não é comum.
Nódulo Reumatóide	Morfologia: nódulos arredondados ou lobulados Número TC: 30 a 60 HU Localização: variável Outras características: podem cavar
Pneumonia em Organização	Morfologia: margens mal definidas Número TC: opacidades periféricas esparsas Localização: variável Outras características: podem ser migratórias
Pneumonia Lipóide	Morfologia: margens irregulares Número TC: -70 HU Localização: conferir Outras características:-
Sarcoidose em fibrose conglomerada	Morfologia: arredondados ou reticulares (fibrose) Número TC:- Localização: bilateral, perihilar e lobos superiores Outras características: Representa estágio avançado de sarcoidose pulmonar.
Talcosis em fibrose conglomerada	Morfologia:- Número TC: alta atenuação devido a deposição de talco Localização: lobos superiores Outras características: usuário de droga intravenosa. Doença ocupacional.
Pneumoconiosis	
Fibrose massiva progressiva em silicose	Morfologia: reticulares Número TC:- Localização: tipicamente perihilar, bilateral dos lobos superiores Outras características: Associada a linfonomegalias hilares e mediastinais calcificadas tipo "casca de ovo". Doença ocupacional.
Neoplásica	
Carcinoma Pulmonar	Morfologia: liso, lobulado ou mais comum com margens espiculadas Número TC: 30 a 60 HU Localização: variável Outras características:-
Tumor carcinóide	Morfologia: margens usualmente lisas Número TC: conferir Localização: variável Outras características: Lesão tipicamente endobrônquica.
Hamartoma	Morfologia: margens lobuladas, calcificação pipoca Número TC: 60% áreas com -70 HU Localização: variável Outras características: -

Metástase	Morfologia: margens lisas e padrão de distribuição hematogênico. Número TC: variável Localização: predominam nos lobos inferiores Outras características: Metástase solitária incomum exceto para sarcoma.
Traumático	
Hematoma	Morfologia: margens lisas Número TC: 20 a 40 HU Localização: periférica Outras características: Lento, porém possui aumento progressivo do seu tamanho.

Fonte: Adaptada de MULLER et. al, 2008, p. 137.

Um NPS não costuma anunciar sua presença clínica com sintomas. Embora possa ser insidiosa, uma variedade de fatores de riscos clínicos, tais como o avanço da idade e ser fumante estão associados a um maior grau de malignidade. A presença de linfonodos mediastinais nas imagens de TC sugere um novo câncer de pulmão primário, em vez de metástase (PATEL et al, 2013).

As doenças pulmonares intersticiais, como fibrose idiopática, asbestose e esclerodermia estão associadas com um aumento de câncer (PATEL et al, 2013).

Em termos de localização, segundo Winer-Muran (2006), o câncer de pulmão é 1,5 vezes mais provável de ocorrer no pulmão direito do que no pulmão esquerdo. De acordo com os mesmos autores, estudos mostraram que 70% dos cânceres de pulmão estão localizadas nos lobos superiores e ocorrem com maior frequência no pulmão direito.

3.3 *Nódulos Benignos e Malignos*

Os quatro sinais mais úteis na avaliação de um nódulo pulmonar na prática cotidiana são: o tamanho; a variação do tamanho; a presença de calcificação e as características da interface tumor-pulmão (HENSCHKE et al, 2002; MULLER et al, 2008) .

A diferenciação de nódulos em benignos ou malignos é realizada através do acompanhamento de alterações de seu tamanho, presença de calcificação ou gordura e opacidade ao longo do tempo. Esta distinção pode ser estabelecida com um grau razoável de confiança baseado na clínica e critérios radiológicos (MULLER et al, 2008; WINER-MURAM, 2006).

Para um diagnóstico diferencial, é usual dividir os nódulos em duas categorias: (1) aqueles que são claramente benignos devido aos sinais radiológicos, e (2) aqueles de natureza

indeterminada, que representam todas as outras lesões. O principal motivo desta separação é que os critérios de benignidade são mais claros do que os sinais radiológicos de malignidade.

A calcificação em nódulos pulmonares é o sinal mais confiável de que se trata de uma lesão benigna. (MULLER et al, 2008; WINER-MURAM, 2006).

O crescimento descontrolado de um nódulo é um aspecto característico de câncer. Um aumento de seu tamanho deve causar preocupação, ao passo que a ausência de variação é confortante. A estabilidade de 2 anos no tamanho do nódulo pode ser considerada um critério de benignidade. Para pequenos nódulos, nos quais uma pequena alteração no tamanho pode ser difícil de ser visualizada, deve-se continuar a monitorá-los (HOOP et all, 2010).

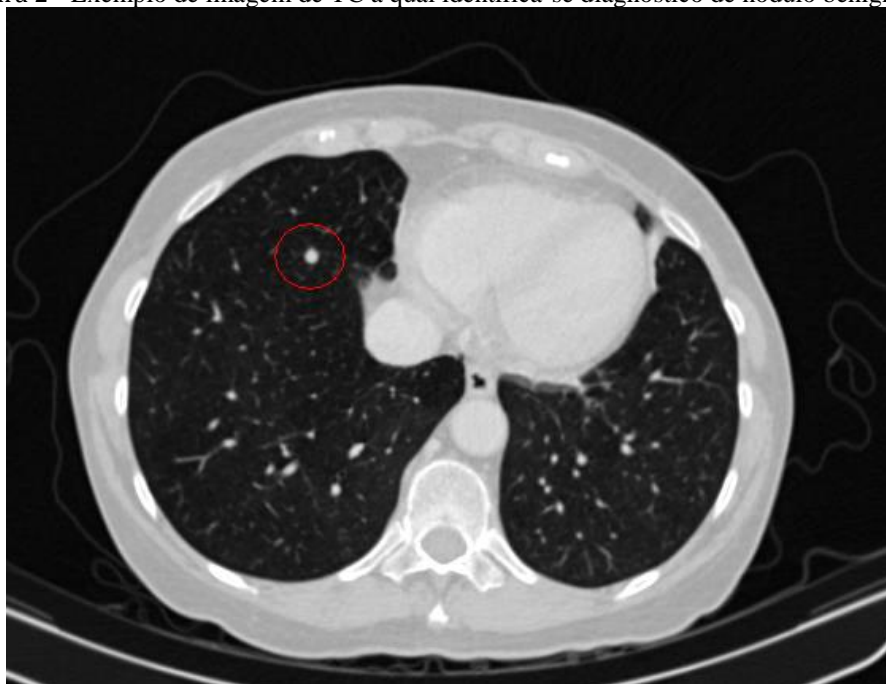
O acompanhamento do crescimento do nódulo deve ser considerado até que o nódulo torne-se estável (isto é, não haja crescimento ou desenvolvimento de um componente sólido), pois seu crescimento é considerado lento; porém, podem demonstrar alterações súbitas no tamanho e densidade. Quando um nódulo aumenta de tamanho, ou se desenvolve de um componente sólido, este deve ser ressecado (HOOP et all, 2010).

Lesões maiores de 3 cm são mais susceptíveis a malignidade. Lesões menores, no entanto, podem ser malignas (PATEL et al, 2013).

Em relação à localização, Patel e colaboradores (2013) mostraram que os nódulos pulmonares solitários localizados nos lobos pulmonares superiores, são diagnosticados como malignos em diversos casos. Isto ocorre porque há uma maior concentração de carcinógenos inalados nos lobos superiores resultantes do cigarro.

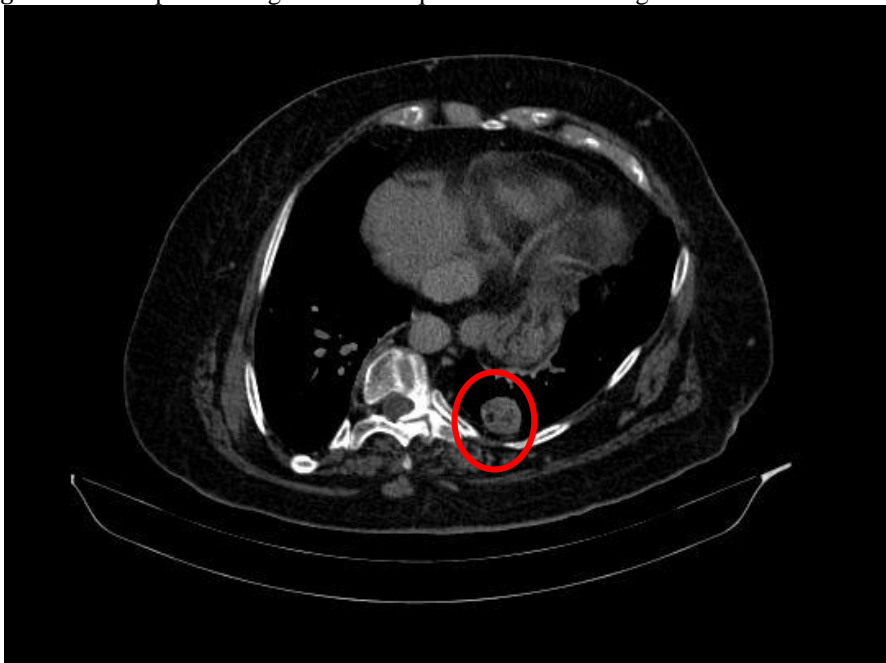
Etiologias benignas mais comuns incluem granulomas infecciosos (Figura 2) e hamartomas (Figura 3), enquanto causas malignas comuns incluem câncer primário de pulmão, tumor carcinóide, e metástases pulmonares (ALBERT and RUSSELL , 2009).

Figura 2 - Exemplo de imagem de TC a qual identifica-se diagnóstico de nódulo benigno



Fonte: Ref. Banco de Imagens do Hospital São Lucas da PUCRS, 2013.

Figura 3 - Exemplo de imagem de TC a qual identifica-se diagnóstico de hamartoma



Fonte: Ref. Banco de Imagens do Hospital São Lucas da PUCRS, 2013

3.4 Atributos da Imagem Radiológica

A análise visual da imagem radiológica para o diagnóstico de nódulos pulmonares, como observado nas subseções anteriores, inclui a avaliação de aspectos relacionados à forma das bordas, dimensão dos nódulos, suas características de opacidade e presença de calcificações, por exemplo. Tais aspectos podem ser quantificados em imagens digitais por meio da análise de características ou atributos quantitativos, também denominados *features*, em inglês.

Os atributos quantitativos em imagens digitais podem ser relacionados à forma ou morfológicos, de intensidade ou de textura, sendo amplamente utilizados em sistemas de auxílio computadorizado ao diagnóstico. As subseções a seguir detalham cada um dos atributos quantitativos utilizados neste trabalho, considerando as imagens radiológicas de interesse: os nódulos pulmonares.

3.4.1 Atributos Morfológicos

Atributos morfológicos são aqueles relacionados à forma dos objetos, tais como perímetro, área, volume, compacidade ou irregularidade, e vem sendo utilizados na classificação das imagens médicas, particularmente na classificação de lesões malignas ou benignas (ALBERT and RUSSELL, 2009).

A volumetria de um nódulo é superior às medidas de diâmetro em nódulos sólidos em termos de precisão e reprodutibilidade, mas as técnicas de medição volumétrica ainda não são regularmente usadas (HOOP et al., 2010).

Características radiológicas, tais como tamanho, forma e taxa de crescimento, muitas vezes ajudam a determinar a probabilidade de malignidade (ALBERT and RUSSELL, 2009; RANGAYAN, 1997).

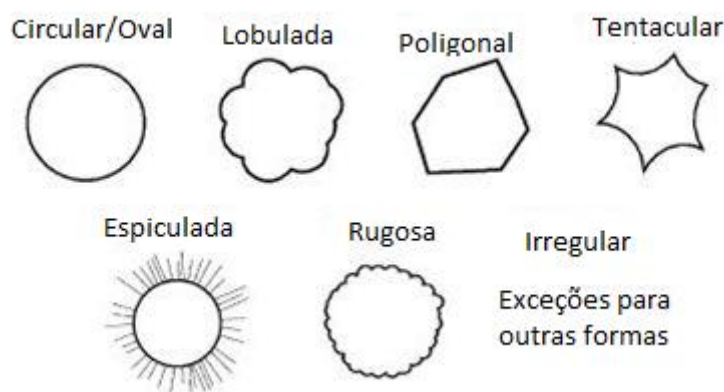
O tamanho do nódulo pode ser caracterizado através das medidas de área, perímetro e os diâmetros máximos e mínimos. Estes recursos também são usados em algumas medidas de forma, tais como características das bordas características com base na energia de flexão do contorno e, características de momento (MCNITT-GRAY et al [A], 1999).

Dentre as características das lesões pulmonares, a forma é a chave para discriminar células normais e anormais. Está bem estabelecido que massas benignas são geralmente arredondadas ou macro-lobuladas na aparência, e bem circunscritas, com contornos suaves.

Por outro lado, os tumores malignos tipicamente possuem limites ásperos ou irregulares, incluindo micro-lobulações, espículas e concavidade.

Bordas irregulares, espiculadas e lobulares indicam características de malignidade (MURAM-WINER, 2006). Em geral, os nódulos redondos, poligonais, tentaculares e rugosos são benignos, enquanto os nódulos espiculados e irregulares são malignos. Nódulos lobulados podem ser tanto benignos como malignos (Figura 4) (IWANO, et al. 2005).

Figura 4 - Sete tipos de bordas/margens de nódulos pulmonares



Fonte: IWANO et. Al. 2005, p. 566.

A maioria dos contornos de massas benignas é suave, ovalar ou possui porções maiores de macrolobulações convexas. Algumas massas benignas podem ter menor concavidade e espicularidade. Por outro lado tumores malignos possuem segmentos côncavos e convexas bem como microlobulações e proeminentes espiculações (IWANO, et al. 2005). O Quadro 2 mostra os tipos de margens descritas e suas possíveis causas:

Quadro 2. Características das margens de um nódulo pulmonar solitário

Margem	Etiologia
Liso	Sugere uma lesão benigna. No entanto, pode ser maligna em até um terço dos casos.
Lobulada	Sugere crescimento desigual, um valor preditivo positivo de 80% para malignidade. Até 25% das lesões benignas, como hamartomas, podem ter margens lobuladas.
Espiculada	É altamente preditiva de malignidade, com 88% a 94%. Algumas exceções poderiam ter margens espiculadas incluem pneumonia lipídica, atelectasia focal, tuberculoma e progressistas maciças de fibrose.
Rugosa	Sugere padrão de crescimento ao longo da parede alveolar; padrão lipídico de adenocarcinoma.
Tentacular ou Poligonal	Visualizado em fibrose, infiltração alveolar e colapso dos alvéolos.

Halo	Nódulo pulmonar solitário cercado por um “halo” de atenuação em vidro fosco, também chamado de “sinal do halo TC”. Visto em aspergilose, sarcoma de Kaposi, granulomatose com poliangiite e angiossarcoma metastático. Adenocarcinoma in situ também pode produzir um halo, devido ao seu crescimento lipídico.
Concavidade	Nódulos com entalhes ou concavidades na margem são vistos em alguns nódulos com o crescimento do tumor. Estes entalhes são frequentemente encontrados em adenocarcinoma com evidente invasão e estão associados com prognóstico reservado.

Fonte: Adaptado de PATEL et. Al, 2003, p. 828

Outra característica importante do nódulo é o crescimento de seu tamanho ao longo do tempo. A medição do diâmetro dos nódulos é realizada em pelo menos duas dimensões, com uma média de duas imagens consecutivas. Uma limitação da medição de um nódulo ocorre quando há inflamação, atelectasia ou cicatrizes adjacentes ao nódulo, fazendo assim a lesão parecer maior do que a realidade (WINER-MURAM, 2006).

Um nódulo que não apresente um crescimento ao longo de, pelo menos, 2 anos, é considerado benigno e é amplamente aceito pelo padrão de tratamento clínico. O tempo para que um nódulo dobre seu volume é conhecido como tempo de duplicação (WINER-MURAM 2006).

Lesões malignas usualmente têm um tempo de duplicação entre um mês e um ano. Assim, um nódulo que duplique de tamanho em menos de um mês, ou mantenha-se estável durante mais do que um ou dois anos, é provavelmente benigno (ALBERT et. al, 2009).

3.4.2 Atributos de Intensidade

Atributos de intensidade são aqueles relacionados à denominada densidade ou opacidade radiológica da lesão, como usualmente nominado pelos radiologistas. Um nódulo de alta densidade é geralmente um indicador de malignidade e está relacionado à presença de calcificações, enquanto lesões de baixa densidade são associadas a nódulos que possuem gordura sendo geralmente benignas (HOOP et al, 2010; MCNITT-GRAY et al, 1999).

A densidade ou opacidade radiológica de uma imagem radiológica está relacionada ao processo de atenuação dos raios X ao interagirem com os tecidos do corpo humano.

No equipamento de TC, o tubo emite raios X gira em torno do paciente em um plano axial. Um conjunto de detectores posicionados capta os fótons de raios X que atravessam o paciente após a interação e um algoritmo de reconstrução tomográfico converte os sinais das projeções medidas pelos detectores em uma imagem da seção transversal da região varrida.

A imagem resultante da reconstrução da TC representa os coeficientes de atenuação dos tecidos que compõem um corte transversal do corpo. Cada pixel da imagem representa uma medida do coeficiente de atenuação de um elemento de volume (voxel), que se estende através da espessura do corte, quantificando a fração de radiação que passa através de uma dada quantidade de um determinado material de uma espessura Δx (MAHESH, 2002; HSIEH, 2003).

A energia dos fótons na TC por raios X é aproximadamente 150 keV. Com essa energia, a seção de choque necessária para a produção de pares ou para o espalhamento coerente tem pequena contribuição para o coeficiente de atenuação linear (μ). Assim, no caso da TC, o coeficiente de atenuação pode ser considerado uma função quase exclusiva da densidade física do tecido, ρ , ou da densidade eletrônica, Ne (MULL, 1984).

No meio médico, usualmente utiliza-se para caracterizar a atenuação dos raios X pelos tecidos, ao invés do valor do coeficiente de atenuação ou densidade, o denominado número TC ou número de Hounsfield, ou número HU. O número TC é uma escala normalizada dos coeficientes de atenuação em relação ao coeficiente de atenuação da água (Equação 1). A unidade do número TC é Hounsfield (HU), em homenagem ao criador da tomografia computadorizada.

$$N^{\circ} CT (HU) = k \left(\frac{\mu - \mu_w}{\mu_w} \right) \quad (1)$$

onde k é uma constante de magnificação (normalmente igual a 1000), dependente do equipamento, μ é o coeficiente de atenuação linear do tecido e μ_w é o coeficiente de atenuação linear da água. Os números de TC são estabelecidos baseados numa referência relativa, no caso água. Em função da definição do número TC, seu valor para a água é zero. Para tecidos menos densos do que a água, como a gordura, o valor do número TC é negativo, para tecidos mais densos, como os ossos e calcificação, o valor é positivo (MECCA, 2009).

Nas estações de trabalhos dos equipamentos, as ferramentas de análise das imagens costumam mostrar os valores dos número TC em regiões de interesse, permitindo uma caracterização do tecido (Figura 5).

Figura 5 - Exemplo de imagem de TC, na qual identificam-se os valores de uma região de interesse em termos de unidades Hounsfield



Fonte: Ref. Banco de Imagens do Hospital São Lucas da PUCRS, 2013.

Para caracterizar os valores da densidade ou opacidade de uma região de interesse de uma imagem de TC, podem ser utilizados atributos de intensidade, tais como valor médio de um região e desvio padrão. Estes valores podem ser utilizados na classificação de lesões, visto que caracterizam a presença de calcificações, gordura e outras substâncias (ALBERT and RUSSELL, 2009).

A detecção de nódulos com densidade semelhante à da água na TC é diagnosticada como uma lesão cística benigna, por exemplo. No entanto, o diagnóstico de uma área de vidro fosco ou consolidação é amplo e pode incluir pneumonia bacteriana, fúngica, viral ou focal, entre outros (MULLER et al, 2008), não possibilitando sua classificação com base nesse atributo.

Diversos pesquisadores vêm realizando estudos sobre as densidades de lesões pulmonares benignas e malignas (SIEGELMAN et al., 1980; ALBERT et al., 2003; WINER-MURAM, 2006; PATEL et al., 2013).

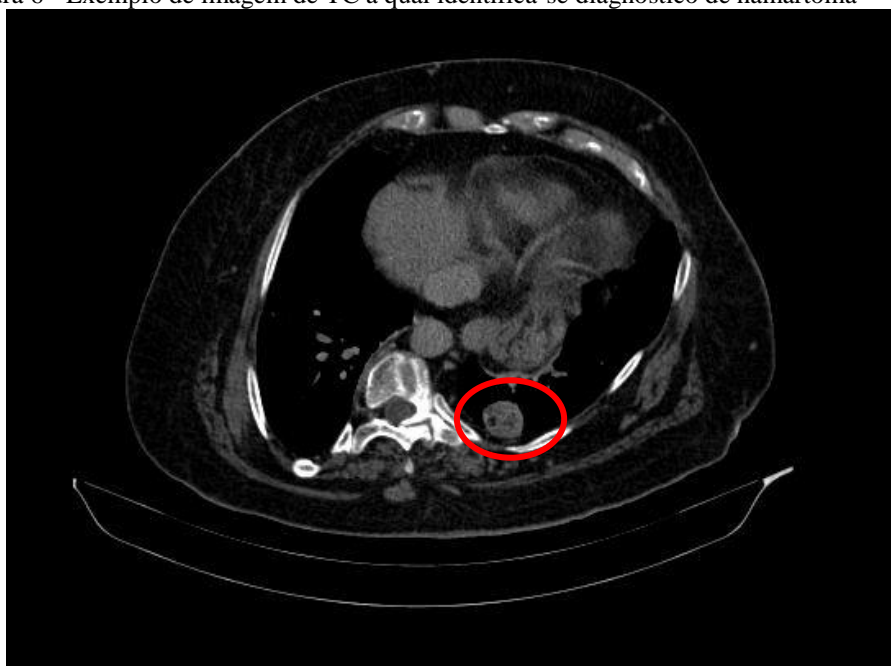
Um estudo realizado por Siegelman et al (1980) mostrou que pacientes diagnosticados com câncer de pulmão primário, possuem lesões malignas com uma média do número TC de (92 ± 18) HU. Segundo o autor, os carcinomas de células escamosas são geralmente mais densos do que outros tipos de células. Os valores de metástases pulmonares variaram de 57 a 147 HU, com média de 98 HU. Os nódulos benignos são divididos em três grupos: no grupo 1, os números de TC são semelhantes aos das lesões malignas, variando de 60 a 142 HU; no segundo grupo, cujos pacientes possuem diagnóstico de granuloma, ou não mostram crescimento do nódulo, o número TC está entre 164 a 860 HU; no terceiro grupo, classificado

como intermediário, estão pacientes que possuem nódulos com números de TC de 147 a 162 HU. Este último grupo foi observado em uma fase inicial do estudo e seus nódulos foram considerados provavelmente benignos e nenhuma das lesões mostrou qualquer crescimento no seguimento de um ano ou 18 meses. Os pacientes com nódulos pulmonares considerados definitivamente calcificados possuem valores de número TC superiores a 600. Lesões com calcificações densas possuem valores de 1200 a 1600 HU.

De acordo com Patel e colaboradores (2013), um valor de atenuação entre -40 e -120 HU sugere a presença de gordura, que está presente em 60% dos hamartomas, nódulos benignos. Causas raras de atenuação de gordura incluem pneumonia lipídica e metástases pulmonares de lipossarcoma (quase sempre lesões de aparência sólida) e carcinoma de células renais (PATEL et al, 2013).

Pode haver dificuldades na determinação de densidades de gordura em nódulos muito pequenos. No entanto, pode-se considerar que a gordura está presente, em casos de hamartoma (Figura 6) ou lipoma. Algumas doenças malignas, tais como: metástase de lipossarcoma ou carcinoma de células renais, ocasionalmente, podem conter gordura. Em pacientes sem doença maligna, a atenuação focal de gordura (-40 a -120 HU) é um indicador confiável de um hamartoma e é visto em mais de 50% dos hamartomas em cortes tomográficos, segundo Winer-Muram (2006).

Figura 6 - Exemplo de imagem de TC a qual identifica-se diagnóstico de hamartoma



Fonte: Ref. Banco de Imagens do Hospital São Lucas da PUCRS, 2013.

Albert e colaboradores (2009) mostraram que, em um estudo com mais de 13.000 pacientes, 26% das lesões predominantemente sólidas eram malignas, 73% das lesões não sólidas que tinham opacidade em vidro fosco eram malignas. Outro estudo (TOZAKI et al, 2005) mostrou que a presença de bordas irregulares estava associada com um aumento de quatro vezes da probabilidade de ocorrência de malignidade.

Uma medida alternativa relacionada à intensidade da imagem, que pode caracterizar os nódulos é a massa, que considera o volume e a densidade do nódulo. Essa medida pode ser avaliada para a identificação de nódulos em vidro fosco com um alto risco de malignidade (HOOP et al, 2010).

A massa de um nódulo pode ser calculada multiplicando o volume de um nódulo pela sua densidade. A atenuação em Unidades de Hounsfield (HU) pode ser traduzida diretamente em densidade física, em miligramas por mililitro, adicionando 1000 ao valor unitário do número TC em Hounsfield (HOOP et al, 2010). Alterações dos nódulos em vidro fosco são normalmente avaliadas usando medidas de diâmetro e avaliação visual da aparência de um componente sólido. A massa pode ser calculada usando os dados da tomografia por causa dos valores de atenuação dos raios X serem proporcionais à densidade do tecido (massa por unidade de volume) (Equação 2).

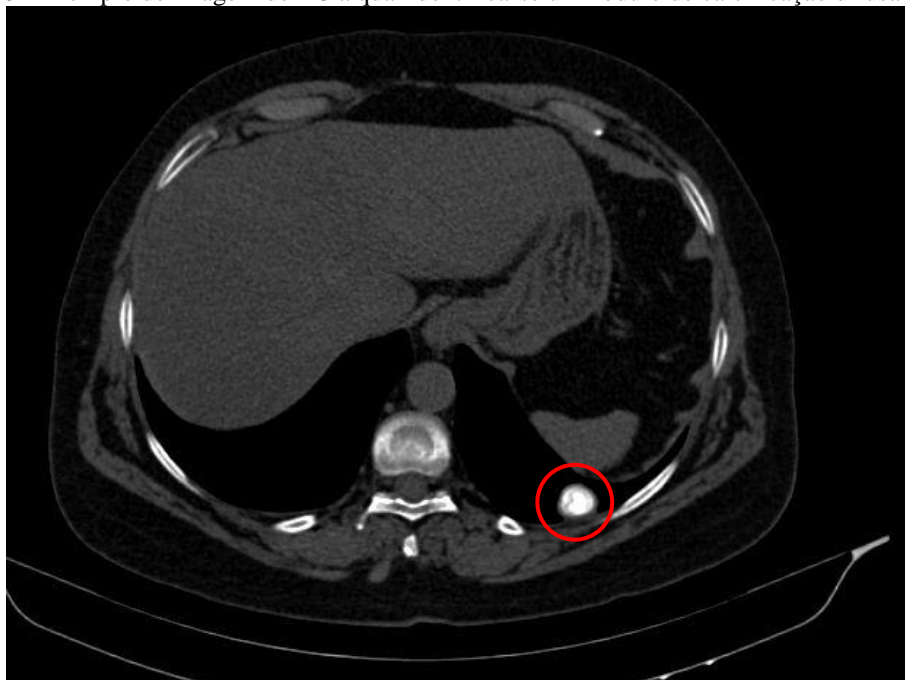
$$massa = Volume \times densidade \quad (2)$$

Calcificações centrais, laminadas ou pipoca (Figura 7) e difusas (Figura 8) são padrões benignos de calcificação. De acordo com Winer-Muram (2006), quando um desses padrões de calcificação é visualizado, a probabilidade de benignidade se aproxima de 100%. Ainda segundo Winer-Muram (2006), calcificações pipocas são observadas em um terço dos hamartomas, e os outros padrões são vistos em histoplasmose e tuberculose. Todos os outros padrões de calcificação não devem ser considerados como um sinal de benignidade (LEUNG et. al, 2007).



Fonte: Ref. LEUNG et. al, 2007

Figura 8 - Exemplo de imagem de TC a qual identifica-se um nódulo de calcificação difusa



Fonte: Ref. Banco de Imagens do Hospital São Lucas da PUCRS, 2013.

A exceção à regra anteriormente descrita é quando os pacientes são conhecidos por terem um tumor primário. Por exemplo, o padrão de calcificação difusa pode ser observado em pacientes com osteossarcoma ou condrossarcoma. Da mesma forma, o padrão central e pipoca podem ser observados em pacientes com tumores do trato gastrointestinal e pacientes que realizaram tratamento de quimioterapia (LEUNG et al, 2007).

Devido ao alto coeficiente de atenuação do cálcio comparado com o tecido normal, muitas características de calcificações são relativamente intensas (brilhantes). Isto faz com que calcificações sejam mais facilmente distinguíveis quando adquiridas em exames de imagens. Calcificações malignas tendem a ser numerosas, agrupadas, pequenas, variando de tamanho e forma angular, forma irregular, e separado em orientações, como por exemplo, na mamografia. Por outro lado, calcificações associadas a condições benignas geralmente são maiores, arredondadas, menores em número, difusamente distribuídas, e homogêneas em tamanho e forma. A diferença entre calcificações malignas e benignas está na rugosidade das suas formas (RANGAYYAN, 2005).

Patel e colaboradores (2013) mostraram que a radiografia simples de tórax não é muito sensível na detecção de calcificação dentro de um nódulo, demonstrando uma sensibilidade de 50% e uma especificidade de 87%. Assim, uma TC sem contraste com cortes de 1 a 3 mm é

recomendada para avaliar o nódulo e a calcificação do nódulo. Valores de atenuação maiores de 200 HU em um nódulo indicam a presença de cálcio.

3.4.3 Atributos de Textura

A textura dá as informações sobre a densidade interna do nódulo e desempenha um papel importante quando se tenta segmentar e caracterizar um nódulo (VINAY et al, 2011).

Atributos de textura são aquelas relacionadas à distribuição espacial dos níveis de cinza dos pixels de uma região como a homogeneidade, entropia e variância. As características de textura são baseadas em estatísticas com distribuição da frequência relativa das intensidades dos pixels (HARALICK et al, 1973).

Medidas de textura são usadas para distinguir nódulos uniformes daqueles que tem distribuições de atenuação irregulares, tais como as causadas pela presença de calcificação ou distribuição de gordura. Medida do nível de cinza espacial dependente da matriz tem sido usada para caracterizar a textura de nódulos pulmonares em TC. Estas medidas de textura são usadas para quantificar a natureza das diferenças locais e levar em conta tanto a distancia como a direção das diferenças (MACCNITT-GRAY et al [B], 1999).

O Quadro 3 resume as características radiológicas sugestivas de benignidade ou malignidade de nódulos pulmonares solitários, descritas anteriormente.

Quadro 3. Características Radiológicas sugestivas de benignidade ou malignidade de nódulos pulmonares solitários

Características Radiológicas	Benignidade	Malignidade
Tamanho	< 5 mm	> 10 mm
Borda	Liso	Irregular ou Espiculado
Densidade	Denso, sólido	Não sólido, “vidro fosco”
Calcificação	Tipicamente características benignas especialmente classificadas como “concêntrica”, “central”, “pipoca” ou padrões “homogêneo”	Tipicamente não calcificado ou calcificação “excêntrica”
Tempo de duplicação	Menor do que um mês; mais do que um ano.o	Um mês a um ano

Fonte: Adaptado de ALBERT et. al, 2009, p. 828

3.5 *Computed Aided Diagnosis – CAD*

O Diagnóstico Auxiliado por Computador, do inglês, *Computed-aided diagnosis – CAD*, pode ser definido como um diagnóstico realizado por um radiologista que utiliza o resultado de análises quantitativas automatizadas como uma segunda opinião para a tomada de decisões diagnósticas (AZEVEDO MARQUES, 2001).

Existem dois tipos de aplicações de sistemas CAD. Um auxilia na detecção das lesões a partir da localização de padrões anormais através da varredura da imagem pelo computador, ou seja, identifica a localização dos nódulos pulmonares em imagens de tórax. O outro auxilia no diagnóstico através da quantificação de características da imagem e sua classificação como correspondendo a padrões normais ou anormais, ou seja, associa a textura dos pulmões com as lesões intersticiais em imagens de tórax (AZEVEDO-MARQUES, 2001). Este último será o método estudado nessa dissertação.

A finalidade do CAD é melhorar a acurácia do diagnóstico, assim como a interpretação da imagem radiológica, através do uso da resposta do computador como referência. Esta resposta pode ser útil, uma vez que o diagnóstico do radiologista é baseado em uma avaliação subjetiva, sujeita a variações intra e interpessoais, bem como perda de informação devido à natureza do achado radiológico, pouca qualidade na imagem, sobreposição de estruturas, fadiga visual ou distração. Uma dupla leitura (por dois radiologistas) pode aumentar a sensibilidade do diagnóstico funcionando como um segundo especialista (AZEVEDO-MARQUES, 2001).

Este procedimento ajuda a definir a variabilidade de interpretação que os radiologistas podem definir para desenvolver um sistema de diagnóstico auxiliado do por computador (CAD). Esta aproximação conduz a um aumento na compreensão do trabalho de classificação característica dos radiologistas em relação aos nódulos que é chamada representação subjetiva característica. A classificação do nódulo é um desafio que envolve a análise subjetiva, mas garante interpretação objetiva (AZEVEDO-MARQUES, 2001).

Os sistemas CAD podem auxiliar na distinção entre lesões malignas e benignas e aumentar a sensibilidade e especificidade do diagnóstico. Para tanto, utilizam atributos extraídos e quantificados de forma automatizada. A vantagem da extração automatizada, pelo computador, é a objetividade e reprodutibilidade da medida dos atributos escolhidos. Por outro lado, os radiologistas utilizam uma grande quantidade de atributos objetivos e subjetivos, os quais são extraídos e interpretados simultânea e instantaneamente. É importante ressaltar que um dos objetivos dos sistemas de auxílio à classificação entre achados benignos

e malignos é a redução do número de casos benignos encaminhados para biópsia. Como o "custo" da perda de uma lesão maligna é muito maior que o de uma classificação errônea de um caso benigno, os sistemas CAD devem ser desenvolvidos com o propósito de aumentar a especificidade, porém sem reduzir a sensibilidade (AZEVEDOMARQUES, 2001).

3.5.1 Detecção de Nódulos Pulmonares em Imagens Digitais

Para detectar nódulos pulmonares em imagens de TC, é necessário separar regiões pulmonares por segmento de outras regiões, tais como músculo, gordura, osso, mediastino e fundo. Todos os passos de processamento subsequentes são restringidos para o interior das regiões pulmonares segmentados (HARALICK et al, 1973).

A segmentação pulmonar é uma tarefa relativamente simples, que normalmente é realizado corte a corte. Na maioria dos sistemas de diagnóstico auxiliado por computador ou CAD, a técnica de detecção inicial é aplicada diretamente às imagens originais de TC (LI 2007). No entanto, essa facilidade não ocorre no caso da segmentação de lesões no tecido pulmonar, particularmente no caso de lesões próximas às paredes da caixa torácica ou nódulos em vidro fosco. Minami e colaboradores (2008) segmentaram regiões de nódulos pulmonares a partir de um volume isotrópico reconstruído a partir de cada uma das séries de imagem.

3.5.2 Quantificação de Atributos em Imagens Digitais

Para quantificar lesões pulmonares através de seu tamanho, forma, intensidade ou textura, torna-se necessário inicialmente segmentar ou separar esse nódulo do resto da imagem radiológica através de processos de segmentação. A partir do nódulo segmentado, inicia-se o processo de extração de seus atributos.

Em geral, em imagens digitais, as formas podem ser representadas pelo contorno num espaço discretizado em termos de coordenadas (x,y) de pontos digitalizados (pixels) ao longo do contorno. Um contorno com N pontos pode ser representado por uma série de coordenadas $\{x(n), y(n)\}$, $n=0, 1, 2, \dots, N-1$. Não há nível de cinza associado com pixels ao longo do contorno e ele pode ser representado como imagem binária ou bi-nível (RANGAYYAN, 2005; RANGAYYAN et al, 1997).

A partir da determinação do contorno de um objeto em uma imagem digital, diversos atributos morfológicos podem ser determinados através de medidas geométricas básicas, como o perímetro e a área. Não há um modelo suficientemente genérico para descrever as formas dos objetos. Normalmente, a forma é definida através da comparação com outros objetos ou através de parâmetros que apresentam o quanto a forma é distante de um objeto padrão, como, por exemplo, um círculo.

A compacidade (C) ou *compactness*, em inglês, é um descritor muito utilizado para caracterizar a forma de contornos que contém uma dada área. Existem diversas definições de compacidade, mas basicamente, esse descritor quantifica o quão próximo da forma do círculo é o contorno do objeto. A compacidade é adimensional e pode ser definida como mostra a equação (3) (BANKMANN, 2000):

$$C = \frac{P^2}{4\pi A} \quad (3),$$

onde P e A representam perímetro e a área incluída no contorno, respectivamente. Quanto menor for a área contida por um contorno de um dado comprimento, maior será o valor de compacidade. A compacidade possui um limite inferior de 4π para um círculo, mas não há limite superior. Ela é invariável à mudanças de escala, rotação e reflexão do contorno. A fim de restringir e normalizar o intervalo de $(0, 1)$, bem como a obtenção de valores crescentes, com o aumento da complexidade da forma, a definição da compacidade pode ser modificada (IWANO, et al. 2005).

A irregularidade (I) também relaciona o perímetro e a área do objeto, sendo definida como:

$$I = \frac{A}{P^2} \quad (4)$$

Além dos atributos descritos anteriormente, outras medidas são úteis para a caracterização de objetos em imagens digitais, como é o caso dos momentos de ordem zero, primeira, segunda e terceira, os quais podem fornecer informações sobre tamanho, forma e orientação de objetos (PRATT, 1991).

Os momentos geométricos regulares são definidos como a forma de uma projeção de função $f(x,y)$, representando a imagem em uma função polinomial do tipo $x^p y^q$. O momento $m_{p,q}$ é definido de acordo com a equação a seguir:

$$m_{p,q} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x,y) dx dy \quad (5),$$

sendo $p, q = 0, 1, \dots, \infty$.

O momento $m_{0,0}$ é a área da região e $m_{0,1}$ e $m_{1,0}$ são as coordenadas do centro de massa da região. As coordenadas do centro de massa da região, normalizadas pela área, são dadas por:

$$\bar{x} = \frac{m_{1,0}}{m_{0,0}} \quad e \quad \bar{y} = \frac{m_{0,1}}{m_{0,0}} \quad (6)$$

Os chamados momentos centrais são centralizados em regiões, e são expressos por:

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) dx dy \quad (7)$$

Como os momentos são aplicados a imagens digitais, a equação pode ser simplificada:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (8)$$

Para o cálculo dos momentos centrais de primeira, segunda e terceira ordens, tem-se:

1ª ordem:

$$\mu_{10} = \mu_{01} = 0 \quad (9)$$

2ª ordem:

$$\mu_{20} = m_{20} - \bar{x}m_{10} \quad (10)$$

$$\mu_{02} = m_{02} - \bar{y}m_{01} \quad (11)$$

$$\mu_{11} = m_{11} - \bar{y}\bar{x}m_{10} \quad (12)$$

3ª ordem:

$$\mu_{12} = m_{12} - 2\bar{y}\bar{x}m_{11} - \bar{x}m_{02} - 2\bar{y}\bar{x}^2m_{10} \quad (13)$$

$$\mu_{21} = m_{21} - 2\bar{x}\bar{y}m_{11} - \bar{y}m_{02} - 2\bar{x}^2\bar{y}m_{01} \quad (14)$$

$$\mu_{30} = m_{30} - 3\bar{x}m_{20} - 2\bar{x}^2m_{10} \quad (15)$$

$$\mu_{03} = m_{03} - 3\bar{y}m_{02} - 2\bar{y}^2m_{01} \quad (16)$$

Os momentos μ_{20} e μ_{02} representam a variância da massa do objeto na direção x (centralidade horizontal) e na direção y (centralidade vertical), respectivamente. A diagonalidade (μ_{11}) é a distribuição da massa sobre os 4 quadrantes. A divergência horizontal (μ_{12}) e vertical (μ_{21}) indicam a simetria do objeto em relação aos eixos x e y . Já os μ_{30} (desequilíbrio horizontal) e μ_{03} (desequilíbrio vertical) representam a localização do centro de massa com relação à metade da largura e altura do objeto (BANKMANN, 2000; PRATT, 1991).

Ao normalizar os momentos centrais pela área, os mesmos se tornam invariantes em escala, sendo definidos por:

$$n_{pq} = \frac{\mu_{pq}}{\mu_{00}^y} \quad (17),$$

sendo $y = \frac{p+q-2}{2}$, para $(p+q) \geq 2$.

A partir da combinação dos momentos centrais normalizados pela área, Hu (1962) definiu 7 funções, denominadas Momentos Invariantes (MI) de Hu (1962), os quais são invariantes à rotação, translação e escala. Os MI são calculados por:

$$\Phi_1 = n_{20} + n_{02} \quad (18)$$

$$\Phi_2 = (n_{20} - n_{02})^2 + 4n_{11}^2 \quad (19)$$

$$\Phi_3 = (n_{30} - 3n_{12})^2 + (3n_{21} - n_{03})^2 \quad (20)$$

$$\Phi_4 = (n_{30} + n_{12})^2 + (n_{21} + n_{03})^2 \quad (21)$$

$$\Phi_5 = (n_{30} - 3n_{12})(n_{30} + n_{12})[(n_{30} + n_{12})^2 - 3(n_{21} - n_{03})^2] + (3n_{21} - n_{03})(n_{21} + n_{03})[3(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2] \quad (22)$$

$$\Phi_6 = (n_{20} - n_{02})[(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2] + 4n_{11}(n_{30} + n_{12})(n_{21} + n_{03}) \quad (23)$$

$$\Phi_7 = (3n_{21} - n_{03})(n_{30} + n_{12})[(n_{30} + n_{12})^2 - 3(n_{21} + n_{03})^2] + (3n_{12} - n_{30})(n_{21} + n_{03})[3(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2] \quad (24)$$

Dentre os atributos quantitativos em imagens digitais, a textura é uma das mais utilizadas. A textura é definida como "uma característica representativa da distribuição espacial dos níveis de cinza dos pixels de uma região" (HARALICK, 1973). Ou seja, um atributo de textura é um valor, calculado a partir da imagem de um objeto, que quantifica características relacionadas com a variação dos níveis de cinza desse objeto. Normalmente, um atributo de textura não depende da posição, orientação, tamanho, forma e brilho do objeto.

Em busca de recursos significativos para descrever imagens, Gonzales (2000) define três abordagens usadas em processamento de imagens na descrição de textura: estatística, estrutural e espectral. A abordagem estrutural trata de arranjos de imagens primitivas, como a descrição de textura baseada em linhas paralelas espaçadas. A abordagem espectral baseia-se em propriedades do espectro de Fourier, usadas basicamente na detecção da periodicidade global em uma imagem através da identificação de picos de alta frequência do espectro. A abordagem estatística considera a distribuição de níveis de cinza e o interrelacionamento entre eles.

Haralick (1973) descreve as características espectrais, texturais e contextuais como os três elementos fundamentais utilizados na interpretação humana de imagens coloridas. Características espectrais descrevem as variações médias de tons em diferentes bandas da parte visível e/ou infravermelhos de um espectro eletromagnético, enquanto que as

características texturais contêm informações sobre a distribuição espacial das variações de tonalidade dentro de uma banda. Características contextuais contêm informações derivadas de blocos de dados pictóricos em torno da área a ser analisada. Quando pequenas áreas da imagem são processadas em preto e branco de forma independente por uma máquina, a textura e o tom são os mais importantes.

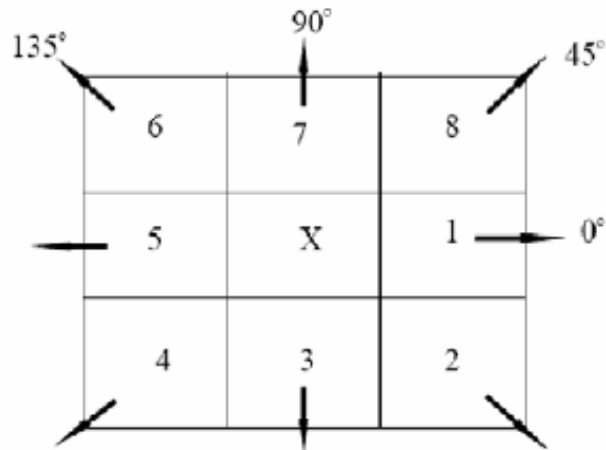
A textura pode ser agrupada em duas categorias: periódica e randômica. Se houver uma repetição de um elemento de textura em intervalos regulares ou quase periódicos, podemos classificar a textura como sendo periódica ou ordenada. Textura também pode estar relacionada a sensações visuais e/ou táteis como finura, aspereza, suavidade, granularidade, periodicidade ou ter uma orientação preferencial. Medidas estatísticas como nível de cinza de matrizes de co-ocorrência e entropia caracterizam a textura num sentido estocástico (RANGAYYAN, 2005).

Para o cálculo de textura, costuma-se determinar a matriz de co-ocorrência (Figura 9) dos níveis de cinza (GLCM, do inglês *Grey Level Co-occurrence Matrix*), que é uma matriz de probabilidade de ocorrência $P(i, j)$ de um determinado par de pixels da imagem, sendo i o nível de cinza do pixel de interesse da matriz e j o nível de cinza do pixel vizinho ao nível de interesse, em relação a uma distância d e um ângulo θ escolhido entre eles. Os cálculos são realizados em quatro direções: 0° , 45° , 90° e 135° , nos dois sentidos da matriz, como ilustrado na Figura 9. As distâncias de cálculos da GLCM variam de acordo com o espaçamento o qual se deseja analisar entre os pares de pixels (HARALICK et al, 1973).

A linha e a coluna da matriz de co-ocorrência variam do menor até o maior nível de cinza da imagem, sendo uma matriz sempre quadrada. A matriz possui simetria em relação a sua diagonal principal e é preenchida com a contagem de quatro pares de combinações que ocorrem em diferentes níveis de cinza (HARALICK et al, 1973).

Após calcular os quatro ângulos, realiza-se a normalização e calcula-se a média das quatro matrizes, transformando-as em uma única matriz que representará quantitativamente a região a ser analisada. (HARALICK et al, 1973).

Figura 9 - Ângulos para o cálculo da matriz de concorrência



Fonte: Adaptado de Trindade, 2009.

A homogeneidade também conhecida como momento inverso de diferenças de ordem k , fornece uma medida da distribuição dos valores de intensidade dos pixels, retornando um valor que representa a proximidade da distribuição dos elementos à diagonal da GLCM. Os possíveis valores encontrados são de 0 a 1 (HARALICK et al, 1973).

$$\text{Homogeneidade: } f_1 = \sum_i \sum_j \{p(i, j)\}^2 \quad (25)$$

O contraste é um momento da diferença da matriz p e é uma medida do contraste ou a quantidade de variações de intensidade presentes na imagem. Uma vez que existe uma grande quantidade de variação local presente na imagem, o contraste para essa imagem possui valores elevados (HARALICK et al, 1973).

$$\text{Contraste: } f_2 = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) \right\} \quad (26)$$

A Variância expressa a variação do fundo da imagem.

$$\text{Variância: } f_4 = \sum_i \sum_j (i - \mu)^2 p(i, j) \quad (27)$$

O Momento do Inverso das Diferenças mede a homogeneidade local realça texturas sutis e pequenas mudanças.

$$\text{Mom. Inv. Diferença: } f_5 = \sum_i \sum_j \frac{1}{1+(i-j)^2} p(i, j) \quad (28)$$

A entropia fornece o grau de dispersão de níveis de cinza de uma imagem, medindo a informação contida em $p(i, j)$, onde $p(i, j)$ é a probabilidade de ocorrência do par de pixels da imagem, i é o nível de cinza do pixel de interesse da matriz, j é o nível de cinza do pixel vizinho ao nível de interesse. Muitos pixels com o mesmo valor de nível de cinza significam uma imagem com pouca informação ou pequena entropia (Haralick et al, 1973).

$$\text{Entropia: } f_3 = -\sum_i \sum_j p(i, j) \log_2 p(i, j) \quad (29)$$

3.6 *Revisão da Literatura: Análise dos Atributos em imagens de TC de pulmão*

Na seção a seguir é apresentada uma revisão bibliográfica dos artigos que tratam dos métodos de análise de atributos quantitativos em imagens de TC de pulmão, em ordem cronológica.

Giger e colaboradores (1994, apud ARMATO et al, 2003, p.1188) desenvolveram uma detecção automática de nódulos baseado em análises de características morfológicas como forma do contorno do nódulo, tamanho e localização, combinado com limiarização de níveis de cinza. Em cada corte, os limites do tórax e pulmão eram detectados, e os recursos dentro dos limites eram submetidos à limiarização dos níveis de cinza. Ao analisar as relações entre as características que surgem em diferentes níveis de limiarização, no que diz respeito à forma, tamanho e localização, cada característica foi atribuída a uma probabilidade de ser um nódulo ou um vaso. O tórax era separado do fundo externo usando as técnicas de limiarização de níveis de cinza. Um perfil dos valores de cinza dos pixels ao longo de uma linha a partir do centro de uma borda da imagem era calculado e utilizado para determinar a limiarização dos níveis de cinza de um corte específico. Para eliminar contornos que surgissem da mesa de exame, cada contorno detectado era examinado em relação a restrições geométricas de localização, tamanho (área) e circularidade. Cada limiarização dos níveis de cinza dos atributos da região interna do pulmão foram rastreados. Sete descritores geométricos de cada atributo foram calculados os quais são: perímetro, área, compacidade, alongamento, circularidade e medida da distância. Perímetro envolve a contagem do número de pixels ao redor da margem do nódulo, levando em consideração as bordas interiores dos pixels. Para calcular a área é basicamente o número de pixels dentro do contorno do nódulo. Compacidade proporciona informações sobre a forma e contorno do nódulo, tal como as medidas de

alongamento e de circularidade. No cálculo da medida da distancia os limites internos e externos do pulmão referem-se às localizações mediais e laterais. Os resultados mostraram uma sensibilidade de 94%. No entanto os níveis de desempenho podem variar com a base de dados examinadas. Para casos em que há anormalidade presentes, o esquema de detecção de nódulos teria de ser combinado com outros métodos de detecção para que pudessem ser sinalizados estes casos antes de utilizar o programa de detecção de nódulos.

Da mesma forma, Toshioka e colaboradores (1997) desenvolveram um algoritmo para detecção de câncer de pulmão com base nas imagens de TC de pulmão. Este algoritmo consiste em uma etapa de análise e outra de procedimentos de diagnóstico. Na etapa de análise são extraídas regiões do pulmão e vasos pulmonares analisando as características dessas regiões através de técnicas de processamento de imagens. No procedimento de diagnóstico são definidas regras com base nas características dessas regiões. As características das regiões que continham câncer foram analisadas e dentre elas, uma atenção especial foi dada para a forma do contorno dos nódulos, valores dos níveis de cinza e as posições de cada imagem que continham câncer. Foram calculados os valores de área que considera os valores de pixels de cada região que contem câncer, espessura que determina a distância máxima ponderada, circularidade que mede a taxa de ocupação no interior do circulo circunscrito de cada região que contem câncer, níveis de cinza que determina a média dos números CT de cada região que contem câncer, valores de variância que determinam a variância dos números CT em cada região que contem câncer e localização dos nódulos. Três regras foram criadas no auxilio da classificação do diagnóstico: Regra 1 – são eliminadas imagens que apresentem nódulos com área e circularidade pequena, valores de níveis de cinza e variância muito alta e existência de densidade óssea nas imagens que contem câncer. Regra 2 – as imagens tumorais sem contato com a parede do pulmão são quantificadas e determinadas suspeitas se forem circulares, possuir um tamanho grande e possuir brilho dependendo dos valores de níveis de cinza e variância. Regra 3 – tumores em contato com a parede do pulmão são considerados suspeitos. Após a criação e a aplicação do algoritmo em 450 dados de pacientes os resultados mostraram que o algoritmo detectou as sombras suspeitas de serem nódulos pulmonares. Contudo, existem alguns casos de falso positivos e falso negativos. Um caso de falso negativo foi encontrado fora da área do pulmão e os casos restantes são inferiores a 5mm. Há cerca de 11 casos de falso-positivos que pertencem a sombras de vasos sanguíneos, difíceis de distinguir.

Kawata e colaboradores (1997) desenvolveram um sistema computacional de análise quantitativa baseado em pequenos nódulos pulmonares com base na morfologia (área,

circularidade, densidade) e classificação do desenvolvimento de lesões pulmonares em cortes finos de imagens de TC. Algoritmos de extração das características foram projetados para extrair os parâmetros característicos de forma tridimensional (3D) de imagens de nódulos utilizando curvaturas de superfície e linha do cume. Os nódulos pulmonares têm números de TC mais elevados do que outras regiões do tórax e é extraído por uma técnica de limiar com um único número TC especificado pelo operador. Uma vez que os vasos sanguíneos e região de brônquios tem números de TC semelhantes a região do nódulo, eles são extraídos simultaneamente com o valor limite especificado. Os vasos sanguíneos e brônquios são removidos por meio de um processo de fusão de imagens 3D. Foi concluído que o aspecto importante deste método é a de que as características baseadas na morfologia quantificam pistas importantes para distinguir nódulos malignos de benignos em imagem 3D do tórax. Através da plotagem de nódulos em características de espaço múltiplo, é demonstrada a capacidade técnica de extração de características, mostrando uma boa evidencia de separação entre nódulos malignos de benignos.

Kawata e colaboradores (1998) desenvolveram um método para quantificar as características da superfície de pequenos nódulos pulmonares em superfícies bem definidas com base em imagens de TC. Foi usada uma técnica de segmentação automática baseada em curvaturas de superfície deformáveis, como as curvas Gaussianas médias. As características da superfície extraídas foram usadas para distinguir nódulos malignos de benignos. Foram analisados atributos topológicos, como as medidas de densidade, distribuição da densidade, diâmetro, área, forma do contorno dos nódulos e textura de lesões suspeitas de serem malignas ou benignas. Os resultados mostraram a viabilidade de plotagem dos recursos propostos para avaliar o potencial de malignidade ou benignidade de pequenos nódulos. Outros estudos estão em andamento para investigar características adicionais com respeito a intensidade e textura interno do nódulo.

Tozaki e colaboradores (1998) aplicaram curvas gaussianas, segmentação e técnicas de detecção de bordas que extraem automaticamente vasos pulmonares e brônquios. O método utilizado para análise foi: a extração das regiões de interesse (ROI) com base na segmentação pulmonar; a divisão entre brônquios e vasos sanguíneos; a identificação da artéria pulmonar e da veia pulmonar por meio de informações anatômicas e processamento de imagens 3D; e uma análise quantitativa das relações espaciais entre nódulos anormais e outros órgãos pulmonares. O método de análise foi realizado a partir da extração do pulmão através de algoritmos morfológicos e segmentação do pulmão, do fundo e das regiões de interesse (brônquio, artéria pulmonar e veia pulmonar). A análise quantitativa para o diagnóstico

diferencial mostrou a necessidade de avaliação mais ampla para determinar a sua utilidade na prática clínica.

McNitt-Gray e colaboradores (1999) desenvolveram uma ferramenta semiautomática que isola nódulos pulmonares em imagens de alta resolução de TC, baseado em características quantitativas da densidade de nódulos, distribuição da densidade, tamanho, forma do contorno dos nódulos e textura. O objetivo deste estudo foi determinar os efeitos dos parâmetros da formação da matriz sobre o desempenho das medidas de textura da matriz de co-ocorrência, na discriminação entre nódulos malignos e benignos, através da investigação detalhada das medidas de textura de Haralick na classificação de nódulos pulmonares solitários. Para tal foram investigados os efeitos da variação dos parâmetros utilizados para formar as matrizes de co-ocorrência na precisão da classificação dos nódulos pulmonares solitários (NPS). Para cada medida de textura individual, foi realizada a análise discriminante para determinar se os atributos eram relevantes, e em seguida para avaliar o desempenho do sistema utilizando a análise discriminante linear com as características específicas dos atributos. Foi concluído neste estudo que as medidas de textura quando usadas em combinações, são muito precisas na classificação de nódulos pulmonares solitários.

Kawata e colaboradores (2000) em continuidade de seus estudos (1998) desenvolveram métodos computadorizados para extração de características e classificação para analisar nódulos malignos e benignos em três dimensões (3D) de imagens do tórax. Este trabalho abordou características do histograma e da textura interna, que, na opinião dos autores, é uma das pistas mais importantes para a diferenciação entre nódulos malignos e benignos. Nesta abordagem, cada voxel foi descrito em termos de índice de formato de derivados de curvaturas do voxel. Os voxels dentro do nódulo foram analisados através de histogramas, de forma a quantificar o quanto estavam presentes no nódulo. Características topológicas foram introduzidas para caracterizar a morfologia do conjunto construído a partir de um conjunto de voxels com a mesma categoria de forma. As propriedades densidade e curvatura foram utilizadas em conjunto com os recursos do histograma. Um classificador linear discriminante foi designado para cada classe separadamente, baseado na estrutura interna dos atributos. Verificou-se que o melhor conjunto de recursos foi selecionado a partir do espaço topológico e recursos do histograma combinado. Foi comparado o desempenho da classificação computadorizada com a experiência dos médicos, concluindo que, o desempenho da classificação baseada na classificação no espaço de características combinadas atingiu o desempenho dos médicos experientes. Os resultados demonstraram a viabilidade do uso de

recursos topológicos e histograma para analisar a textura interna no auxílio aos médicos na tomada de decisões de diagnóstico.

Reeves e colaboradores (2000) verificaram que métodos computadorizados podem ser aplicados a três áreas básicas da radiologia diagnóstica referentes ao câncer de pulmão: visualização, detecção e caracterização. Os autores descreveram as ferramentas fundamentais para detecção de pequenos nódulos pulmonares que podem ser usados para medir com precisão a forma do contorno dos nódulos, tamanho e conseqüentemente a taxa de crescimento de nódulos detectados em tempo hábil. Neste estudo, os colaboradores verificaram que as taxas de crescimento determinadas através do volume dos tumores podem ser um preditor útil de malignidade em pequenos nódulos pulmonares. Teoriza-se que também pode ser possível estimar a probabilidade de nódulo malignos com base no tamanho, forma, densidade e parâmetros determinados a partir de um único exame. Os resultados iniciais deste estudo indicam que os sistemas CAD podem ser eficazes para melhorar o diagnóstico de câncer de pulmão, fornecendo ao radiologista um recurso mais preciso do que medições de tamanho.

Kawata e colaboradores (2003) utilizaram uma abordagem baseada em exemplos para auxiliar a tomada de decisão na classificação de nódulos pulmonares em três dimensões (3D) em imagens de tomografia computadorizadas que utilizam características morfológicas e perfis internos do nódulo. Através da análise das características dos nódulos centrais, pode extrair características de forma, circularidade e estrutura interna dos nódulos. Os eixos principais e a compacidade foram utilizados para caracterizar a forma dos nódulos. As estruturas adjacentes e internas foram representadas pelo padrão de distribuição do valor de número TC e índices da curvatura 3D. Este estudo excluiu nódulos maiores de 3 cm de diâmetro. Para análise, foram utilizadas a análise discriminante linear e uma rede neural artificial, para calcular a probabilidade de malignidade e buscar imagens semelhantes para casos desconhecidos. A avaliação preliminar desta abordagem mostrou que a abordagem foi eficaz para a tomada de decisões diagnósticas na classificação de nódulos pulmonares, usando um banco de dados de imagens de nódulos.

Arimura e colaboradores (2004) desenvolveram um sistema informatizado de detecção automática de nódulos pulmonares para rastreamento de câncer de pulmão. Este sistema foi baseado em uma técnica de diferença entre imagens desenvolvido para realçar nódulos. As regiões dos nódulos foram determinadas por meio de monitoramento das características morfológicas com uma técnica de crescimento de regiões para segmentação. Para cada região do pulmão segmentada foi realizada uma análise discriminante linear (LDA). O diâmetro de

cada nódulo foi definido pelo diâmetro de um círculo com a mesma área. O grau de circularidade foi definido pela fração da área de sobreposição da área do círculo. Em cada nível de cinza, a região do nódulo foi quantificada por meio características morfológicas (diâmetro efetivo, a circularidade, e irregularidade). Verificou-se que o sistema informatizado usado neste estudo atingiu uma sensibilidade de 83%.

Way e colaboradores (2008) desenvolveram um método de diagnóstico por computador para diferenciar nódulos pulmonares de malignos e benignos em tomografia computadorizada. A segmentação da imagem foi realizada através de um método de contorno 3D ativo. Além das características morfológicas, como tamanho, volume, borda, diâmetro, média, desvio padrão, valores máximo e mínimo de número TC e de textura (variância), os autores desenvolveram novas características de superfície de nódulos para caracterizar a suavidade da de nódulos pulmonares com bordas irregulares. Foram comparados o desempenho dos classificadores entre a análise discriminante linear (LDA) e uma máquina de vetor de suporte (SVM). Neste estudo, concluiu-se que a discriminação do sistema CAD entre cânceres primários de pulmão e nódulos benignos foi maior do que entre cânceres metastáticos e nódulos benignos por provável sobreposição entre o aparecimento de nódulos benignos e cânceres metastáticos.

Minami e colaboradores (2008) segmentaram regiões dos nódulos a partir de um volume isotrópico reconstruído a partir de cada uma das séries de imagem. São apresentadas relações entre as características de realce de contraste e tipos de nódulos. As características de reforço de contraste são quantificadas usando o número TC de histograma. O número TC dos histogramas no interior do nódulo segmentados foram calculados em séries de pré e pós contraste. A variação entre os dois histogramas foi calculada subtraindo o histograma da série pós-contraste da série pré-contraste, e dividindo a soma da frequência subtraída de cada compartimento pelo volume do nódulo segmentado na série pré-contraste. Os nódulos foram classificados em 5 tipos (α , β , γ , δ e ϵ) em função das características internas extraídas no número TC do histograma na série pré-contraste. Esta abordagem consiste em segmentar os nódulos, classificar com base no número TC do histograma e calcular o realce do nódulo. As características internas extraídas são de acordo com a média, desvio padrão, assimetria e diferença da distribuição interna do nódulo. As lesões malignas apresentam densidade alta em relação a lesões benignas devido ao aumento da vascularização de nódulos malignos. Os nódulos pré-realce foram categorizados em subtipos de acordo com o diâmetro do nódulo, verificou que o desempenho da classificação da característica de realce depende do tipo e do tamanho do nódulo.

O Quadro 4 apresenta um resumo dos atributos quantitativos, separados por categoria, utilizados nas pesquisas em sistemas computadorizados de diagnóstico de nódulos pulmonares em imagens de TC.

Observa-se que dentre os atributos, o volume é o mais utilizado nas pesquisas que buscam a classificação dos nódulos pulmonares, seguido de um atributo de intensidade, o número TC. Outros atributos recorrentes nos estudos são aqueles morfológicos, como a área, a forma do contorno e a circularidade.

Quadro 4. Atributos quantitativos em nódulos pulmonares, analisados por diferentes autores.

	Giger (1994)	Toshioka (1997)	Kawata (1997)	Kawata (1998)	Tozaki (1998)	McNitt-Gray (1999)	Kawata (2000)	Reeves (2000)	Kawata (2003)	Arimura (2004)	Way (2008)	Minami (2008)	Número de Autores
Atributos Morfológicos													
Diâmetro	X			X		X		X	X	X	X	X	8
Área	X	X	X	X						X			5
Perímetro	X												1
Volume								X			X		2
Forma do Contorno	X	X		X		X		X					5
Circularidade	X	X	X						X	X			5
Irregularidade										X			1
Compacidade									X				1
Espessura		X											1
Atributos de Intensidade													
Número TC (HU)	X		X	X		X	X	X			X	X	7
Níveis de Cinza	X	X								X			3
Histograma							X					X	2
Variância		X									X		2
Atributos de Textura													
Textura				X		X	X						3
Outros Atributos													
Localização	X	X											2
Taxa Crescimento								X					1

Fonte: Autora, 2014.

3.7 Classificadores: redes neurais, análise discriminante e regressão logística

Para a classificação de lesões pulmonares em imagens de tomografia computadorizada, costumam ser utilizados diversos métodos, dentre os quais a análise discriminante linear ou LDA (MCNITT-GARY et al, 1999; KAWATA et al, 2000; ARMATO et al, 2001; ARMATO et al, 2002; GURCAN et al, 2002; AOYAMA et al, 2003; KAWATA et al, 2003; WAY et al, 2008), a regressão logística (MCCARVILLE et al, 2006; CHEN et al, 2012; PATZ et al, 2013; NORDHOLM-CARSTENSEN et al, 2014; PERANDINI et al, 2014) e redes neurais (ASHIZAWA et al, 1999 ; CHEN et al, 2011; AKRAM et al, 2014).

A análise discriminante linear é uma técnica de estatística multivariante, cujo objetivo é determinar a qual grupo, dentre vários definidos *a priori*, pertence um dado indivíduo, com base em características observadas. Cada uma das características, variáveis independentes, contribui para a classificação. A análise discriminante combina as variáveis independentes em uma ou mais funções que determinam para cada indivíduo e índices de classificação. As funções são construídas de tal forma que os índices dos indivíduos de cada grupo se concentram em torno do índice médio do grupo, minimizando a superposição de índices de indivíduos de grupos diferentes (ALTAMAN et al, 1979).

A classificação logística, ou análise *logit*, é uma técnica estatística de análise multivariada, com base em um modelo de regressão logística, que estima a probabilidade de ocorrência de um evento, sendo apropriada para as situações nas quais a variável dependente é categórica e assume um entre dois resultados possíveis (binária), tais como: “benigno ou maligno” ou “normal ou anormal”. A técnica é projetada para problemas nos quais os padrões devem ser classificados em uma das duas classes. Quando a variável de resposta é binária, considerações teóricas e empíricas indicam que a função de resposta é muitas vezes curvilínea (HOSMER & LEMESHOW, 2000; RANGAYYAN et al, 1997). O objetivo da regressão logística é gerar uma função matemática cuja resposta permita estabelecer a probabilidade de uma observação pertencer a um grupo previamente determinado, em razão do comportamento de um conjunto de variáveis independentes (HOSMER & LEMESHOW, 2000).

A regressão logística apresenta certas vantagens em relação à análise discriminante linear, principalmente devido às suas suposições iniciais serem menos rígidas. A análise discriminante linear está baseada em pressupostos bastante restritivos, como a normalidade das variáveis independentes e a igualdade das matrizes de variância-covariância dos grupos de

interesse. Essas suposições podem não ser válidas em muitas situações, principalmente quando há variáveis independentes de natureza não métrica (HAIR et al., 1998).

A classificação também pode ser realizada através de redes neurais, sistemas computacionais construídos através de técnicas que procuram imitar o cérebro humano com suas conexões de elementos de processamento. O cérebro humano tem a capacidade de organizar sua estrutura funcional para a execução de tarefas de processamento muito mais rapidamente que o mais rápido computador digital construído. Uma rede neural artificial tem a capacidade de “aprender” com seu próprio uso, ou seja, produzir saídas adequadas para entradas que não estavam presentes durante o treinamento, de forma a alcançar um objetivo desejado. Essa capacidade de generalização torna possíveis as redes neurais resolverem problemas computacionais complexos. As redes neurais têm sido aplicadas como meios auxiliares de diagnóstico e fornecem um modelo de prognóstico para a análise de sobrevivência. Esta técnica tornou-se extremamente popular em termos de classificação e previsão de imagens médicas (AMBROSIO, 2002), mas ela exige um grande número de exemplos para o treinamento da rede.

Nesta dissertação, o método escolhido para a classificação das imagens dos nódulos de pulmão é a regressão logística utilizando as variáveis relacionadas a atributos extraídas das imagens do nódulos. Essa escolha deve-se às características binárias de classificação dos nódulos em duas classes (benignos e malignos) e às suposições de que os conjuntos de dados relacionados a cada atributo da imagem não precisam apresentar uma distribuição normal.

4 METODOLOGIA

O projeto foi desenvolvido a partir da implementação de um método de análise quantitativa computadorizada de exames de pacientes que realizaram acompanhamento da evolução de nódulos pulmonares por meio de tomografia computadorizada, a partir de bancos de dados. Foram analisadas imagens de alto contraste da região do tórax utilizando um protocolo de tórax de matriz 512 x 512, 120 kV, 241 mAs, pitch de 0,9, espessura de corte de 2,0 mm e resolução de 1,185 pixel/mm, para um equipamento de TC (Siemens, Somatom Emotion 16 canais) instalado no HSL. O anonimato dos pacientes foi preservado, visto que sua identificação não é necessária ao estudo.

Este estudo passou pelo Comitê de Ética e Pesquisa e foi aprovado, sob protocolo de número 12385713.4.0000.5336.

Para a análise quantitativa, os exames foram divididos em dois grupos: (1) imagens de nódulos benignos, identificados a partir de exames de tomografia computadorizada de pulmão de pacientes, avaliados clinicamente por radiologistas, e seus respectivos exames anatomopatológico; (2) imagens de nódulos malignos, identificados a partir de exames de tomografia computadorizada de pulmão de pacientes e seus respectivos exames anatomopatológicos.

4.1 *Seleção das Imagens*

Com a ajuda do Laboratório de Anatomia e Patologia do Hospital São Lucas da PUCRS, foram selecionados 33 pacientes com histórico de adenocarcinoma de grau II e III e carcinoma epidermóide, para análise e seleção das imagens de nódulos malignos e 18 pacientes com histórico de hamartoma, lipoma e condroadenoma, para análise e seleção das imagens de nódulos benignos. Cada exame foi analisado para verificar se a imagem era compatível com este estudo, pois as imagens de interesse eram nódulos menores que 3 cm, passíveis de realização da volumetria e que não estivessem envoltos em estrias atelectásicas, pois isso dificultaria a segmentação das imagens e o cálculo da volumetria.

4.2 Atributos Qualitativos na avaliação do diagnóstico

Para um levantamento preliminar dos atributos considerados relevantes para o diagnóstico das lesões, foi realizado um estudo com três médicos radiologistas, no qual cada médico realizava uma avaliação do grau de importância de uma série de atributos no diagnóstico de malignidade de nódulos pulmonares, conforme a sua experiência. Todos os radiologistas envolvidos no estudo possuíam experiência superior a 5 anos. Cada radiologista avaliou cada atributo utilizando uma escala de importância: de 1, para os atributos que eram considerados menos importantes para o diagnóstico de malignidade de nódulos pulmonares, a 5, para os atributos que eram considerados mais importantes.

A escolha e o levantamento dos atributos avaliados foram realizados de acordo com as descrições da literatura e através de contato da autora com os radiologistas.

Posteriormente, os resultados desse levantamento foram comparados aos atributos quantitativos considerados mais relevantes na análise estatística.

4.3 Interpretação das Imagens pelos Radiologistas

A seguir foi realizado um estudo cego de interpretação das imagens de TC de pulmão com quatro radiologistas, sendo dois deles com experiência superior a 9 anos, e dois residentes, com experiência entre 3 e 5 anos.

Para o estudo cego, os exames dos 51 pacientes com nódulos foram reordenados aleatoriamente e anonimizados. Os médicos não tiveram acesso à história clínica e ou ao diagnóstico que determinava a malignidade dos nódulos. Foi solicitado, a cada médico, que as imagens fossem analisadas e cada nódulo fosse classificado como benigno ou maligno, analisando apenas as características visíveis das imagens e a idade dos pacientes. A análise foi realizada nas estações de trabalho utilizadas para interpretação no serviço de radiodiagnóstico e foi permitido o uso das ferramentas de visualização (zoom, brilho e contraste) e análise (valor médio e desvio padrão de uma região de interesse).

Foi criada uma tabela, para que cada médico pudesse realizar a classificação do nódulo em maligno, benigno ou 'não sei'. A seguir, as respostas de cada médico foram comparadas com o padrão ouro, obtido através dos laudos anatomopatológicos.

Considerando o diagnóstico realizado pelos radiologistas e o padrão ouro, foi determinada a concordância (confiabilidade e acurácia) das interpretações dos médicos e entre os médicos. Para tanto, foi utilizada a medida Kappa, que é uma medida de concordância

entre observadores e mede o grau de concordância além do que seria esperado tão somente pelo acaso (PERROCA e GAIDIZINSK, 2003). Esta medida de concordância tem como valor máximo 1 (um), que representa total concordância; os valores próximos ou abaixo de zero, indicam nenhuma concordância, ou que a concordância foi exatamente igual à esperada devido ao acaso. Segundo Landis e Koch (1977), como regra, coeficientes Kappa entre 0,21 e 0,40 referem-se a concordâncias consideradas razoáveis, entre 0,41 e 0,60 são concordâncias moderadas. Coeficientes Kappa com valores entre 0,61 e 0,80 representam uma concordância substancial e acima de 0,80 considera-se a concordância excelente.

O interesse deste tipo de análise estatística é determinar o grau de concordância individual da interpretação de cada médico radiologista com o laudo anatomopatológico, além do que seria esperado pelo acaso. Posteriormente, utilizou-se o mesmo coeficiente Kappa para avaliar o grau de concordância entre o modelo de classificação proposto neste trabalho e o laudo anatomopatológico.

4.4 *Análise do Desempenho do Modelo na Classificação dos Nódulos*

Para descrever quantitativamente o desempenho de um modelo de classificação dos nódulos, pode-se utilizar a curva ROC (do inglês, *Receiver Operating Characteristic*). Ela é calculada a partir da sensibilidade e a especificidade para todas as observações da amostra. (METZ et al, 1986).

A sensibilidade (S_E) é definida como a probabilidade do modelo sob investigação fornecer um resultado positivo (maligno), dado que o indivíduo é realmente portador da enfermidade. A sensibilidade é calculada por meio da Equação (30) (MARTINEZ et al, 2003).

$$S_E = \frac{VP}{VP + FN} \quad (30)$$

onde VP são considerados os casos em que o teste diagnóstico é verdadeiro positivo, FN são considerados os casos em que o teste é falso negativo.

A especificidade (E_S) é definida como a probabilidade do modelo fornecer um resultado negativo (benigno), dado que o indivíduo está livre da enfermidade. A Equação (31) apresenta o cálculo da especificidade (MARTINEZ et al, 2003).

$$E_S = \frac{VN}{FP + VN} \quad (31)$$

onde VN são considerados os casos em que o teste é verdadeiro negativo e FP são considerados os casos em que o teste é falso positivo.

A especificidade e a sensibilidade são calculadas sobre os mesmos indivíduos, ou seja, no cálculo da S_E utiliza-se apenas os doentes, e no cálculo da E_S utilizam-se somente os não doentes, sendo medidas independentes entre si (MARTINEZ et al, 2003).

O verdadeiro estado de cada nódulo, maligno ou benigno, é determinado por um teste de referência conhecido como padrão ouro, que no caso deste estudo, é o laudo anatomopatológico.

Muitos testes diagnósticos não produzem resultados expressos como mostram as equações 30 e 31, mas produzem uma resposta sob forma de uma variável ordinal ou contínua. Nesse caso, a classificação do indivíduo é baseada em uma regra de decisão, que busca um ponto de corte, de forma que a decisão é resumida em uma resposta dicotômica (binária). Assim, um indivíduo com medidas menores ou iguais ao ponto de corte (*cutoff point*) é classificado como não doente e, um indivíduo com resposta maior do que o ponto de corte é classificado como doente ou vice-versa. Para diferentes pontos de cortes, para uma amplitude de valores possíveis que o teste produz, pode-se estimar a sensibilidade e especificidade utilizando a chamada curva ROC (MARTINEZ et al, 2003).

A curva ROC é obtida registrando em um gráfico a sensibilidade no eixo da abscissa e $(1 - \text{especificidade})$ no eixo da ordenada, para os diversos pontos de corte. A área sob a curva mede a capacidade de discriminação do modelo. Hosmer e Lemeshow (2000, p.162) apresentam uma regra geral para avaliação do desempenho geral através da determinação área sob a Curva ROC: a discriminação é aceitável se a área estiver no intervalo entre 0,7 e 0,8; a discriminação é excelente se a área estiver no intervalo entre 0,8 e 0,9; e a discriminação é excepcional para se a área sob a curva ROC for maior do que 0,9.

4.5 Segmentação dos Nódulos

Para extração dos atributos quantitativos dos nódulos é necessária uma etapa de segmentação, ou seja, a criação de uma imagem na qual apenas a região do nódulo é diferente de zero, ou seja, uma imagem do nódulo sem o fundo.

Diversos testes de segmentação da região do nódulo nas imagens de TC foram realizados com as ferramentas disponíveis na Estação de Trabalho da General Electric HealthCare (Sun 4.6).

Após analisar visualmente cada corte, as imagens de TC e identificar quais os cortes que possuíam nódulos pulmonares, optou-se pela segmentação por meio da delimitação manual do contorno do nódulo. Foram utilizados alguns aplicativos de segmentação semiautomática dos nódulos, mas devido às características das imagens, os resultados não foram consistentes. Dessa forma, optou-se pela segmentação manual de todos os nódulos. As imagens dos nódulos segmentadas foram analisadas para extração dos atributos quantitativos.

4.6 Extração dos Atributos Quantitativos

Para a extração dos atributos quantitativos foram utilizadas diversas ferramentas: a ferramenta de análise de regiões de interesse da estação de trabalho disponível no serviço, utilizada para calcular alguns atributos morfológicos e de intensidade (volume, valores do número TC, desvio padrão e o número do corte que corresponde ao centro geométrico da região do nódulo na imagem); uma ferramenta desenvolvida no programa MATLAB®, por Nunes (2014), cedida e adaptada para uso neste trabalho; e o software de domínio público ImageJ®.

4.6.1 Atributos Morfológicos

Os atributos morfológicos calculados neste trabalho foram: perímetro, área, volume, compacidade, irregularidade e os momentos invariantes, que são calculados a partir dos momentos centrais. Os atributos de perímetro, área, volume, compacidade e irregularidade foram calculadas com o auxílio da estação de trabalho e do programa ImageJ®, os momentos invariantes foram calculados através do programa desenvolvido em MATLAB®, conforme descrição do fluxograma apresentado na Figura 11.

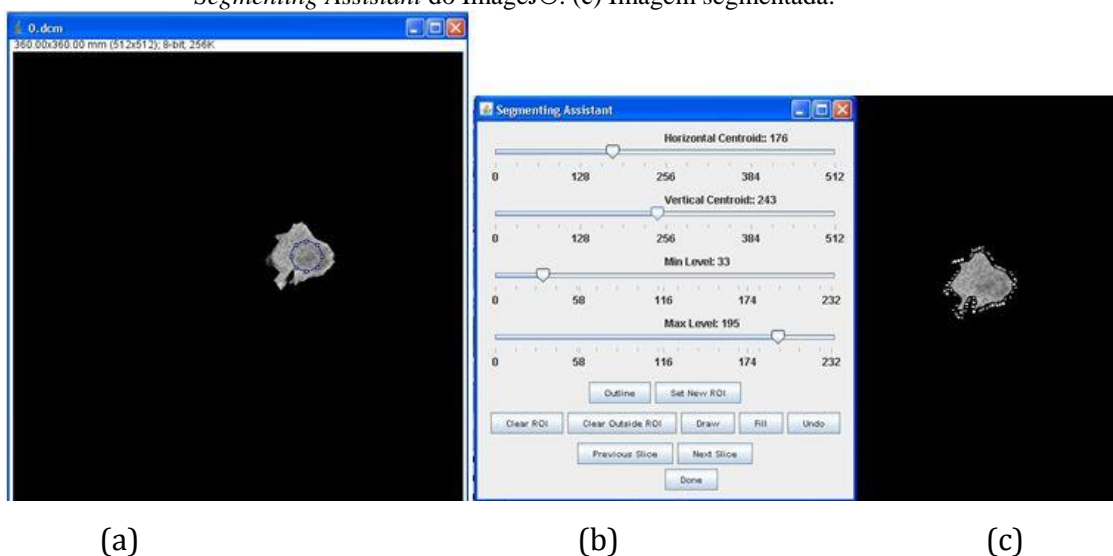
Os momentos invariantes estão relacionados à variância e distribuição da massa do objeto nas direções x e y (Φ_1 , Φ_2), a simetria do objeto em relação aos eixos x e y (Φ_3 , Φ_4), e a localização do centro de massa do objeto (Φ_5 , Φ_6 , Φ_7). Suas equações permitem que o cálculo desses atributos seja invariante em relação à posição do objeto, seu tamanho e orientação, conforme descrito na seção.

A ferramenta disponível na estação de trabalho permite a obtenção do volume do nódulo segmentado manualmente, assim como a média e desvio padrão dos valores de número TC (HU) da imagem central do nódulo. Esta imagem foi salva em DICOM, para que posteriormente fosse calculado os valores de área e perímetro utilizando o *plugin Segmenting*

Assistant disponível no programa ImageJ®. Todas as imagens foram analisadas em 16 bits por pixel.

Para o cálculo dos valores de área e perímetro, uma ROI (do inglês, *Region of Interest*), foi posicionada na imagem central do nódulo, e, através do plugin *Segmenting Assistant* (Figura 10) ImageJ®, verificou-se a distribuição de níveis de cinza, ajustando a semente nas centróides horizontais, verticais e os níveis máximos e mínimos dos níveis de cinza que compõem a imagem. Através da ferramenta *Measure* do ImageJ®, é possível verificar os resultados dos valores da área e perímetro que posteriormente são utilizados para os cálculos de compacidade e irregularidade.

Figura 10 - (a) Exemplo de imagem na qual identifica-se uma ROI no centro de um nódulo. (b) Plugin *Segmenting Assistant* do ImageJ®. (c) Imagem segmentada.



Fonte: Ref. Banco de imagens do Hospital São Lucas da PUCRS, 2013.

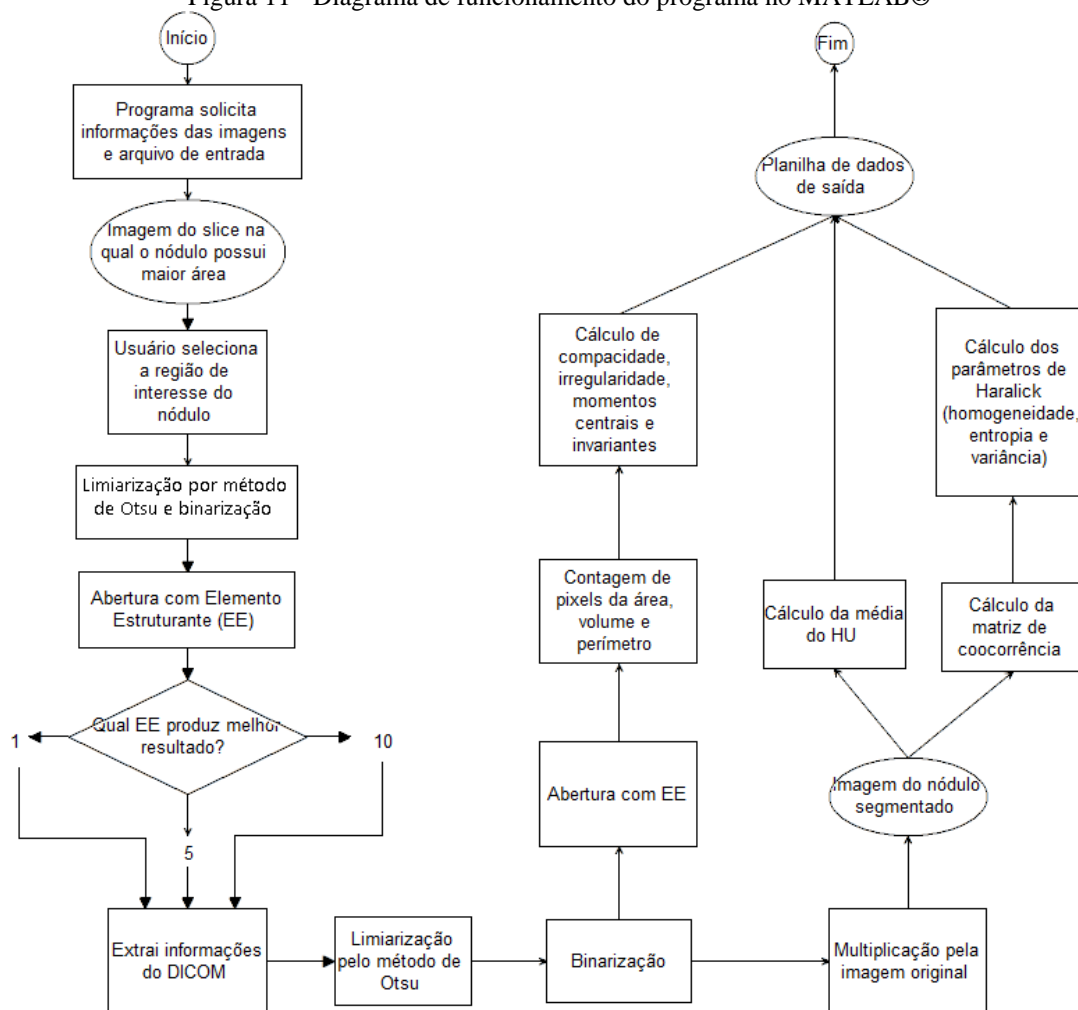
Para calcular os valores de compacidade e irregularidade, foram utilizadas as equações (3) e (4), descritas na seção 1.5.2.

A volumetria foi realizada manualmente, através da estação de trabalho da GE, delimitando o nódulo, corte a corte nas imagens que continham o nódulo pulmonar. A segmentação manual foi necessária, pois muitos nódulos encontravam-se próximos a parede do tórax ou próximos a vasos pulmonares, dificultando a segmentação automática do volume. Deve-se destacar que o problema da segmentação automática ou semi-automática dos nódulos em imagens de TC de pulmão não era o escopo deste trabalho, e trata-se de tema de pesquisa específico e de grande complexidade.

A Figura 11 mostra um diagrama do programa MATLAB, desenvolvido por Nunes (2014) para análise de atributos de imagens de TC de pulmão. O referido programa foi adaptado para a utilização neste trabalho.

O programa funciona da seguinte forma: o aplicativo solicita informações das imagens e do arquivo de entrada DICOM, que são inseridas pelo usuário; a seguir é necessário informar a imagem do corte na qual o nódulo possui maior área; o passo seguinte exige que o usuário informe os cortes que contém as imagens do nódulo; após uma limiarização automática pelo método de Otsu, a imagem é binarizada para produzir uma máscara. Para que a segmentação seja adequada, é aplicada uma operação de abertura morfológica com a opção de modificação do tamanho do elemento estruturante (EE). O usuário pode visualizar as imagens segmentadas após a abertura com diferentes EE (que pode ser igual a 1, 5 ou 10) e escolher a opção que segmenta melhor o nódulo para a formação de uma máscara, que é aplicada sobre a imagem do nódulo. A partir da extração de informações do *header* DICOM, é realizada a contagem dos pixels das áreas e perímetros de todos os cortes do nódulo, o volume total é calculado e os seguintes atributos: compacidade, irregularidade e momentos centrais e invariantes. Todos os dados são transferidos para uma planilha eletrônica de dados Microsoft Excel®. A partir da multiplicação da imagem original do nódulo segmentado com a máscara binária, é realizado o cálculo da média dos números CT de cada corte, em HU. O aplicativo realiza o cálculo da matriz de coocorrência para uso nos cálculos dos parâmetros de textura de Haralick, tais como homogeneidade, entropia e variância, que são transportados para a planilha de dados de saída.

Figura 11 - Diagrama de funcionamento do programa no MATLAB®



Fonte: Dartora, 2014.

4.6.2 Atributos de Intensidade

Os atributos de intensidade calculados neste trabalho são: os valores de média de número TC em HU, a massa, a densidade e dois índices relacionados à presença de gordura ou calcificações nos nódulos. Para o cálculo destes atributos foram utilizados a ferramenta desenvolvida em MATLAB®, descrita na seção anterior e o ImageJ®.

De posse dos valores de média e desvio padrão do número TC, é possível calcular o valor da massa e densidade do nódulo. Como a atenuação dada pelo número TC em HU, segundo Hoop et al (2010), está diretamente relacionada à densidade física, (adiciona-se 1000 ao valor de HU), pode-se obter o valor da densidade e calcular sua massa multiplicando o valor do volume encontrado pela densidade.

Para contemplar algumas características apontadas pelos radiologistas e estudos sobre a densidade de nódulos pulmonares (SIEGELMAN et al., 1980; PATEL et al., 2013),

relevantes na diferenciação dos nódulos, tais como a presença de calcificações ou gordura como indicador de malignidade ou benignidade, foram criados dois atributos de intensidade: índice de gordura (IG), dado pela razão entre a área de gordura e a área total do nódulo; e índice de calcificação (IC), dado pela razão entre a área de calcificação no nódulo e a área total do nódulo. Os pixels de gordura e calcificação no nódulo foram obtidos a partir de uma limiarização simples. Para a segmentação da gordura, a limiarização é realizada para identificar os pixels com número TC entre -80HU e -100 HU e para a segmentação da calcificação, os pixels foram limiarizados em valores de número TC acima de 100 HU.

4.6.3 Atributos de Textura

Os atributos de textura calculados neste trabalho são: a homogeneidade, a entropia e a variância. Esses atributos foram calculados a partir das matrizes de co-ocorrência, a qual descreve a probabilidade de ocorrer um par de pixels (i, j) em um ângulo θ . Neste trabalho foi utilizada a média dos atributos de textura em quatro ângulos (0° , 90° , 180° , 270°). Para os cálculos desses atributos foi utilizado a ferramenta desenvolvida em MATLAB®, descrita na seção anterior.

4.7 Construção de um Modelo para Classificação dos Nódulos

Para a construção do modelo de classificação dos nódulos foram analisados 20 atributos quantitativos extraídos das imagens de TC de pulmão, frequentemente utilizados em estudos dessa natureza, 2 variáveis qualitativas ou categóricas (sexo e localização do nódulo) e 1 variável quantitativa relacionada à idade do paciente.

Os atributos quantitativos dos nódulos extraídos das imagens de tomografia computadorizada de pulmão compreenderam: 12 (doze) atributos morfológicos (volume, área, perímetro, compacidade, irregularidade e os momentos invariantes de ordem 1 a 7), 5 (cinco) atributos de intensidade (massa, densidade, média do número TC, índice de gordura e índice de calcificação) e 3 (três) atributos de textura (homogeneidade, entropia e variância). O Quadro 5 exibe todos os 20 atributos quantitativos das imagens utilizados neste estudo, assim como as equações para seu cálculo.

Quadro 5. Atributos analisados com vistas à construção de um modelo de classificação dos nódulos pulmonares.

Nome do Atributo	Símbolo	Definição
Volume	V	Volume total do nódulo segmentado
Área	A	Área máxima do nódulo segmentado em um corte
Perímetro	P	Perímetro do nódulo segmentado em um corte
Compacidade	C	$C = P^2/(4\pi A)$
Irregularidade	I	$I = A/(P^2)$
Momento Invariante de ordem 1	MI1	$\Phi_1 = n_{20} + n_{02}$
Momento Invariante de ordem 2	MI2	$\Phi_2 = (n_{20} - n_{02})^2 + 4n_{11}^2$
Momento Invariante de ordem 3	MI3	$\Phi_3 = (n_{30} - 3n_{12})^2 + (3n_{21} - n_{03})^2$
Momento Invariante de ordem 4	MI4	$\Phi_4 = (n_{30} + n_{12})^2 + (n_{21} + n_{03})^2$
Momento Invariante de ordem 5	MI5	$\Phi_5 = (n_{30} - 3n_{12})(n_{30} + n_{12})[(n_{30} + n_{12})^2 - 3(n_{21} - n_{03})^2] + (3n_{21} - n_{03})(n_{21} + n_{03})[3(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2]$
Momento Invariante de ordem 6	MI6	$\Phi_6 = (n_{20} - n_{02})[(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2] + 4n_{11}(n_{30} + n_{12})(n_{21} + n_{03})$
Momento Invariante de ordem 7	MI7	$\Phi_7 = (3n_{21} - n_{03})(n_{30} + n_{12})[(n_{30} + n_{12})^2 - 3(n_{21} + n_{03})^2] + (3n_{12} - n_{30})(n_{21} + n_{03})[3(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2]$
Média Número TC	HU	Média do número TC do nódulo, em HU
Densidade	ρ	$[1000 \cdot \text{Número TC (HU)}] \text{mg/cm}^3$
Massa	M	$m = \rho \cdot V$
Índice de Gordura	IG	Razão entre o número de pixels com $-100 \text{ HU} < \text{Número TC} < -80 \text{ HU}$ e o número de pixels total do nódulo
Índice de Calcificação	IC	Razão entre o número de pixels com Número TC $> 100 \text{ HU}$ e o número de pixels total do nódulo
Homogeneidade	Hom	<i>Homogeneidade:</i> $f_1 = \sum_i \sum_j \{p(i,j)\}^2$
Entropia	Ent	<i>Entropia:</i> $f_9 = - \sum_i \sum_j p(i,j) \log_2 p(i,j)$
Variância	Var	<i>Variância:</i> $f_4 = \sum_i \sum_j (i - \mu)^2 p(i,j)$

Fonte: Autora, 2014.

Toda estatística da análise das variáveis para o conjunto de nódulos malignos e benignos foi realizada por meio da técnica de regressão logística (HOSMER & LEMESHOW, 2000), utilizando o programa estatístico SPSS (*Statistical Package for Social Science for Windows*).

Um aspecto que favorece a utilização da regressão logística é que seus resultados podem ser interpretados em termos de probabilidade. Esse fator se mostra particularmente importante neste trabalho, pois possibilita que seja medida, por exemplo, a probabilidade de um determinado nódulo assumir a condição de maligno ou benigno.

Os coeficientes da regressão (b_i) são estimados pelo método da máxima verossimilhança. O modelo da regressão logística assume a seguinte relação:

$$\ln[p/(1-p)] = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k \quad (32)$$

onde p é a probabilidade de ocorrer o evento (nódulo maligno); $(1 - p)$ é a probabilidade de não ocorrer o evento (nódulo benigno); $p/(1 - p)$ a razão de probabilidades; X_i são as variáveis independentes (atributos) e b_i são os coeficientes estimados para cada atributo.

Os coeficientes estimados pelo modelo de regressão logística indicam a importância de cada variável independente para a ocorrência do evento, que neste caso será a malignidade ou benignidade do nódulo.

Na regressão logística, os coeficientes medem o efeito das alterações nas variáveis independentes sobre o logaritmo natural da razão de probabilidades, chamado *logit*. Para avaliar o impacto dos parâmetros sobre a probabilidade de ocorrer o evento, eles devem ser transformados por meio do inverso do logaritmo. A probabilidade P associada à ocorrência do evento de interesse pode ser obtida pela equação 33:

$$P = [1 + e^{-(b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k)}]^{-1} \quad (33)$$

Na regressão logística, para a seleção das variáveis independentes que serão utilizadas no modelo, realiza-se uma análise univariada prévia das variáveis, neste caso, cada um dos 20 atributos quantitativos calculados a partir das imagens de TC de pulmão, a idade e as variáveis categóricas (sexo e localização do nódulo).

Uma variável qualitativa ou categórica é incluída no modelo de classificação a partir da análise do teste de χ^2 (*Chi-square*) de associação entre a variável e o diagnóstico de malignidade ($p < 0,05$) (COSTA & NADANOVSKY, 2005), e uma análise de regressão logística univariada (HOSMER & LEMESHOW, 2000).

O critério para inclusão de uma determinada variável quantitativa independente no modelo de regressão linear múltipla é que, na regressão univariada, o seu valor de

significância $p \leq 0,25$, segundo Hosner e Lemershow (2000, p. 95). Assim, se o valor de p de uma variável é maior do que 0,25, a probabilidade da associação da variável com o diagnóstico estar incorreta é o valor de p , ou seja, mais que 25%.

Para alguns atributos bidimensionais (compacidade, irregularidade, momentos invariantes e atributos de textura) foi analisada a relevância da utilização de um corte específico do volume de imagens. Assim, cada um desses atributos foi estatisticamente avaliado no corte central dos nódulos, nos cortes anteriores e posteriores, a um quarto e três quartos do comprimento total do nódulo na direção longitudinal, e no corte de maior área. Para verificar se havia diferença significativa na inclusão de um mesmo atributo avaliado em diferentes cortes no modelo de classificação, cada atributo foi submetido a uma análise estatística não paramétrica do tipo Kruskal-Wallis. O nível de significância utilizado para esses testes foi de 0,05 (COSTA, NADANOVSKY, 2005).

Finalmente, para avaliar o desempenho do modelo obtido da regressão logística, foi utilizada a curva ROC e a medida da área sob a curva, que mede a capacidade de discriminação do modelo (HOSMER & LEMESHOW, 2000).

5 RESULTADOS E DISCUSSÃO

5.1 *Atributos Relevantes para o Diagnóstico Médico*

A partir de escores de dois radiologistas com experiência superior a 5 anos, foram determinados os atributos com maior grau de importância do diagnóstico de malignidade de nódulos pulmonares em imagens de TC. Para cada atributo, um escore de importância foi atribuído pelos radiologistas: de 1 a 5, onde 1 seriam os atributos que fossem considerados menos importantes e 5 seriam os atributos mais importantes. O Quadro 6 mostra os resultados das médias dos escores para cada atributo, obtido a partir deste levantamento. Os escores finais assinalados nas células em caixa cinza correspondem aos valores escore médio maior do que 3, o que corresponde a atributos considerados mais relevantes para todos os médicos.

O atributo com maior escore (5) para classificação da malignidade do nódulo, unânime entre os médicos, foi a velocidade de crescimento do tumor. Este resultado concorda com os resultados de Reeves e colaboradores (2000), que verificaram que a taxa de crescimento determinada através do volume do tumor pode ser um preditor útil de malignidade em pequenos nódulos pulmonares. Como este estudo não acompanhou o mesmo paciente ao longo do tempo, esse atributo não foi considerado na análise posterior.

A seguir, em nível de importância, menor do que 5 e maior do que 3,5, estão: a volumetria, o diâmetro, a circularidade do nódulo, o halo de infiltração periférico e as calcificações. O volume do nódulo é o atributo mais estudado nas pesquisas de análise quantitativa de lesões pulmonares, como mostrado na seção de revisão de literatura.

Finalmente, para um nível de importância menor do que 3,5 até 3,0 estão: a massa, a intensidade média (número TC) e a forma do contorno; também considerados pelos pesquisadores, atributos de importância na classificação de lesões pulmonares.

Esta avaliação qualitativa dos médicos da importância dos atributos para a classificação da malignidade será confrontada posteriormente com os resultados do modelo de classificação construído a partir da análise estatística dos atributos quantitativos das imagens de TC e variáveis categóricas.

Quadro 6. Classificação dos atributos pelos radiologistas de acordo com o grau de importância para a classificação dos nódulos

Radiologista	Localização do nódulo	Volumetria	Densidade (g/cm ³)	Massa	Intensidade Média (HU)	Área	Perímetro	Diâmetro	Circularidade	Textura interna	Forma do Contorno	Halo de infiltração periférico	Calcificações	Velocidade de crescimento
A	2	5	4	4	4	4	4	4	4	4	3	3	3	5
B	2	3	1	1	5	3	3	3	4	3	4	3	4	5
C	1	3	3	3	3	1	1	2	3	4	5	4	3	5
Média	1,7	3,7	2,7	2,7	4,0	2,7	2,7	3,0	3,7	3,7	4,0	3,3	3,3	5,0
Desvio Padrão	0,6	1,2	1,5	1,5	1,0	1,5	1,5	1,0	0,6	0,6	1,0	0,6	0,6	0,0

Fonte: Autora, 2015.

5.2 Interpretação das Imagens pelos Radiologistas

A seguir estão os resultados do estudo cego de interpretação das imagens de TC de pulmão com quatro médicos radiologistas, denominados como Médicos 1 a 4. O Padrão Ouro refere-se ao laudo anatomopatológico, que define se o nódulo é benigno ou maligno. Os casos válidos são considerados como aqueles nos quais o médico responde que o nódulo é benigno ou maligno, enquanto os casos não válidos referem-se às respostas ‘não sei’.

O Quadro 7 mostra o número de respostas válidas e não válidas para cada médico. Os médicos 1 e 4 tratam-se de radiologistas mais experientes, com 10 e 9 anos de experiência, respectivamente, apresentando um percentual mais elevado de respostas válidas. Este fato revela a importância da experiência médica na interpretação das imagens radiológicas, que permite uma maior número de casos no qual o médico sente-se seguro em fornecer um diagnóstico.

Quadro 7. Número e percentual de casos válidos e não válidos para cada médico avaliador.

	Casos					
	Válidos		Não Válidos		Total	
	N	Percentual	N	Percentual	N	Percentual
Padrão Ouro * Médico 1	47	92,2%	4	7,8%	51	100,0%
Padrão Ouro * Médico 2	39	76,5%	12	23,5%	51	100,0%
Padrão Ouro * Médico 3	41	80,4%	10	19,6%	51	100,0%
Padrão Ouro * Médico 4	41	80,4%	10	19,6%	51	100,0%

Fonte: Autora, 2015.

Para analisar a concordância entre as interpretações dos médicos e o laudo anatomopatológico dos nódulos foi utilizado o coeficiente Kappa. Este coeficiente possibilita a verificação do grau de correspondência, além do que seria esperado pelo acaso, entre as avaliações independentes dos quatro médicos radiologistas que classificaram um mesmo nódulo, utilizando o mesmo instrumento, neste caso, a experiência na interpretação visual dos nódulos em imagens de TC de pulmão.

O Quadro 8 apresenta os coeficientes Kappa e os respectivos intervalos de confiança (95%) para os quatro médicos radiologistas. A última coluna apresenta a acurácia total, determinada a partir do cálculo de $(VP+VN)/N$ de cada médico, onde N é o número de casos válidos (excluídas as respostas ‘não sei’). Deve-se ressaltar que a acurácia não distingue a probabilidade de fornecer as respostas certas devido ao acaso.

Observa-se que apenas o médico 1, com 10 anos de experiência, possui uma concordância significativa (entre 0,61 e 0,80) com o laudo anatomopatológico. O médico 4, com 9 anos de experiência, apresenta uma concordância moderada (entre 0,41 e 0,60) e os médicos 2 (5 anos de experiência) e 3 (3 anos de experiência) apresentam diagnóstico com concordância fraca com o laudo, quando se excluem os acertos devido ao acaso.

No Quadro 8, abaixo do coeficiente Kappa, encontra-se o intervalo de confiança de 95% para cada médico, determinado a partir do erro padrão ($Kappa \pm 1,96*SE$) (COSTA, NADANOVSKY, 2005). Estes valores revelam a dificuldade dos médicos radiologistas em realizarem um laudo exclusivamente baseado apenas na interpretação de características visuais das imagens de TC de pulmão. Observa-se como a maior experiência na interpretação radiológica (médico 1) fornece resultados significativamente melhores de concordância do diagnóstico.

Quadro 8. Coeficientes Kappa e acurácia para os quatro médicos radiologistas.

		Benigno	Maligno	N (Casos Válidos)	Coeficiente Kappa	Acurácia Total
Médico 1	Benigno	11	5	16	0,653 [0,420; 0,886]	0,851
	Maligno	2	29	31		
Médico 2	Benigno	6	4	10	0,369 [0,051; 0,687]	0,744
	Maligno	6	23	29		
Médico 3	Benigno	5	7	12	0,396 [0,084; 0,708]	0,780
	Maligno	2	27	29		
Médico 4	Benigno	6	5	11	0,424 [0,110; 0,738]	0,780
	Maligno	4	26	30		

Fonte: Autora, 2015.

5.3 Perfil da Amostra

A seguir estão descritos os resultados da análise de frequência da idade dos pacientes, das variáveis categóricas independentes de sexo e localização do nódulo nos lóbulos pulmonares.

A amostra foi constituída de imagens de TC de pulmão de 51 pacientes, sendo 21 do sexo feminino (F) e 30 do sexo masculino (M). As idades variaram de 18 a 83 anos. Os nódulos pulmonares estão localizados em diferentes lóbulos pulmonares: lobo superior esquerdo (LSE), lobo superior direito (LSD), lobo inferior esquerdo (LIE), lobo inferior direito (LID) e lobo médio direito (LMD).

O Quadro 9 apresenta as estatísticas dos pacientes cujos laudos anatomopatológicos possuem nódulos malignos e benignos por faixa etária.

Quadro 9. Número de pacientes por faixa etária, para os nódulos benignos e malignos.

Faixa Etária	Número de Pacientes
≤ 18 anos	1
19-27	0
28-36	0
37-45	4
46-55	8
56-64	14
65-74	18
>74	6
Total	51

A média da idade dos pacientes com nódulos benignos nessa amostra foi de (63,1±12,0) e média da idade dos pacientes com nódulos malignos (60,8±12,9), com nível de p=0,548, ou seja, para esse amostra, a idade não está associada com a malignidade do nódulo. Desta forma, a variável ‘idade’ não foi incluída no modelo de classificação.

O Quadro 10 apresenta as estatísticas de sexo dos pacientes cujos laudos anatomopatológicos possuem nódulos malignos e benignos.

Quadro 10. Número de pacientes por sexo, para os nódulos benignos e malignos.

	F	M	Total
Benigno	9	9	18
Maligno	21	12	33
Total	30	21	51

No teste do χ^2 , o nível de significância para a associação entre a variável categórica ‘sexo’ e a malignidade do nódulo foi de p = 0,344 para a amostra estudada. Assim, pode-se afirmar que não há associação entre o sexo dos pacientes e a malignidade dos nódulos pulmonares neste conjunto de imagens.

O Quadro 11 apresenta as estatísticas da variável ‘localização’ dos nódulos benignos e malignos. As siglas no quadro significam: lobo superior esquerdo (LSE), lobo superior direito (LSD), lobo inferior esquerdo (LIE), lobo inferior direito (LID) e lobo médio direito (LMD).

Quadro 11. Número de pacientes por localização do nódulo nos lóbulos pulmonares, para os nódulos benignos e malignos.

	LID	LIE	LMD	LSD	LSE	Total
Benigno	3	6	2	3	4	18
Maligno	13	1	4	6	9	33
Total	16	7	6	9	13	51

Para a variável categórica independente ‘localização’, observou-se que há evidências de que os nódulos benignos estão associados ao LIE (lobo inferior esquerdo), pois 85,7% dos nódulos benignos estão nessa localização, com resíduo ajustado é 3,0 (maior do que 1,96).

No teste do χ^2 , o nível de significância para a associação entre a variável categórica ‘localização’ e a malignidade do nódulo foi de $p = 0,043$ ($p < 0,05$). Dessa forma, optou-se por conservar a variável ‘localização’ no modelo de regressão logística. Este resultado é coerente com os resultados de Winer-Muran (2006), que mostraram que 70% dos cânceres de pulmão estão localizadas nos lobos superiores e ocorrem com maior frequência no pulmão direito.

5.4 Análise dos Atributos Quantitativos das Imagens para o Modelo de Classificação

Na regressão logística, para a seleção das variáveis que seriam utilizadas no modelo de regressão multivariada, realizou-se inicialmente uma análise univariada de todas as variáveis independentes, neste caso, cada um dos 20 atributos de imagens quantitativos e a variável categórica ‘localização’. O critério para inclusão de uma determinada variável independente no modelo de regressão linear múltipla é que, na regressão univariada, o valor de significância seja menor ou igual a 0,25, segundo Hosner e Lemershow (2000, p. 95). Assim, neste trabalho, as variáveis cujo valor de p na regressão logística univariada foi maior do que 0,25 foram descartadas do modelo de classificação, pois a probabilidade de associação de tais variáveis com o diagnóstico seria incorreta em mais de 25% dos casos estudados.

Para alguns atributos bidimensionais (compacidade, irregularidade, momentos invariantes e atributos de textura) foi analisada a relevância da utilização de um corte específico do volume de imagens. Assim, cada um desses atributos foi estatisticamente avaliado no corte central dos nódulos, nos cortes anteriores e posteriores, a um quarto e três quartos do comprimento total do nódulo na direção longitudinal, e no corte de maior área. Para verificar se havia diferença significativa na inclusão de um mesmo atributo avaliado em diferentes cortes no modelo de classificação, cada atributo foi submetido a uma análise estatística não paramétrica do tipo Kruskal-Wallis. O nível de significância utilizado para esses testes foi de 0,05.

O Quadro 12 mostra os valores de significância de cada variável bidimensional avaliada no corte anterior (CA), corte central (CC), corte de maior área (CMA) e corte posterior (CP). A hipótese nula assume que os atributos calculados nos diferentes cortes provêm de regiões que apresentam as mesmas características e portanto, não apresentam diferenças significativas.

Quadro 12. Variáveis independentes bidimensionais analisadas pelo teste de Kruskal-Wallis para verificar a diferença significativa na inclusão de um mesmo atributo avaliado em diferentes cortes.

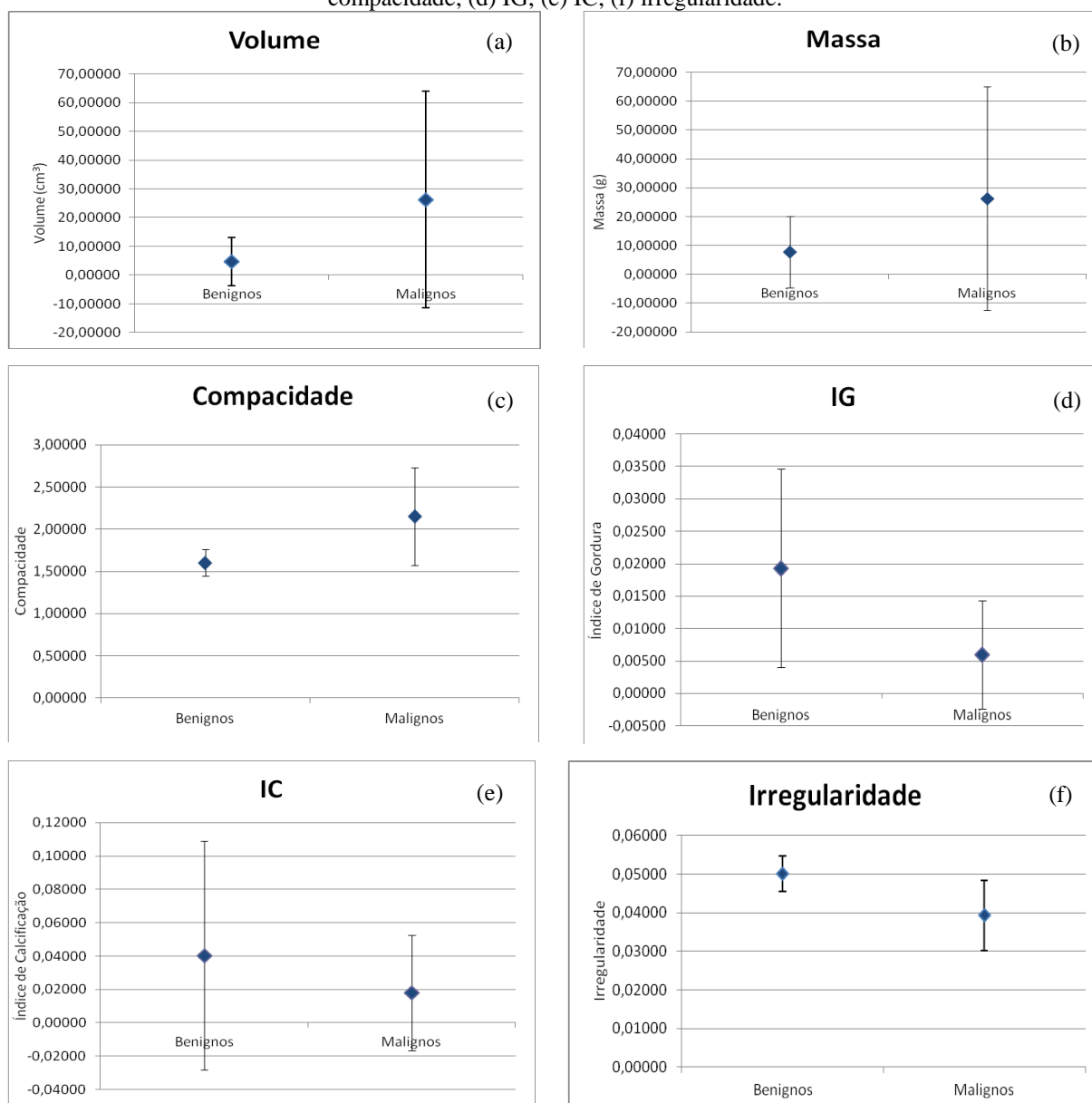
Variável Independente	Chi-Square	DF	p	Ho
Compacidade	0,91596	3	0,82157	Não há diferenças significativas
Irregularidade	0,92046	3	0,82049	Não há diferenças significativas
M01	0,15017	3	0,9852	Não há diferenças significativas
M02	0,49064	3	0,92094	Não há diferenças significativas
M03	0,96635	3	0,80939	Não há diferenças significativas
M04	2,39356	3	0,49483	Não há diferenças significativas
M05	1,85552	3	0,60293	Não há diferenças significativas
M06	4,14406	3	0,24632	Não há diferenças significativas
M07	5,03948	3	0,16893	Não há diferenças significativas
Homogeneidade	33,64413	3	$2,35527 \cdot 10^{-7}$	Há diferenças significativas
Entropia	31,77365	3	$5,84101 \cdot 10^{-7}$	Há diferenças significativas
Variância	29,98521	3	$1,38998 \cdot 10^{-6}$	Há diferenças significativas

Observa-se que não há diferença significativa entre os atributos morfológicos (compacidade, irregularidade e momentos invariantes), quando calculados nos diferentes cortes. Para os atributos de textura, há diferenças significativas entre os valores de cada corte, o que se deve possivelmente ao fato dos nódulos serem muito pequenos e ao cálculo desses atributos ser feito a partir da média em todas as direções. Considerando esses resultados e a facilidade de implementação computacional para a busca e análise do corte central, optou-se por utilizar todos atributos bidimensionais calculados apenas no corte central no modelo de regressão logística multivariada.

O Anexo A mostra os resultados finais da regressão logística univariada realizada com cada um dos 23 atributos analisados, sendo eles: sexo, idade, localização, 12 (doze) atributos morfológicos (volume, área, perímetro, compacidade, irregularidade e os momentos invariantes de ordem 1 a 7), 5 (cinco) atributos de intensidade (massa, densidade, média do número TC, índice de gordura e índice de calcificação), e 3 (três) atributos de textura (homogeneidade, entropia e variância), conforme saída do programa SPSS.

As figuras 12(a) a 12(f) mostram a média e desvio-padrão de algumas variáveis independentes quantitativas.

Figura 12 - Gráficos da média e desvio padrão das variáveis quantitativas: (a) volume; (b) massa; (c) compacidade; (d) IG; (e) IC; (f) irregularidade.



Observa-se que as variáveis possuem grandes valores de desvio-padrão, mostrando a grande variabilidade desses valores dentro da amostra de nódulos benignos e malignos.

Como a irregularidade está altamente correlacionada à compacidade ($r = -0,98$), na avaliação do modelo de classificação, optou-se por excluir a variável irregularidade, mantendo apenas a compacidade.

O Quadro 13 mostra o resumo das variáveis submetidas à regressão logística univariada do tipo *step-wise* e seu nível de significância. No caso dos atributos bidimensionais, o cálculo foi realizado apenas para o corte central do nódulo.

Quadro 13. Variáveis independentes analisadas no modelo de regressão logística. As variáveis com $p < 0,25$ foram incluídas no modelo de regressão logística multivariada (HOSNER & LEMERSHOW, 2000, p. 95).

Variável Independente	<i>p</i>	<i>Inclusão no Modelo</i>
Localização LIE	0,014	SIM
Volume	0,068	SIM
Densidade (g/cm ³)	0,308	NÃO
Massa (g)	0,070	SIM
Média Número TC (HU)	0,308	NÃO
Índice de Gordura	0,085	SIM
Índice de Calcificação	0,143	SIM
Área Corte Central (cm ²)	0,056	SIM
Perímetro Corte Central (cm)	0,018	SIM
Compacidade Corte Central	0,003	SIM
M01 Corte Central	0,250	NÃO
M02 Corte Central	0,433	NÃO
M03 Corte Central	0,595	NÃO
M04 Corte Central	0,583	NÃO
M05 Corte Central	0,510	NÃO
M06 Corte Central	0,782	NÃO
M07 Corte Central	0,573	NÃO
Homogeneidade Corte Central	0,647	NÃO
Entropia Corte Central	0,071	SIM
Variância Corte Central	0,117	SIM

As variáveis independentes com $p < 0,25$ foram incluídas no modelo de regressão logística multivariada (HOSNER & LEMERSHOW, 2000, p. 95).

Para avaliar o melhor modelo de classificação da malignidade dos nódulos a partir de seus atributos, foram avaliados diversas combinações de modelos por regressão logística múltipla, usando as seguintes variáveis ou atributos: localização LIE, volume, massa, índice de gordura, índice de calcificação, área, perímetro, compacidade, entropia e variância.

Para avaliar o desempenho de cada modelo, foi determinada e a medida da área sob essa curva, que mede a capacidade de discriminação do modelo (HOSMER & LEMESHOW, 2000). A escolha do melhor modelo de classificação levou em consideração não apenas a maior área da curva ROC, mas a inclusão de atributos que tivessem, em seu conjunto, os menores valores de p , ou seja, cuja probabilidade da associação das variáveis com o diagnóstico estar incorreta fosse a menor possível.

O Quadro 14 mostra os valores dos coeficientes estimados nos oito modelos avaliados pela regressão logística multivariada neste trabalho, com os respectivos atributos e seus níveis de significância. A última linha do quadro mostra a medida da área sob a curva ROC.

O Modelo 7 está destacado no Quadro 14 pois ele foi escolhido como aquele que melhor classifica os nódulos malignos, em virtude da combinação do melhor desempenho, dado pela área da curva ROC, e os melhores níveis de significância de todas as variáveis.

Quadro 14 - Valores dos coeficientes estimados (b_i) nos modelos avaliados pela regressão logística multivariada neste trabalho, com os respectivos atributos e seus níveis de significância.

Variável Independente	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Modelo 7*	Modelo 8
Localização LIE	-2,479 (p=0,139)	-1,900 (p=0,156)	-1,700 (p=0,205)	-2,003 (p=0,117)	-1,927 (p=0,101)	-2,142 (p=0,086)	-2,267 (p=0,066)	-2,052 (p=0,079)
Volume	-1,731 (p=0,204)	-						
Massa (g)	1,846 (p=0,186)	0,096 (p=0,425)						
Índice de Gordura	7,942 (p=0,957)	-74,042 (p=0,495)	-112,576 (p=0,220)	-30,309 (p=0,655)				
Índice de Calcificação	-40,342 (p=0,019)	-33,748 (p=0,032)	-29,046 (p=0,029)	-23,381 (p=0,035)		-24,316 (p=0,028)	-21,724 (p=0,036)	
Área (cm ²)	-0,019 (p=0,165)	-0,010 (p=0,263)						
Perímetro (cm)	0,169 (p=0,092)	0,090 (p=0,182)						
Compacidade	2,374 (p = 0,639)	6,363 (p=0,183)	10,032 (p=0,011)	7,351 (p=0,012)	5,465 (p=0,009)	7,683 (p=0,009)	7,773 (p=0,006)	5,634 (p=0,007)
Entropia	11,147 (p=0,266)	12,814 (p=0,140)	10,997 (p=0,168)	-4,359 (p=0,294)	-3,028 (p=0,403)	-4,882 (p=0,225)		
Variância	0,000 (p=0,052)	0,000 (p=0,032)	0,000 (p=0,037)					
Área Curva ROC	0,965	0,951	0,939	0,918	0,894	0,921	0,923	0,887

*O Modelo 7 está destacado pois foi escolhido como aquele que melhor classifica os nódulos malignos.

Após a regressão logística multivariada e os testes com modelos considerando diferentes conjuntos de atributos, o modelo escolhido que melhor classifica os nódulos malignos foi aquele que inclui as variáveis localização, índice de calcificação e compacidade foram consideradas relevantes. Este modelo (número 7 no Quadro 14) revelou um alto valor preditivo positivo (86,3%) para essa amostra.

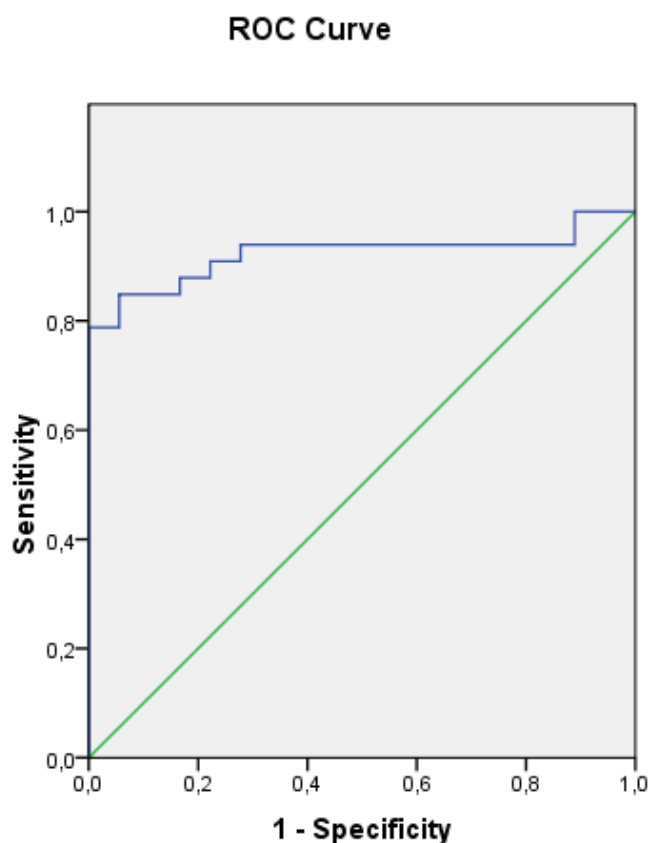
Assim, a função que representa o modelo resultante de classificação dos nódulos possui a seguinte equação:

$$P = [1 + e^{(-12,001 - 2,267.L - 21,724.IC + 5,886.C)}]^{-1}$$

onde P é o valor preditivo de classificação correta da malignidade do nódulo pulmonar, L é um termo binário (1 ou 0) que representa se a localização do nódulo está no LIE, IC é valor do índice de calcificação de todo volume do nódulo e C é o valor da compacidade calculada no corte central.

Para avaliar o desempenho do modelo foi construída a curva ROC (Figura 13) deste modelo de classificação, resultante deste trabalho.

Figura 13 - Curva ROC do modelo de regressão logística proposto neste trabalho



O Quadro 15 mostra o desempenho do modelo a partir da área da curva ROC do modelo de classificação proposto neste trabalho, bem como seu desvio padrão, significância e intervalo de confiança. Observa-se que a compacidade, um atributo morfológico relacionado à forma do nódulo é altamente preditivo da malignidade do nódulo, como confirmado por diversos trabalhos de classificação de nódulos pulmonares.

Quadro 15. Estatística da área da curva ROC do modelo proposto neste trabalho, seu desvio padrão, nível de significância e intervalo de confiança.

			Intervalo de 95% de confiança	
Área	Desvio Padrão ^a	Significância ^b	Mínimo	Máximo
0,923	0,040	0,000	0,843	1,000

a. Sob a hipótese não paramétrica

b. Hipótese nula: área verdadeira = 0,5

Observa-se que o modelo desenvolvido inclui um atributo qualitativo de localização, um atributo de intensidade relacionado ao índice de calcificação e um atributo morfológico, a compacidade. Todos esses atributos estão de acordo com a importância atribuída pelos médicos radiologistas na interpretação das imagens.

Apesar das limitações do modelo proposto, observa-se que o valor preditivo do modelo (86,3%) é muito maior do que o valor preditivo baseado na avaliação visual do médico mais experiente (65,3%), sem considerar o acaso, estando próximo da acurácia desse mesmo médico (85,1%), incluído o acerto devido ao acaso.

6 CONCLUSÕES

O objetivo geral deste estudo foi avaliar a utilização de dados de pacientes e atributos quantitativos de nódulos pulmonares em imagens de tomografia computadorizada de pulmão para a construção de um modelo de classificação em termos de malignidade.

Os atributos quantitativos, obtidos a partir de um banco de imagens com pacientes que foram para biópsia, possuíam o diagnóstico confirmado por meio da análise anatomopatológica do tecido pulmonar, permitindo uma classificação de malignidade e benignidade considerada verdadeira, que foi considerada neste estudo como o padrão ouro.

A amostra foi constituída de imagens de TC de pulmão de 51 pacientes, sendo 21 do sexo feminino (F) e 30 do sexo masculino (M). As idades variam de 18 a 83 anos.

O estudo foi dividido em uma etapa de interpretação diagnóstica realizada a partir de um estudo cego com médicos radiologistas e a utilização da regressão logística para análise do poder preditivo de variáveis qualitativas e quantitativas extraídas das imagens de tomografia computadorizada de pulmão.

Na análise do estudo cego realizado pelos radiologistas, observou-se que apenas o médico com maior experiência apresentou uma concordância com o laudo anatomopatológico ($Kappa = 0,645$), quando se excluem os acertos devido ao acaso.

Para a construção do modelo de classificação dos nódulos foram analisados 20 atributos quantitativos extraídos das imagens de TC de pulmão, frequentemente utilizados em estudos dessa natureza, 2 variáveis qualitativas ou categóricas (sexo e localização do nódulo) e 1 variável quantitativa relacionada à idade do paciente.

Os atributos quantitativos dos nódulos extraídos das imagens de tomografia computadorizada de pulmão foram constituídos de: 12 (doze) atributos morfológicos (volume, área, perímetro, compacidade, irregularidade e os momentos invariantes de ordem 1 a 7), 5 (cinco) atributos de intensidade (massa, densidade, média do número TC, índice de gordura e índice de calcificação) e 3 (três) atributos de textura (homogeneidade, entropia e variância).

Concluiu-se que o modelo que melhor classifica a malignidade dos nódulos incluir em sua equação um atributo qualitativo de localização, um atributo de intensidade relacionado ao índice de calcificação e um atributo morfológico relacionado à compacidade do nódulo, em concordância com a importância atribuída pelos médicos radiologistas a algumas dessas características na interpretação das imagens de tomografia computadorizada.

Apesar das limitações do modelo proposto, o valor preditivo do modelo (86,3%) foi muito maior do que o valor preditivo baseado na avaliação visual do médico mais experiente (65,3%), sem considerar o acaso, estando próximo da acurácia desse mesmo médico (85,1%), incluído o acerto devido ao acaso. Este fato revela a importância deste tipo de pesquisa na busca de um modelo quantitativo, baseado em imagens de tomografia computadorizada, que permita a classificação dos nódulos pulmonares.

No entanto, maiores estudos devem ser desenvolvidos no sentido de ampliar a amostra para incluir imagens de nódulos benignos que não tenham ido para biópsia. Considerando que a biópsia é indicada pelo médico apenas nos casos de suspeita de malignidade, pode-se afirmar que os casos mais clássicos de nódulos benignos, de bordas mais arredondadas, com conteúdo maior de gordura e calcificações, não compuseram a amostra estudada.

Além da ampliação da amostra, novas investigações devem ser realizadas no sentido de incluir atributos relacionados à textura da vizinhança do nódulo, visto que nódulos em meios opacos representam sinais de malignidade. Ressalta-se ainda a necessidade da inclusão de outras variáveis categóricas do diagnóstico clínico no modelo, como a presença/ausência do fumo e o acompanhamento temporal dos atributos, uma vez que os radiologistas consideram esses elementos como altamente relevantes.

7 REFERÊNCIAS

AKRAM S; JAVED M. Y; QAMAR U; KHANUM A. and HASSAN A. (2014). Artificial Neural Network based Classifications Lungs Nodule using Hybrid Features from Computerized Tomographic Images. **Applied Mathematics & Information Sciences**, v. 9, n.1, p. 183-195.

ALBERT H. ROSS; RUSSELL J. JOHN (2009). **Evaluation of the Solitary Pulmonary Nodule**. American Family Physician, V. 80, n:8.

ALTAMN E. I; BAIDYA T. K. N; DIAS L. M. R. (1979). Previsão de Problemas Financeiros em Empresas. **R. Adm. Emp**, 19(1), p. 17-28.

AMBROSIO, Paulo Eduardo. **Redes neurais artificiais no apoio ao diagnóstico diferencial de lesões intersticiais pulmonares**. 2002. 86f. Dissertação (Programa de Pós Graduação em Física Aplicada à Medicina e Biologia) – Universidade de São Paulo, São Paulo, 2002.

AOYAMA M; LI Q; KATSURAGAWA S; LI F; SONE S and DOI K. (2003). Computerized scheme for determination of the likelihood measure of malignancy for pulmonary nodules on low-dose CT images. **Medical Physics**, v.30, n.3.

ARIMURA H; KATSURAGAWA S; SUZUKI K; LI F; SHIRAISHI J; SONE S; DOI K. (2004). Computerized Scheme for Automated Detection of Lung Nodules in Low-Dose Computed Tomography Images for Lung Cancer Screening. **Acad Radiol**; 11:617-629.

ARMATO S. G; GIGER M L; and MACMAHON H. (2001). Automated detection of lung nodules in CT scans: Preliminary results. **Medical Physics**, v.28, n.8.

ARMATO S. G; LI F; MACMAHON H; SONE S; DOI K. (2002). Lung Cancer: Performance of Automated Lung Nodule Detection Applied to Cancers Missed in a CT Screening Program. **Radiology**, v.225, p.685-692.

ARMATO S. G; ALTMAN M. B; WILKIE J; LI F; DOI K e ROY A. S. (2003). Automated lung nodule classification following automated nodule detection on CT: A serial approach. **Medical Physics**, 0 (6).

ASHIZAWA K; ISHIDA T; MACMAHON H; VYBORNÝ C. J; KATSURAGAWA S and DOI K. (1999). Artificial Neural Networks in Chest Radiography: Application to the Differential Diagnosis of Interstitial Lung Disease. **Acad Radiol**; 6:2-9.

AZEVEDO-MARQUES, P. (2001). Diagnóstico Auxiliado por Computador na Radiologia. **Radiologia Brasileira** 34(5): 285-293.

BANKMANN, I. N(2000) Handbook of Medical Imaging. San Diego: Academic Press, 2000.

BRITO S; A; G. e NETO A; A (2008). Modelo de Classificação de risco de crédito de empresas. **R. Cont. Financeira**, v. 19, n° 46, p. 18-29.

CHEN H; WU W; DU J; YANG M; MA B. (2011). Classification of Pulmonary Nodules Using Neural Network Ensemble. **Lecture Notes in Computer Science**, v. 6677, p. 460-466.

CHEN H; ZHANG J; XU Y; CHEN B; ZHANG K. (2012). Performance comparison of artificial neural network and logistic regression model for differentiating lung nodules on CT scans. **Expert Systems with Applications**, 39: 11503-11509.

CORREA DA SILVA L. **Conduas em Pneumologia**. Rio de Janeiro: Revinter, 2001.

COSTA, A. J. L. E NADANOVSKY, P. **Epidemiologia e Bioestatística na Pesquisa Odontológica**. São Paulo: Atheneu, 2005.

EL-BAZ, A. AND SURI, J. S. **Lung Imaging and Computer-Aided Diagnosis**. United States: Taylor & Francis Group, 2012.

FISHER, R. A. (1936) The use of multiple measurements in taxonomic problems. **Annals Eugen** 7:179-188.

GONZALES R. C; WOODS R. E. (2000). **Processamento de Imagens Digitais**. 1 ed. Edgard Blucher, São Paulo.

GOULD M. K.; FLETCHER J.; IANNETTONI M. D.; LYNCH W. R.; MIDTHUN D. E.; NAIDICH D. P.; OST D. E. (2007). Evaluations of Patients With Pulmonary Nodules: When Is It Lung Cancer? **Chest**, 132:108S-130S.

GROSKIN S. A. **O Pulmão, correlações radiológicas e patológicas**. Rio de Janeiro: MEDSI, 1997.

GURCAN M. N; SAHINER B; PETRICK N; CAHN H-P; KAZEROONI E. A; CASCADE P. N; and HADIJISKI L. (2002). Lung nodule detection on thoracic computed tomography images: Preliminary evaluation of computer-aided diagnosis system. **Medical Physics**, v.29, n.11.

HAESEGAWA M.; SONE S.; TAKASHIMA S.; LI F.; YANG Z-G.; MURAYAMA Y.; WATANABE T. (2000). Growth rate of small lung cancers detected on mass ct screening. **Radiology** 73: 1252-1259.

HAIR JR., J. F. et al. **Multivariate data analysis**. 5th ed. New Jersey: Prentice Hall, 1998.
HARALICK R. M.; SHANMUGAM K.; DISTEINS I. (1973). Textural Features for Image Classification. **IEEE Transactions on systems, man and cybernetics**, vol. SMC-3, p. 610-621.

HENSCHKE C.I., YANKELEVITZ D.F., MIRTICHEVA R., MCGUINNESS G., MCCAULEY .D, MIETTINEN O.S. (2002). CT screening for lung cancer: frequency and significance of part-solid and nonsolid nodules. **AJR Am J Roentgenol**; 178(5):1053–1057.

HOOP B., GLETEMA H., VORST V. S., MURPHY K., KLAVEREN V. J. R., PROKOP M. (2010) Pulmonary Ground-Glass Nodules: Increase in Mass as Early Indicator of Growth. **Radiology**. 255(1);199-206.

HOSMER, D. W.; LEMESHOW, S (1980). A goodness-of-fit test for the multiple logistic regression model. **Communications in Statistics**, A10, p.1043-1980.

HOSMER, D. W.; LEMESHOW, S. **Applied Logistic Regression**. 2. Ed. New York: John Wiley & Sons, 2000.

HSIEH, J. (2003). "Computed Tomography: Principles, Design, Artifacts, and Recent Advances, 1st ed. SPIE Publications.

HU M-K. Visual Patterns Recognition by Moment Invariants (1962). **IRE Transactions on Information Theory**.

IWANO S., NAKAMURA T., KAMIOKA Y., ISHIGAKI T. (2005). Computer-aided diagnosis: A shape classification of pulmonary nodules imaged by high-resolution CT. **Computerized Medical Imaging and Graphics** 29: 565-570.

KANEKO M , EGUCHI K , OHMATSU H , KAKINUMA R., NARUKE T., SUEMASU K., MORIYAMA N. (1996). Peripheral lung cancer: screening and detection with low-dose spiral CT versus radiography. **Radiology**; 201(3):798–802.

KAWATA Y; NIKI N. (1998). Quantitative Surface Characterization of Pulmonary Nodules Bases on Thin-Section CT Images. **IEEE Transactions on Nuclear Science**. 45, 2132-2138.

KAWATA Y; NIKI N; KAKINUMA R; EGUCHI K; KANEKO M; MORIYAMA. (1997). Classification of Pulmonary Nodules in Thin-section CT Images Based on Shape Characterization. **IEEE Transactions on Nuclear Science**.

KAWATA Y; NIKI N; OHMATSU H; KUSUMOTO M; KAKINUMA R; MORI K; NISHIYAMA H; EGUCHI K; KANEKO M; MORIYAMA N. (2000). Quantitative analysis of internal texture for classification of pulmonary nodules in three-dimensional thoracic images. **Proceedings of SPIE**, vol. 3979.

KAWATA Y; NIKI N; OHMATSU H; MORIYAMA N. (2003). Example-Based Assisting Approach for Pulmonary Nodule Classification in Three-Dimensional Thoracic Computed Tomography Images. **Academic Radiology**, vol 10, No 12.

LANDIS J. R. & KOCH G. G. (1977). The measurement of observer agreement for categorical data. **Biometrics**, vol. 33, N° 1.

LEE, S. M.; PARK, C. M.; GOO, J. M.; LEE H. J.; LEE C. J.; KIM K. G.; KANG M-J.; LEE I. S.. (2010) Transient Part-Solid Nodules Detected at Screening Thin-Section CT for Lung Cancer: Comparison with Persistent Part-Solid Nodules. **Radiology**, 255(1):242-251.

LEUNG A. and SMITHUIS R. (2007). Solitary pulmonary nodule: benign versus malignant Differentiation with CT and PET-CT. Radiology Assistant, **Department of Radiology, Stanford University Medical Center, Stanford California**.

LEVI C.; GRAY J. E.; MCCULLOUGH E. C.; HATTERY R. R. (1982). The Unreability of CT Numbers as Absolute Values. **AJR** 139:443-447.

LI QIANG (2007). Recent Progress in Computer-Aided Diagnosis of Lung Nodules on Thin-Section CT. **Computed Med Imaging Graph** 31(4-5): 248-257.

MACMAHON H., AUSTIN J. H. M., GAMSU G., HEROLD C. J., JETT J. R., NAIDICH D. P., PATZ E. F., SWENSEN S. J. (2005) Guidelines for Management of Small Pulmonary Nodules Detected on CT Scans: A Statement from the Fleischner Society. **Radiology**, 237:395–400.

MAHADEVAPPA MAHESH (2002). The AAPM/RSNA Physics Tutorial for Residents. Search for Isotropic Resolution in CT from Conventional through Multiple-Row Detector. **Radiographics** 22:949 – 962.

MARTINEZ E. Z; LOUZADA-NETO F; PEREIRA B. B (2003). A curva ROC para testes diagnósticos. **Cadernos Saúde Coletiva** 11 (1): 7-31.

MCCARVILLE M. B; LEDERMAN H. M; SANTANA V. M; DAW N. C; SCHOCHAT S. J; LI C-S; KAUFMAN R. A. (2006). Distinguishing Bening from Malignant Pulmonary Nodules with Helical Chest CT in Children with Malignant Solid Tumors. **Radiology**, v.239, n. 2.

MCNITT-GRAY M. F; HART E. M; WYCKOFF N; SAYRE J. W; GOLDIN J. G; ABERLE D. R. [A] (1999) A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: Preliminary results. **Medical Physics** 26:880-888.

MCNITT-GRAY M. F; WYCKOFF N; SAYRE J. W; GOLDIN J. G; ABERLE D. R. [B] (1999) The effects of co-occurrence matrix based texture parameters on the classification of solitary pulmonary nodules imaged on computed tomography. **Computerized Medical Imaging and Graphics**. 23: 339-348.

MECCA, Fernando Augusto. **Estudo da Qualidade da imagem e índices de dose em Tomografia Computadorizada**. 2009. 61 f. Dissertação (Mestrado em Ciências em Engenharia Elétrica) – Faculdade de Engenharia Nuclear, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009.

MELO CARVALHO C. B., PERFEITO J. A. J., DAUD D. F., JUNIO A. S. C., SANTORO I. L. LEÃO L. E. V. (2012). Análise e validação de modelos probabilísticos de malignidade de nódulo pulmonar solitário em uma população no Brasil. **J. Bras Pneumol**; 38(5): 550 – 565.

METZ C. E. (1986). ROC methodology in radiologic imaging. **Invest. Radiol**, v. 21(9), p. 720-33.

MINAMI K; KAWATA Y; NIKI N; OHMATSU H; MORI K; YAMADA K; EGUCHI K; KANEKO M; MORIYAMA N. (2008). Classifying pulmonary nodules using dynamic enhanced CT images based on CT number histogram. **SPIE Proceedings**, vol. 6915.

MULL R. T. (1984) Mass Estimates by Computed Tomography: Physical Density from CT Numbers. **AJR** 143:110 –1104.

MULLER, N. L; SILVA C. I. S. **Imaging of the Chest**. Philadelphia: Saunders Elsevier, 2008.

MURAM-WINER H. T. (2006). The Solitary Pulmonary Nodule. **Radiology** 239:1.

NORDHOLM-CARSTENSEN A; JORGENSEN L. N; WILLE-JORGENSEN P. A; HANSEN H; HARLING H. (2014). Indeterminate Pulmonary Nodules in Colorectal-Cancer: Do Radiologists Agree? **Annals of Surgical Oncology**.

NUNES Lara Marins. **Segmentação e Caracterização de nódulos pulmonares**. 2014. 95 f. Monografia (Graduação em Engenharia Biomédica) – Faculdade de Engenharia Elétrica, Universidade Federal de Uberlândia, Uberlândia, 2014.

PATEL V. K.; NAIK S. K.; NAIDICH D. P.; TRAVIS W. D.; WEINGARTEN J. A.; LAZZARO R.; GUTTERMAN D. D.; WENTOWSKI C.; GROSU H. B.; RAOOF S. (2013). A Practical Algorithmic Approach to the Diagnosis and Management of Solitary Pulmonary Nodules. **Chest Imaging**, 143(3): 825 – 839.

PATZ E. F; CAMPA M. J; GOTTLIN E. G; TROTTER P. R; HERDON J. E; KAFADER D; GRANT R. P and EISENBERG M. (2013). Biomarkers to Help Guide Management of Patients with Pulmonary Nodules. **American Journal of Respiratory and Critical Care Medicine**, v. 188.

PERANDINI S; SOARDI G. A; MOTTON M; DALLASERRA C; MONTEMEZZI S. (2014). Limited Value of Logistic Regression Analysis in Solid Solitary Pulmonary Nodules Characterization: A Single-Center Experience on 288 Consecutive Cases. **Journal of Surgical Oncology**, n. 110, p. 883-887.

PERROCA M. G; GAIDZINSKI R. R. (2003). Avaliando a confiabilidade inteavaliadores de um instrumento para classificação de pacientes – coeficiente Kappa. **Rev. Esc. Enferm.** 37(1): 72-80.

PHELPS, M. E; GADO, M. H; HOFFMAN, E. J. (1975). Correlation of Effective Number and Electron Density with Attenuation Coefficients Measured with Polychromatic X Rays. **Radiology**, n. 117, p.585-588.

PRATT E. K. (1991). Digital Image Processing. Wiley Interscience, New York.
RANGAYYAN R. M. (2005). Biomedical Image Analysis. The biomedical Engineering. Series Editor Michael R. Neuman. CRC Press

RANGAYYAN R. M; EL-FARAMAWY N. M; DESAUTELS J. E. L; ALIM O. A. (1997). Measures of Acutance and Shape for Classification of Breast Tumors. **IEEE Transactions on Medical Imaging**. Vol 16(6).

REEVES A. P; KOSTIS W. J. (2000). Computer-Aided Diagnosis of Small Pulmonary Nodules. **Seminars in Ultrasound, CT and MRI**, Vol 21 No 2: p 116-128.

SIEGELMAN S. STANLEY; ZERTHOUNI A. ELIAS; LEO P. FRANCK; KHOURI F. NAGI; STITIK P. FREDERICK (1980). CT of the Solitary Pulmonary Nodule. **AJR** 135:1-13.

SLUIMER INGRID C.; WARS PAUL F.; VIERGEVER MAX. A.; GINNEKEN BRAM. (2003). Computer-aided diagnosis in high resolution CT of the lungs. **Medical Physics** 30 (12).

SWENSE S.; JETT J. R.; HARTMAN T. E.; MIDTHUN D. E.; SYKES A. M.; AUGHENBAUGH G. L.; CLEMEN M. A. (2003). Lung Cancer Screening with CT: Mayo Clinic Experience. **Radiology** 226: 756 – 761.

TOSHIOKA S; KANAZAWA K; NIKI N; SATOH H; OHMATSU H; EGUCHI K; MORIYAMA N. (1997). Computer Aided Diagnosis System for Lung Cancer Based on Helical CT Images. **SPIE Proceedings** 3034:975-984.

TOZAKI M; ICHIBA N; FUKUDA K. (2005). Dynamic magnetic resonance imaging of solitary pulmonary nodules: utility of kinetic patterns in differential diagnosis. **J Comput Assist Tomography**, 29(1): 13-19.

TOZAKI T; KAWATA Y; NIKI N; OHMATSU H; KAKINUMA R; EGUCHI K; KANEKO M; MORIYAMA N. (1998). Pulmonary Organs Analysis for Differential Diagnosis Based on Toracic Thin-section CT Images. **IEEE Transactions on Nuclear Science**, 45, 3075-3082.

TRINDADE, Fernanda Rocha. **Desenvolvimento de uma metodologia de classificação de padrões em imagens de tomografia computadorizada do pulmão**. 2009.58f. Monografia (Graduação em Física) – Faculdade de Física, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2009.

VINAY K., RAO A., KUMAR G. H. (2011). Computerized Analysis of Classification of Lung Nodules and Comparison between Homogeneous and Heterogeneous Ensemble of Classifier Model. **IEEE** 978-0-7695-4599-8/11.

WAY T. W; SAHINER B; CHAN P-H; HADJIISKI L; CASCADE P. N; CGUGHTAI A; BOGOT N; KAZEROONI E. (2009). Computer-aided diagnosis of pulmonary nodules on CT scans: Improvement of classification performance with nodule surface features. **Medical Physics**, 36(7).

WINER-MURAM H.(2006). The Solitary Pulmonary Nodule. **Radiology** 239: Number 1.

ANEXO A – Análise de Regressão Logística Univariada de cada Atributo

IDADE DO PACIENTE

LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER Idade /CONTRAST (Localização)=Indicator /CONTRAST (Sexo)=Indicator /SAVE=PRED /PRINT=GOODFIT /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

Logistic Regression

		Notes	
	Output Created		18-Jun-2015 17h47min52s
Input	Comments		
	Data	E:\Mestrado Ana Paula\Dados corrigidos.sav	
	Active Dataset	DataSet1	
	Filter	<none>	
	Weight	<none>	
	Split File	<none>	
	N of Rows in Working Data File		51
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing	
	Syntax	LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER Idade /CONTRAST (Localização)=Indicator /CONTRAST (Sexo)=Indicator /SAVE=PRED /PRINT=GOODFIT /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).	
Resources	Processor Time		0:00:00.031
	Elapsed Time		0:00:00.043
Variables Created or Modified	PRE_28	Predicted probability	

[DataSet1] E:\Mestrado Ana Paula\Dados corrigidos.sav

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	51	100,0
	Missing Cases	0	,0
Total		51	100,0
Unselected Cases		0	,0
	Total	51	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
Benigno	0
Maligno	1

Block 0: Beginning Block

Classification Table^{a,b}

Observed			Predicted		
			Padrão Ouro		Percentage Correct
			Benigno	Maligno	
Step 0	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
		Overall Percentage			64,7

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	,606	,293	4,279	1	,039	1,833

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Idade	,378	1	,539
		Overall Statistics	,378	1	,539

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	,386	1	,534
	Block	,386	1	,534
	Model	,386	1	,534

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	65,837 ^a	,008	,010

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	4,421	8	,817

Contingency Table for Hosmer and Lemeshow Test

		Padrão Ouro = Benigno		Padrão Ouro = Maligno		Total
		Observed	Expected	Observed	Expected	
Step 1	1	3	2,085	2	2,915	5
	2	2	1,973	3	3,027	5
	3	1	1,901	4	3,099	5
	4	1	1,118	2	1,882	3
	5	1	1,834	4	3,166	5
	6	2	2,136	4	3,864	6

7	3	1,745	2	3,255	5
8	1	1,993	5	4,007	6
9	2	1,556	3	3,444	5
10	2	1,659	4	4,341	6

Classification Table^a

Observed		Predicted			
		Padrão Ouro		Percentage Correct	
		Benigno	Maligno		
Step 1	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
		Overall Percentage			64,7

a. The cut value is ,500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a						
Idade	-,015	,025	,374	1	,541	,985
Constant	1,537	1,558	,972	1	,324	4,649

a. Variable(s) entered on step 1: Idade.

SEXO

LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER Sexo /CONTRAST (Localização)=Indicator /CONTRAST (Sexo)=Indicator /SAVE=PRED /PRINT=GOODFIT /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

Logistic Regression

Notes

	Output Created	18-Jun-2015 17h47min4s
	Comments	
Input	Data	E:\Mestrado Ana Paula\Dados corrigidos.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	51
Missing Value Handling	Definition of Missing Syntax	User-defined missing values are treated as missing LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER Sexo /CONTRAST (Localização)=Indicator /CONTRAST (Sexo)=Indicator /SAVE=PRED /PRINT=GOODFIT /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
Resources	Processor Time	0:00:00.031
	Elapsed Time	0:00:00.020
Variables Created or Modified	PRE_27	Predicted probability

[DataSet1] E:\Mestrado Ana Paula\Dados corrigidos.sav

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	51	100,0
	Missing Cases	0	,0
	Total	51	100,0
	Unselected Cases	0	,0
	Total	51	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
Benigno	0
Maligno	1

Categorical Variables Codings

		Frequency	Parameter coding (1)
Sexo do Paciente	F	30	1,000
	M	21	,000

Block 0: Beginning Block

Classification Table^{a,b}

Observed			Predicted		
			Padrão Ouro		Percentage Correct
	Benigno	Maligno			
Step 0 Padrão Ouro	Benigno	0	18	,0	
	Maligno	0	33	100,0	
	Overall Percentage			64,7	

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	,606	,293	4,279	1	,039	1,833

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Sexo(1)	,894	1	,344
		Overall Statistics	,894	1	,344

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
--	------------	----	------

Step 1	Step	,889	1	,346
	Block	,889	1	,346
	Model	,889	1	,346

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	65,334 ^a	,017	,024

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	,000	0	.

Contingency Table for Hosmer and Lemeshow Test

		Padrão Ouro = Benigno		Padrão Ouro = Maligno		Total
		Observed	Expected	Observed	Expected	
Step 1	1	9	9,000	12	12,000	21
	2	9	9,000	21	21,000	30

Classification Table^a

Observed			Predicted		
			Padrão Ouro		Percentage Correct
			Benigno	Maligno	
Step 1	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
Overall Percentage					64,7

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Sexo(1)	,560	,594	,887	1	,346	1,750
	Constant	,288	,441	,426	1	,514	1,333

a. Variable(s) entered on step 1: Sexo.

LOCALIZAÇÃO

LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER Localização /CONTRAST (Localização)=Indicator /CONTRAST (Sexo)=Indicator /SAVE=PRED /PRINT=GOODFIT /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

Logistic Regression

Notes

Input	Output Created	18-Jun-2015 17h48min42s
	Comments	
	Data	E:\Mestrado Ana Paula\Dados corrigidos.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>

	N of Rows in Working Data File		51
Missing Value Handling	Definition of Missing Syntax	User-defined missing values are treated as missing	
		LOGISTIC REGRESSION VARIABLES PO	
		/METHOD=ENTER Localização	
		/CONTRAST (Localização)=Indicator	
		/CONTRAST (Sexo)=Indicator	
		/SAVE=PRED	
		/PRINT=GOODFIT	
		/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20)	
		CUT(0.5).	
Resources	Processor Time		0:00:00.031
	Elapsed Time		0:00:00.031
Variables Created or Modified	PRE_29	Predicted probability	

[DataSet1] E:\Mestrado Ana Paula\Dados corrigidos.sav

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	51	100,0
	Missing Cases	0	,0
	Total	51	100,0
Unselected Cases		0	,0
	Total	51	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
Benigno	0
Maligno	1

Categorical Variables Codings

		Frequency	Parameter coding			
			(1)	(2)	(3)	(4)
Localização do Nódulo	LID	16	1,000	,000	,000	,000
	LIE	7	,000	1,000	,000	,000
	LMD	6	,000	,000	1,000	,000
	LSD	9	,000	,000	,000	1,000
	LSE	13	,000	,000	,000	,000

Block 0: Beginning Block

Classification Table^{a,b}

			Predicted		
			Padrão Ouro		Percentage Correct
			Benigno	Maligno	
Observed	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0

Overall Percentage			64,7
--------------------	--	--	------

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	,606	,293	4,279	1	,039	1,833

Variables not in the Equation

	Score	df	Sig.
Step 0 Variables			
Localização	9,852	4	,043
Localização(1)	2,794	1	,095
Localização(2)	9,032	1	,003
Localização(3)	,011	1	,915
Localização(4)	,018	1	,892
Overall Statistics	9,852	4	,043

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Step 1 Step	9,896	4	,042
Block	9,896	4	,042
Model	9,896	4	,042

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	56,328 ^a	,176	,243

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	,000	2	1,000

Contingency Table for Hosmer and Lemeshow Test

		Padrão Ouro = Benigno		Padrão Ouro = Maligno		Total
		Observed	Expected	Observed	Expected	
Step 1	1	6	6,000	1	1,000	7
	2	5	5,000	10	10,000	15
	3	4	4,000	9	9,000	13
	4	3	3,000	13	13,000	16

Classification Table^a

	Observed	Predicted		
		Padrão Ouro		Percentage Correct
		Benigno	Maligno	
Step 1	Padrão Ouro Benigno	6	12	33,3

Maligno	1	32	97,0
Overall Percentage			74,5

a. The cut value is ,500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a Localização			6,788	4	,148	
Localização(1)	,655	,878	,557	1	,456	1,926
Localização(2)	-2,603	1,236	4,434	1	,035	,074
Localização(3)	-,118	1,054	,012	1	,911	,889
Localização(4)	-,118	,928	,016	1	,899	,889
Constant	,811	,601	1,821	1	,177	2,250

a. Variable(s) entered on step 1: Localização.

VOLUME

LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER vol /SAVE=PRED /PRINT=GOODFIT /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

Logistic Regression

Notes

	Output Created	18-Jun-2015 17h31min35s
	Comments	
Input	Data	E:\Mestrado Ana Paula\Dados corrigidos.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	51
Missing Value Handling	Definition of Missing Syntax	User-defined missing values are treated as missing LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER vol /SAVE=PRED /PRINT=GOODFIT /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
Resources	Processor Time	0:00:00.032
	Elapsed Time	0:00:00.030
Variables Created or Modified	PRE_19	Predicted probability

[DataSet1] E:\Mestrado Ana Paula\Dados corrigidos.sav

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	51	100,0
	Missing Cases	0	,0
	Total	51	100,0
Unselected Cases		0	,0
	Total	51	100,0

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	51	100,0
	Missing Cases	0	,0
	Total	51	100,0
	Unselected Cases	0	,0
	Total	51	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
Benigno	0
Maligno	1

Block 0: Beginning Block

Classification Table^{a,b}

Observed			Predicted		
			Padrão Ouro		Percentage Correct
			Benigno	Maligno	
Step 0	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
		Overall Percentage			64,7

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	,606	,293	4,279	1	,039	1,833

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	vol	3,825	1	,050
		Overall Statistics	3,825	1	,050

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	5,720	1	,017
	Block	5,720	1	,017
	Model	5,720	1	,017

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	60,503 ^a	,106	,146

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	60,503 ^a	,106	,146

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	8,812	8	,358

Contingency Table for Hosmer and Lemeshow Test

		Padrão Ouro = Benigno		Padrão Ouro = Maligno		Total
		Observed	Expected	Observed	Expected	
Step 1	1	4	2,355	1	2,645	5
	2	3	2,333	2	2,667	5
	3	3	2,309	2	2,691	5
	4	0	2,267	5	2,733	5
	5	2	2,177	3	2,823	5
	6	1	2,073	4	2,927	5
	7	2	1,945	3	3,055	5
	8	2	1,445	3	3,555	5
	9	1	,746	4	4,254	5
	10	0	,350	6	5,650	6

Classification Table^a

Observed	Padrão Ouro	Predicted			
		Padrão Ouro		Percentage Correct	
		Benigno	Maligno		
Step 1	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
		Overall Percentage			64,7

a. The cut value is ,500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a						
vol	,037	,020	3,329	1	,068	1,037
Constant	,090	,367	,060	1	,806	1,094

a. Variable(s) entered on step 1: vol.

DENSIDADE

LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER den /SAVE=PRED /PRINT=GOODFIT /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

Logistic Regression**Notes**

Output Created	18-Jun-2015 17h33min0s
Comments	
Input Data	E:\Mestrado Ana Paula\Dados corrigidos.sav

	Active Dataset	DataSet1	
	Filter	<none>	
	Weight	<none>	
	Split File	<none>	
	N of Rows in Working Data File		51
Missing Value Handling	Definition of Missing Syntax	User-defined missing values are treated as missing LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER den /SAVE=PRED /PRINT=GOODFIT /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).	
Resources	Processor Time		0:00:00.016
	Elapsed Time		0:00:00.030
Variables Created or Modified	PRE_20	Predicted probability	

[DataSet1] E:\Mestrado Ana Paula\Dados corrigidos.sav

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	51	100,0
	Missing Cases	0	,0
	Total	51	100,0
Unselected Cases		0	,0
	Total	51	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
Benigno	0
Maligno	1

Block 0: Beginning Block

Classification Table^{a,b}

Observed			Predicted		
			Padrão Ouro		Percentage Correct
			Benigno	Maligno	
Step 0	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
		Overall Percentage			64,7

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	,606	,293	4,279	1	,039	1,833

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	den	1,059	1	,303
Overall Statistics			1,059	1	,303

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	1,052	1	,305
	Block	1,052	1	,305
	Model	1,052	1	,305

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	65,172 ^a	,020	,028

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	8,614	8	,376

Contingency Table for Hosmer and Lemeshow Test

		Padrão Ouro = Benigno		Padrão Ouro = Maligno		Total
		Observed	Expected	Observed	Expected	
Step 1	1	4	2,450	1	2,550	5
	2	2	2,214	3	2,786	5
	3	1	2,002	4	2,998	5
	4	0	1,837	5	3,163	5
	5	2	1,723	3	3,277	5
	6	2	1,644	3	3,356	5
	7	3	1,571	2	3,429	5
	8	1	1,507	4	3,493	5
	9	2	1,456	3	3,544	5
	10	1	1,596	5	4,404	6

Classification Table^a

			Predicted		
			Padrão Ouro		Percentage Correct
Observed			Benigno	Maligno	
Step 1	Padrão Ouro	Benigno	1	17	5,6
		Maligno	0	33	100,0
	Overall Percentage				66,7

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	den	2,328	2,283	1,040	1	,308	10,262
	Constant	-1,486	2,062	,519	1	,471	,226

a. Variable(s) entered on step 1: den.

MASSA

LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER massa /SAVE=PRED /PRINT=GOODFIT /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

Logistic Regression

Notes

	Output Created	18-Jun-2015 17h35min16s
Input	Comments	
	Data	E:\Mestrado Ana Paula\Dados corrigidos.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	51
Missing Value Handling	Definition of Missing Syntax	User-defined missing values are treated as missing LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER massa /SAVE=PRED /PRINT=GOODFIT /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
Resources	Processor Time	0:00:00.031
	Elapsed Time	0:00:00.030
Variables Created or Modified	PRE_21	Predicted probability

[DataSet1] E:\Mestrado Ana Paula\Dados corrigidos.sav

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	51	100,0
	Missing Cases	0	,0
	Total	51	100,0
	Unselected Cases	0	,0
	Total	51	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
Benigno	0
Maligno	1

Block 0: Beginning Block

Classification Table^{a,b}

Observed			Predicted		
			Padrão Ouro		Percentage Correct
			Benigno	Maligno	
Step 0	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
		Overall Percentage			64,7

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	,606	,293	4,279	1	,039	1,833

Variables not in the Equation

	Score	df	Sig.
Step 0 Variables massa	3,738	1	,053
Overall Statistics	3,738	1	,053

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Step 1 Step	5,668	1	,017
Block	5,668	1	,017
Model	5,668	1	,017

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	60,556 ^a	,105	,145

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	11,716	8	,164

Contingency Table for Hosmer and Lemeshow Test

		Padrão Ouro = Benigno		Padrão Ouro = Maligno		Total
		Observed	Expected	Observed	Expected	
Step 1	1	3	2,341	2	2,659	5
	2	4	2,322	1	2,678	5
	3	3	2,295	2	2,705	5
	4	0	2,255	5	2,745	5
	5	2	2,176	3	2,824	5
	6	1	2,100	4	2,900	5
	7	1	1,958	4	3,042	5
	8	3	1,444	2	3,556	5

9	1	,766	4	4,234	5
10	0	,342	6	5,658	6

Classification Table^a

Observed			Predicted		
			Padrão Ouro		Percentage Correct
	Benigno	Maligno			
Step 1	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
		Overall Percentage			64,7

a. The cut value is ,500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a massa	,036	,020	3,294	1	,070	1,037
Constant	,109	,362	,091	1	,764	1,115

a. Variable(s) entered on step 1: massa.

MÉDIA DO NÚMERO DE TC (TODO O NÓDULO)

LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER M_HU_TN /SAVE=PRED /PRINT=GOODFIT /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

Logistic Regression**Notes**

	Output Created	18-Jun-2015 17h35min56s
Input	Comments	
	Data	E:\Mestrado Ana Paula\Dados corrigidos.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	51
Missing Value Handling	Definition of Missing Syntax	User-defined missing values are treated as missing LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER M_HU_TN /SAVE=PRED /PRINT=GOODFIT /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
Resources	Processor Time	0:00:00.063
	Elapsed Time	0:00:00.060
Variables Created or Modified	PRE_22	Predicted probability

[DataSet1] E:\Mestrado Ana Paula\Dados corrigidos.sav

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	51	100,0

Missing Cases	0	,0
Total	51	100,0
Unselected Cases	0	,0
Total	51	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
Benigno	0
Maligno	1

Block 0: Beginning Block

Classification Table^{a,b}

Observed			Predicted		
			Padrão Ouro		Percentage Correct
			Benigno	Maligno	
Step 0	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
		Overall Percentage			64,7

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	,606	,293	4,279	1	,039	1,833

Variables not in the Equation

	Score	df	Sig.
Step 0 Variables M_HU_TN	1,059	1	,303
Overall Statistics	1,059	1	,303

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Step 1 Step	1,052	1	,305
Block	1,052	1	,305
Model	1,052	1	,305

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	65,172 ^a	,020	,028

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	8,614	8	,376

Contingency Table for Hosmer and Lemeshow Test

		Padrão Ouro = Benigno		Padrão Ouro = Maligno		Total
		Observed	Expected	Observed	Expected	
Step 1	1	4	2,450	1	2,550	5
	2	2	2,214	3	2,786	5
	3	1	2,002	4	2,998	5
	4	0	1,837	5	3,163	5
	5	2	1,723	3	3,277	5
	6	2	1,644	3	3,356	5
	7	3	1,571	2	3,429	5
	8	1	1,507	4	3,493	5
	9	2	1,456	3	3,544	5
	10	1	1,596	5	4,404	6

Classification Table^a

Observed	Padrão Ouro	Predicted		
		Padrão Ouro		Percentage Correct
		Benigno	Maligno	
Step 1	Benigno	1	17	5,6
	Maligno	0	33	100,0
	Overall Percentage			66,7

a. The cut value is ,500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a M_HU_TN	,002	,002	1,040	1	,308	1,002
Constant	,842	,382	4,849	1	,028	2,321

a. Variable(s) entered on step 1: M_HU_TN.

ÍNDICE DE GORDURA (IG)

LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER IG /CONTRAST (Localização)=Indicator /SAVE=PRED /PRINT=GOODFIT /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

Logistic Regression

Notes

	Output Created	18-Jun-2015 17h46min27s
Input	Comments	
	Data	E:\Mestrado Ana Paula\Dados corrigidos.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	51
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing

Syntax	LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER IG /CONTRAST (Localização)=Indicator /SAVE=PRED /PRINT=GOODFIT /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).		
Resources	Processor Time		0:00:00.031
	Elapsed Time		0:00:00.040
Variables Created or Modified	PRE_26	Predicted probability	

[DataSet1] E:\Mestrado Ana Paula\Dados corrigidos.sav

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	51	100,0
	Missing Cases	0	,0
	Total	51	100,0
Unselected Cases		0	,0
	Total	51	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
Benigno	0
Maligno	1

Block 0: Beginning Block

Classification Table^{a,b}

Observed			Predicted		
			Padrão Ouro		Percentage Correct
			Benigno	Maligno	
Step 0	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
		Overall Percentage			64,7

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	,606	,293	4,279	1	,039	1,833

Variables not in the Equation

	Score	df	Sig.
Step 0 Variables IG	3,176	1	,075
Overall Statistics	3,176	1	,075

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	3,259	1	,071
	Block	3,259	1	,071
	Model	3,259	1	,071

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	62,965 ^a	,062	,085

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	5,524	8	,700

Contingency Table for Hosmer and Lemeshow Test

		Padrão Ouro = Benigno		Padrão Ouro = Maligno		Total
		Observed	Expected	Observed	Expected	
Step 1	1	2	2,882	3	2,118	5
	2	4	2,452	1	2,548	5
	3	2	2,197	3	2,803	5
	4	1	1,991	4	3,009	5
	5	2	1,735	3	3,265	5
	6	2	1,625	3	3,375	5
	7	2	1,481	3	3,519	5
	8	2	1,387	3	3,613	5
	9	1	1,220	4	3,780	5
	10	0	1,029	6	4,971	6

Classification Table^a

Observed		Predicted			
		Padrão Ouro		Percentage Correct	
		Benigno	Maligno		
Step 1	Padrão Ouro	Benigno	4	14	22,2
		Maligno	3	30	90,9
		Overall Percentage			66,7

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	IG	-71,462	41,501	2,965	1	,085	,000
	Constant	1,841	,797	5,335	1	,021	6,302

a. Variable(s) entered on step 1: IG.

ÍNDICE DE CALCIFICAÇÃO (IC)

LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER IC /CONTRAST (Localização)=Indicator /SAVE=PRED /PRINT=GOODFIT /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

Logistic Regression

		Notes
Input	Output Created	18-Jun-2015 17h45min35s
	Comments	
	Data	E:\Mestrado Ana Paula\Dados corrigidos.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	51
Missing Value Handling	Definition of Missing Syntax	User-defined missing values are treated as missing LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER IC /CONTRAST (Localização)=Indicator /SAVE=PRED /PRINT=GOODFIT /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
Resources	Processor Time	0:00:00.031
	Elapsed Time	0:00:00.012
Variables Created or Modified	PRE_25	Predicted probability

[DataSet1] E:\Mestrado Ana Paula\Dados corrigidos.sav

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	51	100,0
	Missing Cases	0	,0
Total		51	100,0
Unselected Cases		0	,0
	Total	51	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
Benigno	0
Maligno	1

Block 0: Beginning Block

Classification Table^{a,b}

Observed	Predicted	
	Padrão Ouro	

			Benigno	Maligno	Percentage Correct
Step 0	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
		Overall Percentage			64,7

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	,606	,293	4,279	1	,039	1,833

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	IC	2,381	1	,123
		Overall Statistics	2,381	1	,123

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	2,280	1	,131
	Block	2,280	1	,131
	Model	2,280	1	,131

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	63,943 ^a	,044	,060

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	23,040	7	,002

Contingency Table for Hosmer and Lemeshow Test

		Padrão Ouro = Benigno		Padrão Ouro = Maligno		Total
		Observed	Expected	Observed	Expected	
Step 1	1	4	3,155	1	1,845	5
	2	1	2,142	4	2,858	5
	3	2	1,738	3	3,262	5
	4	0	1,631	5	3,369	5
	5	1	1,527	4	3,473	5
	6	0	1,507	5	3,493	5
	7	1	1,503	4	3,497	5
	8	0	1,500	5	3,500	5
	9	9	3,298	2	7,702	11

Classification Table^a

Observed			Predicted		
			Padrão Ouro		Percentage Correct
			Benigno	Maligno	
Step 1	Padrão Ouro	Benigno	4	14	22,2
		Maligno	1	32	97,0
		Overall Percentage			70,6

a. The cut value is ,500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a IC	-8,812	6,022	2,142	1	,143	,000
Constant	,848	,342	6,145	1	,013	2,335

a. Variable(s) entered on step 1: IC.

ÁREA CORTE CENTRAL

LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER A_CC /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

Logistic Regression

Notes

	Output Created	09-Jul-2015 17h23min9s
Input	Comments	
	Data	E:\Mestrado Ana Paula\Dados corrigidos.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	51
Missing Value Handling	Definition of Missing Syntax	User-defined missing values are treated as missing LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER A_CC /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
Resources	Processor Time	0:00:00.000
	Elapsed Time	0:00:00.000

[DataSet1] E:\Mestrado Ana Paula\Dados corrigidos.sav

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	51	100,0
	Missing Cases	0	,0
Total		51	100,0
Unselected Cases		0	,0
	Total	51	100,0

a. If weight is in effect, see classification table for the total number of cases.

**Dependent Variable
Encoding**

Original Value	Internal Value
Benigno	0
Maligno	1

Block 0: Beginning Block

Classification Table^{a,b}

Observed			Predicted		
			Padrão Ouro		Percentage Correct
			Benigno	Maligno	
Step 0	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
		Overall Percentage			64,7

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	,606	,293	4,279	1	,039	1,833

Variables not in the Equation

	Score	df	Sig.
Step 0 Variables A_CC	4,332	1	,037
Overall Statistics	4,332	1	,037

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Step 1 Step	5,497	1	,019
Block	5,497	1	,019
Model	5,497	1	,019

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	60,727 ^a	,102	,141

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

Classification Table^a

Observed			Predicted		
			Padrão Ouro		Percentage Correct
			Benigno	Maligno	
Step 1	Padrão Ouro	Benigno	1	17	5,6
		Maligno	0	33	100,0
		Overall Percentage			66,7

Classification Table^a

Observed		Predicted			
		Padrão Ouro		Percentage Correct	
		Benigno	Maligno		
Step 1	Padrão Ouro	Benigno	1	17	5,6
		Maligno	0	33	100,0
		Overall Percentage			66,7

a. The cut value is ,500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a A_CC	,001	,001	3,648	1	,056	1,001
Constant	-,067	,417	,026	1	,872	,935

a. Variable(s) entered on step 1: A_CC.

PERÍMETRO CORTE CENTRAL

LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER P_CC /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

Logistic Regression

Notes

	Output Created	09-Jul-2015 17h28min41s
	Comments	
Input	Data	E:\Mestrado Ana Paula\Dados corrigidos.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	51
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing
	Syntax	LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER P_CC /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
Resources	Processor Time	0:00:00.000
	Elapsed Time	0:00:00.000

[DataSet1] E:\Mestrado Ana Paula\Dados corrigidos.sav

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	51	100,0
	Missing Cases	0	,0
	Total	51	100,0
Unselected Cases		0	,0
	Total	51	100,0

a. If weight is in effect, see classification table for the total number of cases.

**Dependent Variable
Encoding**

Original Value	Internal Value
Benigno	0
Maligno	1

Block 0: Beginning Block

Classification Table^{a,b}

Observed			Predicted		
			Padrão Ouro		Percentage Correct
Benigno	Maligno				
Step 0	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
		Overall Percentage			64,7

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	,606	,293	4,279	1	,039	1,833

Variables not in the Equation

	Score	df	Sig.
Step 0 Variables P_CC	7,065	1	,008
Overall Statistics	7,065	1	,008

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Step 1 Step	8,684	1	,003
Block	8,684	1	,003
Model	8,684	1	,003

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	57,539 ^a	,157	,215

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

Classification Table^a

Observed			Predicted		
			Padrão Ouro		Percentage Correct
Benigno	Maligno				
Step 1	Padrão Ouro	Benigno	9	9	50,0
		Maligno	6	27	81,8
		Overall Percentage			70,6

Classification Table^a

Observed		Predicted			
		Padrão Ouro		Percentage Correct	
		Benigno	Maligno		
Step 1	Padrão Ouro	Benigno	9	9	50,0
		Maligno	6	27	81,8
		Overall Percentage			70,6

a. The cut value is ,500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a P_CC	,015	,007	5,603	1	,018	1,016
Constant	-,945	,654	2,089	1	,148	,389

a. Variable(s) entered on step 1: P_CC.

COMPACIDADE CORTE CENTRAL

LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER C_CC /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

Logistic Regression

Notes

	Output Created	09-Jul-2015 17h26min4s
	Comments	
Input	Data	E:\Mestrado Ana Paula\Dados corrigidos.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	51
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing
	Syntax	LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER C_CC /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
Resources	Processor Time	0:00:00.000
	Elapsed Time	0:00:00.000

[DataSet1] E:\Mestrado Ana Paula\Dados corrigidos.sav

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	51	100,0
	Missing Cases	0	,0
	Total	51	100,0
Unselected Cases		0	,0
	Total	51	100,0

a. If weight is in effect, see classification table for the total number of cases.

**Dependent Variable
Encoding**

Original Value	Internal Value
Benigno	0
Maligno	1

Block 0: Beginning Block

Classification Table^{a,b}

Observed			Predicted		
			Padrão Ouro		Percentage Correct
			Benigno	Maligno	
Step 0	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
		Overall Percentage			64,7

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	,606	,293	4,279	1	,039	1,833

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	C_CC	12,178	1	,000
		Overall Statistics	12,178	1	,000

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	21,561	1	,000
	Block	21,561	1	,000
	Model	21,561	1	,000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	44,663 ^a	,345	,474

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than ,001.

Classification Table^a

Observed			Predicted		
			Padrão Ouro		Percentage Correct
			Benigno	Maligno	
Step 1	Padrão Ouro	Benigno	12	6	66,7
		Maligno	7	26	78,8

	Overall Percentage			74,5
--	--------------------	--	--	------

a. The cut value is ,500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a C_CC	5,907	1,963	9,056	1	,003	367,742
Constant	-9,822	3,316	8,774	1	,003	,000

a. Variable(s) entered on step 1: C_CC.

MOMENTO INVARIANTE DE ORDEM 1 CORTE CENTRAL

LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER MI01_CC /SAVE=PRED /PRINT=GOODFIT CI(95) /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

Logistic Regression

Notes

	Output Created	09-Jul-2015 19h41min15s
Input	Comments	
	Data	E:\Mestrado Ana Paula\Dados corrigidos.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	51
Missing Value Handling	Definition of Missing Syntax	User-defined missing values are treated as missing LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER MI01_CC /SAVE=PRED /PRINT=GOODFIT CI(95) /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
Resources	Processor Time	0:00:00.031
	Elapsed Time	0:00:00.030
Variables Created or Modified	PRE_20	Predicted probability

[DataSet1] E:\Mestrado Ana Paula\Dados corrigidos.sav

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	51	100,0
	Missing Cases	0	,0
Total		51	100,0
Unselected Cases		0	,0
Total		51	100,0

a. If weight is in effect, see classification table for the total number of cases.

**Dependent Variable
Encoding**

Original Value	Internal Value
----------------	----------------

Benigno	0
Maligno	1

Block 0: Beginning Block**Classification Table^{a,b}**

Observed			Predicted		
			Padrão Ouro		Percentage Correct
			Benigno	Maligno	
Step 0	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
		Overall Percentage			64,7

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	,606	,293	4,279	1	,039	1,833

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	MI01_CC	1,385	1	,239
		Overall Statistics	1,385	1	,239

Block 1: Method = Enter**Omnibus Tests of Model Coefficients**

		Chi-square	df	Sig.
Step 1	Step	1,505	1	,220
	Block	1,505	1	,220
	Model	1,505	1	,220

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	64,719 ^a	,029	,040

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	10,775	8	,215

Contingency Table for Hosmer and Lemeshow Test

		Padrão Ouro = Benigno		Padrão Ouro = Maligno		Total
		Observed	Expected	Observed	Expected	
Step 1	1	2	2,440	3	2,560	5
	2	2	2,098	3	2,902	5
	3	4	1,989	1	3,011	5
	4	1	1,920	4	3,080	5

5	0	1,893	5	3,107	5
6	3	1,803	2	3,197	5
7	2	1,657	3	3,343	5
8	2	1,520	3	3,480	5
9	2	1,426	3	3,574	5
10	0	1,253	6	4,747	6

Classification Table^a

Observed		Predicted			
		Padrão Ouro		Percentage Correct	
		Benigno	Maligno		
Step 1	Padrão Ouro	Benigno	0	18	,0
		Maligno	1	32	97,0
		Overall Percentage			62,7

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	MI01_CC	18,806	16,349	1,323	1	,250	1,470E8
	Constant	-1,000	1,407	,505	1	,477	,368

a. Variable(s) entered on step 1: MI01_CC.

MOMENTO INVARIANTE DE ORDEM 2 CORTE CENTRAL

LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER MI02_CC /SAVE=PRED /PRINT=GOODFIT CI(95) /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

Logistic Regression

Notes

	Output Created	09-Jul-2015 19h42min5s
Input	Comments	
	Data	E:\Mestrado Ana Paula\Dados corrigidos.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	51
Missing Value Handling	Definition of Missing Syntax	User-defined missing values are treated as missing LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER MI02_CC /SAVE=PRED /PRINT=GOODFIT CI(95) /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
Resources	Processor Time	0:00:00.000
	Elapsed Time	0:00:00.010
Variables Created or Modified	PRE_21	Predicted probability

[DataSet1] E:\Mestrado Ana Paula\Dados corrigidos.sav

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	51	100,0
	Missing Cases	0	,0
	Total	51	100,0
	Unselected Cases	0	,0
	Total	51	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
Benigno	0
Maligno	1

Block 0: Beginning Block**Classification Table^{a,b}**

Observed			Predicted		
			Padrão Ouro		Percentage Correct
			Benigno	Maligno	
Step 0	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
		Overall Percentage			64,7

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	,606	,293	4,279	1	,039	1,833

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	MI02_CC	,630	1	,427
		Overall Statistics	,630	1	,427

Block 1: Method = Enter**Omnibus Tests of Model Coefficients**

		Chi-square	df	Sig.
Step 1	Step	,640	1	,424
	Block	,640	1	,424
	Model	,640	1	,424

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square

1	65,583 ^a	,012	,017
---	---------------------	------	------

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	10,851	8	,210

Contingency Table for Hosmer and Lemeshow Test

		Padrão Ouro = Benigno		Padrão Ouro = Maligno		Total
		Observed	Expected	Observed	Expected	
Step 1	1	1	2,307	4	2,693	5
	2	2	1,936	3	3,064	5
	3	2	1,876	3	3,124	5
	4	4	1,831	1	3,169	5
	5	0	1,783	5	3,217	5
	6	1	1,729	4	3,271	5
	7	3	1,681	2	3,319	5
	8	2	1,642	3	3,358	5
	9	2	1,586	3	3,414	5
	10	1	1,629	5	4,371	6

Classification Table^a

Observed	Padrão Ouro		Predicted		
			Padrão Ouro		Percentage Correct
			Benigno	Maligno	
Step 1	Padrão Ouro	Benigno	1	17	5,6
		Maligno	1	32	97,0
		Overall Percentage			64,7

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
		Step 1 ^a	MI02_CC	9,331	11,905	,614	1
	Constant	-,149	,998	,022	1	,881	,862

a. Variable(s) entered on step 1: MI02_CC.

MOMENTO INVARIANTE DE ORDEM 3

LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER MI03_CC /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

Logistic Regression

Notes

Input	Output Created	09-Jul-2015 17h24min40s
	Comments	
	Data	E:\Mestrado Ana Paula\Dados corrigidos.sav
	Active Dataset	DataSet1

	Filter	<none>	
	Weight	<none>	
	Split File	<none>	
	N of Rows in Working Data File		51
Missing Value Handling	Definition of Missing Syntax	User-defined missing values are treated as missing LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER MI03_CC /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).	
Resources	Processor Time		0:00:00.000
	Elapsed Time		0:00:00.000

[DataSet1] E:\Mestrado Ana Paula\Dados corrigidos.sav

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	51	100,0
	Missing Cases	0	,0
Total		51	100,0
Unselected Cases		0	,0
Total		51	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
Benigno	0
Maligno	1

Block 0: Beginning Block

Classification Table^{a,b}

			Predicted		
			Padrão Ouro		Percentage Correct
Observed		Benigno	Maligno		
Step 0	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
		Overall Percentage			64,7

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	,606	,293	4,279	1	,039	1,833

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	MI03_CC	,287	1	,592

Variables not in the Equation

	Score	df	Sig.
Step 0 Variables MI03_CC	,287	1	,592
Overall Statistics	,287	1	,592

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Step 1 Step	,295	1	,587
Block	,295	1	,587
Model	,295	1	,587

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	65,929 ^a	,006	,008

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Classification Table^a

Observed			Predicted		Percentage Correct
			Padrão Ouro		
	Benigno	Maligno			
Step 1 Padrão Ouro Benigno	0	18		,0	
Maligno	0	33		100,0	
Overall Percentage				64,7	

a. The cut value is ,500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a MI03_CC	6,483	12,181	,283	1	,595	653,932
Constant	,579	,297	3,799	1	,051	1,784

a. Variable(s) entered on step 1: MI03_CC.

MOMENTO INVARIANTE DE ORDEM 4 CORTE CENTRAL

LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER MI04_CC /SAVE=PRED /PRINT=GOODFIT CI(95) /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

Logistic Regression

Notes

Output Created	09-Jul-2015 19h43min45s
Comments	
Input Data	E:\Mestrado Ana Paula\Dados corrigidos.sav
Active Dataset	DataSet1
Filter	<none>
Weight	<none>
Split File	<none>
N of Rows in Working Data	51
File	

Missing Value Handling	Definition of Missing Syntax	User-defined missing values are treated as missing LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER MI04_CC /SAVE=PRED /PRINT=GOODFIT CI(95) /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
Resources	Processor Time	0:00:00.031
	Elapsed Time	0:00:00.039
Variables Created or Modified	PRE_23	Predicted probability

[DataSet1] E:\Mestrado Ana Paula\Dados corrigidos.sav

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	51	100,0
	Missing Cases	0	,0
	Total	51	100,0
Unselected Cases		0	,0
	Total	51	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
Benigno	0
Maligno	1

Block 0: Beginning Block

Classification Table^{a,b}

Observed			Predicted		
			Padrão Ouro		Percentage Correct
			Benigno	Maligno	
Step 0	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
		Overall Percentage			64,7

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	,606	,293	4,279	1	,039	1,833

Variables not in the Equation

	Score	df	Sig.
Step 0 Variables MI04_CC	,313	1	,576
Overall Statistics	,313	1	,576

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	,337	1	,562
	Block	,337	1	,562
	Model	,337	1	,562

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	65,886 ^a	,007	,009

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	13,915	8	,084

Contingency Table for Hosmer and Lemeshow Test

		Padrão Ouro = Benigno		Padrão Ouro = Maligno		Total
		Observed	Expected	Observed	Expected	
Step 1	1	0	2,026	5	2,974	5
	2	2	1,825	3	3,175	5
	3	2	1,806	3	3,194	5
	4	3	1,800	2	3,200	5
	5	4	1,793	1	3,207	5
	6	3	1,785	2	3,215	5
	7	0	1,778	5	3,222	5
	8	2	1,760	3	3,240	5
	9	1	1,714	4	3,286	5
	10	1	1,714	5	4,286	6

Classification Table^a

Observed			Predicted		
			Padrão Ouro		Percentage Correct
			Benigno	Maligno	
Step 1	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
		Overall Percentage			64,7

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	MI04_CC	11,123	20,282	,301	1	,583	67684,755
	Constant	,553	,306	3,266	1	,071	1,739

a. Variable(s) entered on step 1: MI04_CC.

MOMENTO INVARIANTE DE ORDEM 5 CORTE CENTRAL

LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER MI05_CC /SAVE=PRED /PRINT=GOODFIT CI(95) /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

Logistic Regression

		Notes
	Output Created	09-Jul-2015 19h44min42s
Input	Comments	
	Data	E:\Mestrado Ana Paula\Dados corrigidos.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	51
Missing Value Handling	Definition of Missing Syntax	User-defined missing values are treated as missing LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER MI05_CC /SAVE=PRED /PRINT=GOODFIT CI(95) /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
Resources	Processor Time	0:00:00.016
	Elapsed Time	0:00:00.019
Variables Created or Modified	PRE_24	Predicted probability

[DataSet1] E:\Mestrado Ana Paula\Dados corrigidos.sav

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	51	100,0
	Missing Cases	0	,0
	Total	51	100,0
Unselected Cases		0	,0
	Total	51	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
Benigno	0
Maligno	1

Block 0: Beginning Block

Classification Table^{a,b}

Observed	Predicted	
	Padrão Ouro	

			Benigno	Maligno	Percentage Correct
Step 0	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
		Overall Percentage			64,7

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	,606	,293	4,279	1	,039	1,833

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	MI05_CC	,448	1	,503
		Overall Statistics	,448	1	,503

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	,473	1	,491
	Block	,473	1	,491
	Model	,473	1	,491

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	65,750 ^a	,009	,013

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	11,924	8	,155

Contingency Table for Hosmer and Lemeshow Test

		Padrão Ouro = Benigno		Padrão Ouro = Maligno		Total
		Observed	Expected	Observed	Expected	
Step 1	1	0	2,116	5	2,884	5
	2	3	1,923	2	3,077	5
	3	1	1,889	4	3,111	5
	4	4	1,840	1	3,160	5
	5	2	1,777	3	3,223	5
	6	3	1,759	2	3,241	5
	7	2	1,735	3	3,265	5
	8	1	1,702	4	3,298	5
	9	1	1,645	4	3,355	5
	10	1	1,614	5	4,386	6

Classification Table^a

Observed			Predicted		
			Padrão Ouro		Percentage Correct
			Benigno	Maligno	
Step 1	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
		Overall Percentage			64,7

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	MI05_CC	46,598	70,662	,435	1	,510	1,728E20
	Constant	,564	,299	3,566	1	,059	1,758

a. Variable(s) entered on step 1: MI05_CC.

MOMENTO INVARIANTE DE ORDEM 6 CORTE CENTRAL

LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER MI06_CC /SAVE=PRED /PRINT=GOODFIT CI(95) /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

Logistic Regression

Notes

	Output Created	09-Jul-2015 19h45min45s
Input	Comments	
	Data	E:\Mestrado Ana Paula\Dados corrigidos.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	51
Missing Value Handling	Definition of Missing Syntax	User-defined missing values are treated as missing LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER MI06_CC /SAVE=PRED /PRINT=GOODFIT CI(95) /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
Resources	Processor Time	0:00:00.016
	Elapsed Time	0:00:00.020
Variables Created or Modified	PRE_25	Predicted probability

[DataSet1] E:\Mestrado Ana Paula\Dados corrigidos.sav

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	51	100,0
	Missing Cases	0	,0
	Total	51	100,0

Unselected Cases	0	,0
Total	51	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
Benigno	0
Maligno	1

Block 0: Beginning Block

Classification Table^{a,b}

Observed			Predicted		
			Padrão Ouro		Percentage Correct
	Benigno	Maligno			
Step 0	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
		Overall Percentage			64,7

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	
Step 0	Constant	,606	,293	4,279	1	,039	1,833

Variables not in the Equation

	Score	df	Sig.		
Step 0	Variables	MI06_CC	,077	1	,781
	Overall Statistics		,077	1	,781

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.	
Step 1	Step	,078	1	,780
	Block	,078	1	,780
	Model	,078	1	,780

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	66,145 ^a	,002	,002

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	15,850	8	,045

Contingency Table for Hosmer and Lemeshow Test

		Padrão Ouro = Benigno		Padrão Ouro = Maligno		Total
		Observed	Expected	Observed	Expected	
Step 1	1	0	1,890	5	3,110	5
	2	1	1,853	4	3,147	5
	3	2	1,829	3	3,171	5
	4	3	1,808	2	3,192	5
	5	4	1,786	1	3,214	5
	6	1	1,760	4	3,240	5
	7	1	1,738	4	3,262	5
	8	4	1,722	1	3,278	5
	9	1	1,702	4	3,298	5
	10	1	1,913	5	4,087	6

Classification Table^a

Observed			Predicted		
			Padrão Ouro		Percentage Correct
Benigno	Maligno				
Step 1	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
		Overall Percentage			64,7

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	MI06_CC	20,638	74,540	,077	1	,782	9,187E8
	Constant	,592	,297	3,977	1	,046	1,808

a. Variable(s) entered on step 1: MI06_CC.

MOMENTO INVARIANTE DE ORDEM 7 CORTE CENTRAL

LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER MI07_CC /SAVE=PRED /PRINT=GOODFIT CI(95) /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

Logistic Regression**Notes**

	Output Created	09-Jul-2015 19h47min3s
Input	Comments	
	Data	E:\Mestrado Ana Paula\Dados corrigidos.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	51
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing

Syntax	LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER MI07_CC /SAVE=PRED /PRINT=GOODFIT CI(95) /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).		
Resources	Processor Time		0:00:00.031
	Elapsed Time		0:00:00.020
Variables Created or Modified	PRE_26	Predicted probability	

[DataSet1] E:\Mestrado Ana Paula\Dados corrigidos.sav

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	51	100,0
	Missing Cases	0	,0
	Total	51	100,0
Unselected Cases		0	,0
	Total	51	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
Benigno	0
Maligno	1

Block 0: Beginning Block

Classification Table^{a,b}

Observed			Predicted		
			Padrão Ouro		Percentage Correct
			Benigno	Maligno	
Step 0	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
		Overall Percentage			64,7

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	,606	,293	4,279	1	,039	1,833

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	MI07_CC	1,926	1	,165
		Overall Statistics	1,926	1	,165

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	2,435	1	,119
	Block	2,435	1	,119
	Model	2,435	1	,119

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	63,789 ^a	,047	,064

a. Estimation terminated at iteration number 8 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	9,201	8	,326

Contingency Table for Hosmer and Lemeshow Test

		Padrão Ouro = Benigno		Padrão Ouro = Maligno		Total
		Observed	Expected	Observed	Expected	
Step 1	1	2	2,754	3	2,246	5
	2	3	1,872	2	3,128	5
	3	1	1,770	4	3,230	5
	4	1	1,728	4	3,272	5
	5	1	1,685	4	3,315	5
	6	4	1,636	1	3,364	5
	7	2	1,605	3	3,395	5
	8	1	1,571	4	3,429	5
	9	2	1,552	3	3,448	5
	10	1	1,827	5	4,173	6

Classification Table^a

Observed			Predicted		
			Padrão Ouro		Percentage Correct
			Benigno	Maligno	
Step 1	Padrão Ouro	Benigno	2	16	11,1
		Maligno	0	33	100,0
		Overall Percentage			68,6

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	MI07_CC	-67,678	120,168	,317	1	,573	,000
	Constant	,891	,507	3,087	1	,079	2,437

a. Variable(s) entered on step 1: MI07_CC.

HOMOGENEIDADE CORTE CENTRAL

LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER H_CC /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

Logistic Regression

		Notes
	Output Created	09-Jul-2015 17h28min3s
Input	Comments	
	Data	E:\Mestrado Ana Paula\Dados corrigidos.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	51
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing
	Syntax	LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER H_CC /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
Resources	Processor Time	0:00:00.016
	Elapsed Time	0:00:00.017

[DataSet1] E:\Mestrado Ana Paula\Dados corrigidos.sav

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	51	100,0
	Missing Cases	0	,0
	Total	51	100,0
Unselected Cases		0	,0
	Total	51	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
Benigno	0
Maligno	1

Block 0: Beginning Block

Classification Table^{a,b}

			Predicted		
			Padrão Ouro		Percentage Correct
Observed		Benigno	Maligno		
Step 0	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
	Overall Percentage				64,7

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	,606	,293	4,279	1	,039	1,833

Variables not in the Equation

	Score	df	Sig.
Step 0 Variables H_CC	,211	1	,646
Overall Statistics	,211	1	,646

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Step 1 Step	,215	1	,643
Block	,215	1	,643
Model	,215	1	,643

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	66,008 ^a	,004	,006

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Classification Table^a

Observed			Predicted		
			Padrão Ouro		Percentage Correct
	Benigno	Maligno			
Step 1 Padrão Ouro Benigno	0	18		,0	
Maligno	0	33		100,0	
Overall Percentage				64,7	

a. The cut value is ,500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a H_CC	1,486	3,242	,210	1	,647	4,421
Constant	-,236	1,854	,016	1	,899	,790

a. Variable(s) entered on step 1: H_CC.

ENTROPIA CORTE CENTRAL

LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER E_CC /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

Logistic Regression

Notes

Output Created	09-Jul-2015 17h27min29s
Comments	

Input	Data	E:\Mestrado Ana Paula\Dados corrigidos.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	51
Missing Value Handling	Definition of Missing Syntax	User-defined missing values are treated as missing LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER E_CC /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
Resources	Processor Time	0:00:00.016
	Elapsed Time	0:00:00.017

[DataSet1] E:\Mestrado Ana Paula\Dados corrigidos.sav

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	51	100,0
	Missing Cases	0	,0
	Total	51	100,0
Unselected Cases		0	,0
	Total	51	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
Benigno	0
Maligno	1

Block 0: Beginning Block

Classification Table^{a,b}

Observed			Predicted		
			Padrão Ouro		Percentage Correct
			Benigno	Maligno	
Step 0	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
		Overall Percentage			64,7

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	,606	,293	4,279	1	,039	1,833

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	E_CC	3,494	1	,062
Overall Statistics			3,494	1	,062

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	3,728	1	,054
	Block	3,728	1	,054
	Model	3,728	1	,054

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	62,495 ^a	,070	,097

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Classification Table^a

Observed			Predicted		
			Padrão Ouro		Percentage Correct
			Benigno	Maligno	
Step 1	Padrão Ouro	Benigno	4	14	22,2
		Maligno	2	31	93,9
Overall Percentage					68,6

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	E_CC	-5,747	3,188	3,250	1	,071	,003
	Constant	4,818	2,387	4,073	1	,044	123,705

a. Variable(s) entered on step 1: E_CC.

VARIÂNCIA CORTE CENTRAL

LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER V_CC /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

Logistic Regression

Notes

Input	Output Created	09-Jul-2015 17h26min51s
	Comments	
	Data	E:\Mestrado Ana Paula\Dados corrigidos.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	51

Missing Value Handling	Definition of Missing Syntax	User-defined missing values are treated as missing LOGISTIC REGRESSION VARIABLES PO /METHOD=ENTER V_CC /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
Resources	Processor Time	0:00:00.015
	Elapsed Time	0:00:00.016

[DataSet1] E:\Mestrado Ana Paula\Dados corrigidos.sav

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	51	100,0
	Missing Cases	0	,0
	Total	51	100,0
Unselected Cases		0	,0
	Total	51	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
Benigno	0
Maligno	1

Block 0: Beginning Block

Classification Table^{a,b}

			Predicted		
			Padrão Ouro		Percentage Correct
Observed		Benigno	Maligno		
Step 0	Padrão Ouro	Benigno	0	18	,0
		Maligno	0	33	100,0
Overall Percentage					64,7

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	,606	,293	4,279	1	,039	1,833

Variables not in the Equation

		Score	df	Sig.
Step 0	Variables	V_CC	2,550	,110
		Overall Statistics	2,550	,110

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	2,570	1	,109
	Block	2,570	1	,109
	Model	2,570	1	,109

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	63,653 ^a	,049	,068

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Classification Table^a

Observed			Predicted		
			Padrão Ouro		Percentage Correct
			Benigno	Maligno	
Step 1	Padrão Ouro	Benigno	3	15	16,7
		Maligno	3	30	90,9
		Overall Percentage			64,7

a. The cut value is ,500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	
Step 1 ^a	V_CC	,000	,000	2,456	1	,117	1,000
	Constant	1,927	,912	4,463	1	,035	6,866

a. Variable(s) entered on step 1: V_CC.