

**UNIVERSIDADE ESTADUAL DE SANTA CRUZ**

**ELINALDO SANTOS DE GOES JUNIOR**

**CLASSIFICAÇÃO DE NÓDULOS MAMÁRIOS UTILIZANDO CÁLCULO DE  
DIMENSÃO FRACTAL**

**ILHÉUS – BAHIA  
2015**

**ELINALDO SANTOS DE GOES JUNIOR**

**CLASSIFICAÇÃO DE NÓDULOS MAMÁRIOS UTILIZANDO CÁLCULO DE  
DIMENSÃO FRACTAL**

Dissertação apresentada como exigência parcial para obtenção do título de Mestre em Modelagem Computacional em Ciência e Tecnologia da Universidade Estadual de Santa Cruz.

Orientador: Prof. Paulo Eduardo Ambrósio

**ILHÉUS – BAHIA  
2015**

G598 Goes Júnior, Elinaldo Santos de.  
Classificação de nódulos mamários utilizando  
cálculo de dimensão fractal / Elinaldo Santos de Goes  
Júnior. – Ilhéus, BA: UESC, 2015.  
xvii, 91 f. : il.

Orientador: Paulo Eduardo Ambrósio.  
Dissertação (mestrado) – Universidade Estadual  
de Santa Cruz. Programa de Pós-Graduação em Mo-  
delagem Computacional em Ciência e Tecnologia.  
Referências: f. 87-91.

1. Mamas – Câncer. 2. Mamas – Câncer – Diag-  
nóstico. 3. Mamas – Câncer – Tratamento. 4. Proces-  
samento de imagens. 5. Fractais. I. Título.

CDD 616.994

ELINALDO SANTOS DE GOES JUNIOR

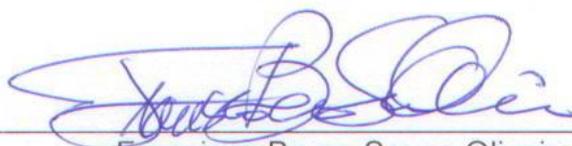
CLASSIFICAÇÃO DE NÓDULOS MAMÁRIOS UTILIZANDO CÁLCULO DE  
DIMENSÃO FRACTAL

Ilhéus, 26/01/2015.



---

Paulo Eduardo Ambrósio – Orientador  
UESC/PPGMC



---

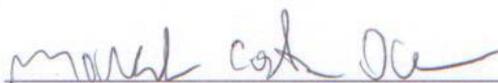
Francisco Bruno Souza Oliveira  
UESC/PPGMC



Susana M. Iglesias  
Mat. 73438702-5

---

Susana Marrero Iglesias  
UESC/PPGMC



---

Marcelo Costa Oliveira  
UFAL

Aos meus pais, por todo ensinamento que me deram e continuam me dando, e por estarem sempre ao meu lado.

A Camila Moura por que sempre torceu por mim e me fez acreditar que eu era capaz.

## AGRADECIMENTOS

Aos meus pais, Elinaldo Santos de Goes e Ariadalva Pereira Novais, pelo amor incondicional e por sempre acreditarem em mim, mesmo nos momentos mais difíceis. Serei sempre grato.

Ao meu avô, Estevam, por me ensinar que nos momentos difíceis é possível sorrir verdadeiramente, e à minha avó Lourdes, por me ensinar que não é vergonha chorar nos momentos felizes.

Aos meus irmãos Igor e Roberto, que me incentivaram a correr atrás dos meus objetivos, sendo, além de irmãos, grandes amigos. Obrigado pelos momentos de descontração e cumplicidade.

A todos os meus familiares, que mesmo quando longe, sempre estiveram presentes, me dando forças e torcendo pelo meu sucesso.

À minha namorada Camila Moura, por sempre me surpreender com seu carinho e sempre me estender a mão nos momentos que mais preciso.

A todos os meus amigos, por sempre torcerem pelo meu sucesso.

Ao meu orientador Paulo Eduardo Ambrósio, pelo empenho, paciência, orientação e oportunidades que me proporcionou.

A todos os docentes, discentes e funcionários do PPGMC, pelos ensinamentos e convivência durante esse período.

Obrigado.

*“A caverna que você tem medo de entrar guarda o tesouro que procura.”*

(Joseph Campbell)

## RESUMO

Nos últimos anos o número de pessoas afetadas pelo câncer de mama tem aumentado progressivamente, em destaque indivíduos do gênero feminino. Portanto, diante dessa elevação da incidência do câncer de mama, aliada a possíveis erros de diagnóstico e/ou diagnóstico tardio, faz-se necessário a realização de pesquisas que visem oferecer uma ferramenta que tenha a possibilidade de auxiliar e potencializar o diagnóstico médico por meio da técnica de processamento de imagens. Deste modo, o objetivo dessa pesquisa centrou-se em verificar se a técnica de processamento de imagens com base no método de *box-counting* seria eficaz para a análise de imagens de exames de mamografia. Para a realização dos experimentos computacionais foram utilizadas 142 imagens mamográficas com diagnóstico previamente estabelecido por especialistas. Estas imagens foram escolhidas com base em sua classificação de anormalidade, disponíveis nas bases *MIAS* e *BancoWeb*, as quais possuem acesso livre. Em seguida foram aplicadas a segmentação por limiar simples e por crescimento de regiões. Depois foi aplicado o filtro de detecção de bordas pelo algoritmo de *Canny* em ambos os métodos e por fim foi empregado o método *box-counting*. Avaliação do sistema foi baseada na análise de curvas *ROC*, com o intuito de conferir uma maior fidedignidade aos dados obtidos. A segmentação por limiarização foi capaz de detectar área correspondente ao nódulo em 49% dos nódulos benignos e 62,2% dos nódulos malignos. O método de crescimento de regiões apresentou uma taxa de 62,8% em relação aos nódulos benignos e 71,1% em relação aos nódulos malignos. O cálculo da dimensão fractal do contorno de nódulos pelo método de *box-counting* para um *cut off* = 1, foi capaz de classificar corretamente aproximadamente 62,7%, para o *cut off* = 1,1 o método obteve resultados satisfatório, notabilizando um desempenho de aproximadamente 77,9%, evidenciando que o método pode ser uma alternativa promissora para a análise nódulos mamários.

Palavras-chave: Dimensão fractal. Diagnóstico auxiliado por computador. Processamento de imagens. Câncer de mama.

## ABSTRACT

In recent years the number of people affected by breast cancer has increased progressively highlighted individuals females. Therefore, with this increased incidence of breast cancer, combined with possible errors in diagnosis and/or late diagnosis, it is necessary to conduct a survey aimed to offer a tool that is able to assist and enhance the medical diagnostic means of image processing technique. Thus, the goal of this research has focused on checking if the image processing technique based on the box-counting method would be effective for the analysis of mammography images. For the realization of computational experiments were used 142 mammographic images with diagnosis was established by experts. These images were chosen based on their abnormality classification, available in the *MIAS* and *BancoWeb* base, which have free access. They were then applied to targeting threshold and growth regions. After we applied the edge detection filter by *Canny* algorithms in both methods and finally we used the box-counting method. System evaluation was based on the analysis of ROC curves in order to provide greater reliability to the data obtained. Targeting threshold was able to detect the area corresponding to the node in 49% of benign nodules and 62.2% of malignant nodules. The zone growth method showed a rate of 62.8% compared to benign tumors and 71.1% from malignant nodules. The calculation of the fractal dimension of the boundary nodes by the box-counting method for a cut off = 1, was able to correctly classify approximately 62.7%. To cut off = 1.1 the method achieved satisfactory results, featuring a performance of approximately 77.9%, showing that the method may be a promising alternative for the analysis breast lumps.

Keywords: Fractal dimension. Computer-Aided Diagnosis. Image processing. Breast cancer.

## LISTA DE FIGURAS

Figura 1 - Características de contorno que definem nódulos malignos (KOPANS, 2000).....	20
Figura 2 - a) Pixels em tons de cinza; b) Pixels em valores numéricos .....	24
Figura 3 - Processo iterativo para criação de fractais artificiais; a) Ilha de Gosper; b) Floco de neve de Koch; c) Tapete de Sierpinski; d) Triângulo de Sierpinski. ....	27
Figura 4 - Elementos naturais que apresentam padrão fractal; a) Galhos de uma árvore; b) Relâmpagos; c) Floco de neve; d) Repolho. ....	28
Figura 5 - Processo iterativo para gerar a curva de Koch em 5 iterações.....	29
Figura 6 - Aplicação do método de box counting no contorno do mapa do Brasil.....	32
Figura 7 - Histograma de uma imagem em tons de cinza. O eixo x representa as 256 tonalidades de cinza e o eixo y representa a quantidade de pixels que tem essa tonalidade .....	37
Figura 8 - Limiarização pelo método de <i>thresholding</i> . a) Histograma da imagem; b) histograma da imagem binarizada .....	41
Figura 9 - Imagem com borda bem definida, caracterizada pela mudança brusca entre as tonalidades de cinza; b) Borda mal definida, caracterizada pela transição suave entre os níveis de cinza. ....	43
Figura 10 - Curvas ROC de três estudos diferentes .....	49
Figura 11 - Mamografia em que existe a necessidade de remover ruídos da imagem, como o identificador no canto superior direito. ....	53
Figura 12 – a) Histograma original - A esquerda podemos observar a imagem e seu histograma com baixo nível de contraste; b) Histograma ajustada - A direita podemos ver a imagens com nível de contraste ajustado e seu histograma bem distribuído. ....	54
Figura 13 - Primeira limiarização realizada na imagem para destaque da região da mama. <i>Threshold</i> de 0,22. ....	55
Figura 14 - Aplicação do filtro morfológico de abertura .....	56
Figura 15 – a) imagem inicial; b) imagem após a passagem do filtro morfológico; c) imagem final. Multiplicação ponto a ponto entre a imagem inicial e a imagem filtrada, resultando na imagem final sem ruídos.....	57

Figura 16 - Imagem equalizada.....	58
Figura 17 - Limiarização da mamografia para identificação do nódulo. Além do nódulo, parte do músculo peitoral (canto superior esquerdo) e parte do tecido glandular foram selecionados. ....	59
Figura 18 - Imagem após os métodos de remoção de pequenas e grandes áreas ..	60
Figura 19- a) Imagem original, em formato PGM; b) Imagem pré-processada, em formato PNG e com eixo y invertido.....	62
Figura 20 - Escolha das coordenadas x e y que determinam a semente para a segmentação por crescimento de regiões. Para essa imagem a semente está localizada na posição 582x265. ....	62
Figura 21 - Segmentação do nódulo por crescimento de regiões.....	64
Figura 22 - Detecção de bordas do nódulo utilizando o método de Canny.....	65
Figura 23- Mamografia mdb204 - Exemplo de teste em que não foi reconhecido nenhum nódulo na mama. a) Mamografia original; b) Primeira segmentação; c) Eliminação de áreas que não pertencem à mama; d) Segunda segmentação; e) Remoção de áreas pequenas; f) Remoção de áreas grandes, resultando na não identificação do nódulo. ....	68
Figura 24 - Mamografia mdb314. Exemplo de teste em que foi detectada uma área diferente da área do nódulo. a) Mamografia original; b) Primeira segmentação; c) Eliminação de áreas que não pertencem à mama; d) Segunda segmentação; e) Remoção de áreas pequenas; f) Remoção de áreas grandes, resultando na identificação de uma área não correspondente ao nódulo. ....	69
Figura 25 - Mamografia mdb315. Exemplo em que o nódulo mamário foi segmentado corretamente, porém uma área não correspondente ao nódulo também foi identificada. a) Mamografia original; b) Primeira segmentação; c) Eliminação de áreas que não pertencem à mama; d) Segunda segmentação; e) Remoção de áreas pequenas; f) Remoção de áreas grandes, resultando na identificação do nódulo e de outras regiões. ....	69
Figura 26 - Mamografia mdb025. Exemplo de segmentação em que apenas o nódulo foi localizado ao final do processo. a) Mamografia original; b) Primeira segmentação; c) Eliminação de áreas que não pertencem à mama; d) Segunda segmentação; e) Remoção de áreas pequenas; f) Remoção de áreas grandes, resultando na identificação do nódulo. ....	70

Figura 27 - Mamografia mdb075. Aplicação do teste em mamografia contendo nódulo maligno, onde só foi identificado o músculo peitoral, caracterizando erro. a) Mamografia original; b) Primeira segmentação; c) Eliminação de áreas que não pertencem à mama; d) Segunda segmentação; e) Remoção de áreas pequenas; f) Remoção de áreas grandes, resultando na não identificação do nódulo.....	71
Figura 28 - Mamografia mdb095. Aplicação em que além do nódulo foram identificadas outras áreas da mama. a) Mamografia original; b) Primeira segmentação; c) Eliminação de áreas que não pertencem à mama; d) Segunda segmentação; e) Remoção de áreas pequenas; f) Remoção de áreas grandes, resultando na identificação do nódulo e de outras áreas.....	72
Figura 29 - mamografia mdb184. Aplicação do teste onde o nódulo foi identificado corretamente de acordo com o diagnóstico a) Mamografia original; b) Primeira segmentação; c) Eliminação de áreas que não pertencem à mama; d) Segunda segmentação; e) Remoção de áreas pequenas; f) Remoção de áreas grandes, resultando na identificação correta do nódulo. ....	73
Figura 30 - Mamografia mdb015. Segmentação por crescimento de regiões em que a região segmentada não corresponde a área do nódulo. a) Mamografia original; b) Indicação da região do nódulo; c) Escolha da semente; d) Segmentação.....	74
Figura 31 - Mamografia mdb002. Segmentação em que o resultado acertou parcialmente a área ocupada pelo nódulo.....	75
Figura 32 - Mamografia mdb126. Resultado em que a segmentação acerta totalmente na detecção do nódulo. ....	75
Figura 33 - Mamografia mdb023. Segmentação por crescimento de regiões em que a região segmentada não corresponde a área do nódulo maligno. ....	76
Figura 34 - Mamografia mdb002. Resultado em que a área do nódulo foi selecionada, porém passando um pouco do limite. ....	76
Figura 35 - Mamografia 126. A segmentação ocorreu exatamente na região do nódulo especificado. ....	77
Figura 36 - a) Nódulo maligno; Detecção de bordas pelos métodos de: b) Prewitt; c) Canny; d) Roberts; c) Sobel .....	77
Figura 37 - Dispersão das dimensões fractais de nódulos benignos .....	79
Figura 38 - Dispersão das dimensões fractais de nódulos malignos .....	80

Figura 39 - Gráfico da curva ROC, estabelecendo relação entre Sensibilidade e Especificidade.....	82
Figura 40 - Curva ROC para valor de <i>cut off</i> = 1,1. ....	83

## LISTA DE TABELAS

Tabela 1 - Relação de mamografias analisadas, separadas por suas características. .....	66
Tabela 2 - Percentagem de benignas e malignas separadas por suas características .....	66
Tabela 3 - Distribuições das mamografias da base BancoWeb com base em sua classificação BIRADS .....	67
Tabela 4 - Resultado da segmentação pelo método de <i>threshold</i> para nódulos benignos .....	67
Tabela 5 - Resultado da segmentação pelo método de <i>threshold</i> para nódulos malignos .....	71
Tabela 6 - Resultado da segmentação pelo método de crescimento de regiões para nódulos benignos .....	74
Tabela 7 - Resultado da segmentação pelo método de crescimento de regiões para nódulos malignos .....	74
Tabela 8 - Dimensão fractal de nódulos benignos, segmentados pelo método de detecção de bordas de Canny. Maior DF: mdb198; Menor DF: mdb099. ....	78
Tabela 9 - Dimensão fractal de nódulos malignos, segmentados pelo método de detecção de bordas de Canny. Maior DF: mdb148; Menor DF: dmb206; .....	78
Tabela 10 – Matriz de confusão para limiar igual a 1 .....	80
Tabela 11 - Análise das probabilidades em relação à Sensibilidade ( $P+$ ) e Especificidade ( $P-$ ) .....	81
Tabela 12 - Matriz de confusão para <i>cut off</i> = 1,1. ....	83

**LISTA DE QUADROS**

Quadro 1 - Classificações BIRADS e suas descrições .....	23
Quadro 2 - Matriz de confusão .....	47
Quadro 3 – Conjunto de características a respeito dos exames de mamografias da base MIAS .....	51

**LISTA DE SIGLAS**

BIRADS	<i>Breast Image Reporting and Data System</i>
CAD	<i>Computer-aided diagnosis</i>
DF	Dimensão Fractal
FN	Falso Negativo
FP	Falso Positivo
IBM	<i>International Business Machines</i>
INCA	Intituto Nacional do Câncer
MIAS	<i>Mammographic Image Analysis Society</i>
OMS	Organização Mundial da Saúde
PGM	<i>Portable Gray Map</i>
PNG	<i>Portable Network Graphics</i>
ROC	<i>Receiver Operating Characteristic</i>
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo
TIFF	<i>Tagged Image File Format</i>

## SUMÁRIO

INTRODUÇÃO.....	16
1. BASE CONCEITUAL.....	19
1.1. Câncer de Mama .....	19
1.1.1. Câncer de Mama no Brasil .....	20
1.1.2. Fatores de risco.....	21
1.2. Mamografia.....	22
1.2.1. Classificação BIRADS .....	22
1.2.2. Características da Imagem Mamográfica.....	23
1.2.3. Qualidade da Imagem do Mamograma.....	24
1.3. Fractais.....	26
1.3.1. Espaço Fractal.....	28
1.3.2. Dimensão Fractal .....	30
1.3.3. Box-Counting.....	31
1.4. Apoio a Diagnóstico por Computador.....	34
1.4.1. Aquisição das Imagens.....	35
1.4.2. Pré-processamento .....	35
1.4.3. Segmentação .....	39
1.4.4. Classificação .....	44
1.4.5. Avaliação do Sistema .....	46
2. materiais e métodos .....	50
2.1. Obtenção das imagens .....	50
2.1.1. Base MIAS .....	50
2.1.2. Base BancoWeb.....	52
2.2. Método Automático para Classificação de Mamogramas.....	52

2.2.1. Identificação da área da mama.....	52
2.2.2. Identificação dos nódulos .....	57
2.3. Método Semi-Automático de Classificação de Mamogramas.....	61
2.3.1. Pré-processamento .....	61
2.3.2. Escolha da semente .....	62
2.3.3. Segmentação das mamografias .....	63
2.4. Filtro de Detecção de Bordas.....	64
2.5. Aplicação do <i>Box-Counting</i> .....	65
3. RESULTADOS E DISCUSSÕES.....	66
4. CONCLUSÃO.....	85
4.1. Trabalhos Futuros .....	86
5. REFERÊNCIAS.....	87

## INTRODUÇÃO

Avanços computacionais têm colaborado para o melhor entendimento em diversas áreas do conhecimento. Na saúde esses avanços vêm permitindo que decisões sejam tomadas em menos tempo e com um nível de acerto cada vez maior, o que em muitos casos pode ser indispensável, por se tratar da vida de seres humanos. O setor de saúde tem sido constantemente alvo de preocupação e de investimentos por parte de vários órgãos nacionais e internacionais. O principal objetivo desses investimentos é, além de prevenir doenças, tornar cada vez mais rápido e confiável o diagnóstico médico, permitindo a detecção precoce das mais diversas enfermidades, principalmente as que têm melhores chances de cura quando diagnosticadas em seu estágio inicial, como o câncer.

O câncer de mama é uma enfermidade que se manifesta com maior frequência em pessoas do gênero feminino, tendo como maior incidência mulheres com idade por volta dos 50 anos (VIEIRA *et al.*, 2012). A elevada incidência, associada com a alta taxa de mortalidade e custo elevado do tratamento, são variáveis que convergem para que essa doença torne-se um complexo e preocupante quadro clínico tanto para os sistemas de saúde de países desenvolvidos, como em desenvolvimento.

Os meios mais eficazes para combater o câncer de mama são baseados em sua detecção precoce que pode ser realizada através do auto-exame, exame clínico e mamografia. Dentre essas técnicas, a mamografia é a que permite um diagnóstico mais preciso, pois por meio desse exame é possível detectar lesões muito pequenas, imperceptíveis ao toque até mesmo de um especialista (CRISAN *et al.*, 2007). Apesar de sua eficácia, ainda existe certo grau de incerteza agregado a esse exame, devido a falhas que podem ocorrer na identificação dos nódulos.

Durante a realização do diagnóstico, o especialista pode interpretar equivocadamente um exame, caracterizando um nódulo benigno, onde não haveria necessidade de biópsia, como um nódulo maligno, levando o paciente à realização de uma biópsia desnecessária. A esse fato dá-se o nome de falso positivo. O segundo caso é ainda mais prejudicial para o paciente, e ocorre quando o especialista identifica como positiva, uma lesão maligna. Essa situação é conhecida por sua vez como falso negativo.

Ambos os casos são muito prejudiciais ao paciente pelo caráter da doença e pelo diagnóstico errôneo, o que pode custar uma biópsia desnecessária, em caso de falsos positivos, ou um tratamento tardio que pode levar o indivíduo a óbito, no caso dos falsos negativos.

Por esses motivos, e pelas características do câncer de mama, existe uma comoção no intuito de promover o desenvolvimento de ferramentas capazes de contribuir com o diagnóstico, fornecendo subsídios ao especialista, para que esse possa tomar uma decisão mais acertada (DOI, 2007). Esses sistemas são sistemas de apoio ao diagnóstico (*CADx – Computer Aided Diagnosis*).

Através de sistemas *CADx*, que também se utilizam de técnicas de processamento de imagens, é possível detectar e extrair características não perceptíveis ao especialista, em relação a uma região de interesse da mamografia, e a partir delas aplicar métodos de classificação para diferenciação entre nódulos malignos e benignos, oferecendo um recurso adicional ao radiologista na detecção do câncer.

Diante desse cenário, essa pesquisa foi realizada visando validar um método que possibilite o auxílio e a potencialização de acertos no diagnóstico médico por meio da técnica de processamento de imagens para identificação do nódulo, e para a classificação, por meio de técnicas de análise de dimensão fractal. Portanto esta pesquisa tem como um dos focos verificar, através dos conceitos de fractais, a utilização da técnica de *boxcounting* para cálculo de dimensão, visando a classificação de nódulos mamários, em mamografias obtidas através de bases de imagens livres.

O segundo foco desse trabalho consistiu no desenvolvimento e utilização de técnicas de processamento de imagens para segmentação automática e semiautomática de mamografias, e na aplicação da técnica desenvolvida em (MARTINS *et al.*, 2014) para segmentação por crescimento de regiões, incorporando ao sistema desenvolvido funcionalidades de auxílio ao diagnóstico.

A fundamentação teórica que oferece sustentação a esse estudo foi constituída e organizada em quatro subcapítulos. O primeiro subcapítulo foi destinado à explanação de conteúdos relacionados ao câncer de mama, tais como conceito, classificação dos nódulos, fatores de risco e panorama do câncer de mama no Brasil.

Já o segundo subcapítulo apresenta a discussão acerca da mamografia, tendo como principais fatores destacados o conceito, características, caracterização das imagens e a qualidade da imagem do mamograma.

No terceiro subcapítulo são apresentados os conceitos de fractais, aplicações na área médica, espaços fractais, dimensão dos fractais, bem como foi explanado acerca da técnica do *boxcounting*.

Por fim, foram discutidos os fatores atrelados a técnica de processamento de imagens tais como: finalidade, tipos e etapas de processamento, aquisição das imagens, pré-processamento, segmentação das imagens, classificação e avaliação do sistema.

## **1. BASE CONCEITUAL**

Este capítulo está dividido em quatro seções. A primeira seção descreve brevemente o câncer de mama, suas causas e consequências. A segunda seção define características acerca do exame de mamografia. A terceira descreve fractais e apresenta noções sobre análise de dimensão fractal. Por fim, a quarta seção explica técnicas de processamento utilizadas nos trabalhos relacionadas a sistemas CADx.

### **1.1. Câncer de Mama**

O câncer consiste no crescimento acelerado e desordenado das células de uma região, que resulta no surgimento de nódulos (INCA, 2014a). O câncer de mama se caracteriza por esse crescimento no tecido da mama. O nódulo é caracterizado como maligno quando suas células possuem potencial para se espalharem por outras partes do corpo, processo conhecido com metástase (SANTOS, 2002).

Quando células malignas adentram na corrente sanguínea podem contaminar tecidos e órgãos saudáveis, causando novos nódulos. Esse processo é crítico e pode levar o indivíduo a óbito (NORVARTIS ONCOLOGY).

Existem casos em que os nódulos não possuem a capacidade de realizar metástase, ou seja, são nódulos benignos. A taxa de crescimento desse tipo de nódulo é mais lenta e frequentemente apresentam uma dimensão menor que os nódulos considerados malignos.

Apesar de os nódulos benignos não serem cancerígenos como os nódulos malignos, eles se caracterizam inicialmente pelo crescimento de suas células. Esse agrupamento de células, quando identificados inicialmente, podem aumentar significativamente as chances de cura em caso de câncer de mama.

Para o diagnóstico de nódulos malignos o especialista deve analisar três principais características: tamanho, densidade e contorno (TABAR & DEAN, 1985). Dentre essas variáveis, contorno e forma dos nódulos são as principais características avaliadas nesse trabalho.

A Figura 1 ilustra a relação entre o contorno e a associação da malignidade do nódulo:

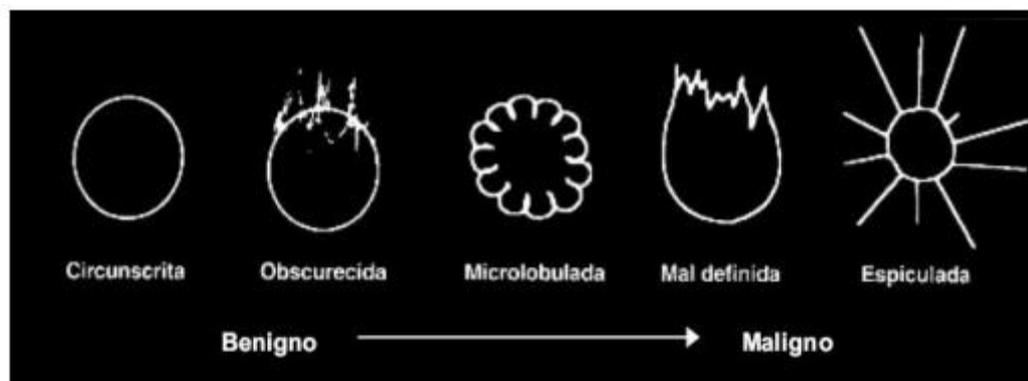


Figura 1 - Características de contorno que definem nódulos malignos (KOPANS, 2000)

Como pode ser visto na Figura 1 os nódulos com formato circular e contornos lisos são geralmente associados a casos benignos. Porém, quanto mais irregular essa borda se torna, maiores são as chances desse nódulo ser maligno (MEDICA, 2007).

Apesar da definição das características do contorno do nódulo e dos exames de diagnósticos, existem limitações quando se trata da diferenciação de lesões benignas e malignas, através do exame, pois em muitos casos as bordas dos nódulos não são bem definidas na imagem, levando a muitos diagnósticos por vezes equivocados.

#### 1.1.1. Câncer de Mama no Brasil

O índice de pessoas acometidas com o câncer de mama vem ascendendo progressivamente ao longo dos anos. Atualmente é o segundo tipo de câncer mais frequente em todo o mundo, sendo caracterizado como “de extrema importância para a saúde pública em nível mundial” (SCLOWITZ *et al.*, 2005).

No Brasil, os números do câncer de mama também são alarmantes e vêm causando preocupação aos órgãos de saúde pública nacionais, graças a sua alta incidência e mortalidade, e aos elevados custos em seu tratamento. Segundo o Instituto Nacional do Câncer (INCA), forma estimados cerca de 57.120 casos só no

ano de 2014, com um risco estimado de 56,09 novos casos de câncer de mama a cada 100 mil mulheres no Brasil. Já em 2011 foram computados aproximadamente 13.000 registros de óbitos em decorrência da doença (INCA, 2014b).

O aumento do número de casos tem sido associado a diversos fatores, que serão discutidos na próxima seção.

### 1.1.2. Fatores de risco

Existem vários fatores que aumentam a probabilidade do aparecimento de tumores malignos na mama. Esses fatores são conhecidos como fatores de risco, e ajudam a complementar a compreensão, não apenas sobre as patologias da mama, pois em geral estão associados também a outros tipos de cânceres. Dentre os fatores existentes alguns podem ser controlados, como o consumo de álcool, falta de atividade física, uso de tabaco e má alimentação. Ao passo que outros não, como faixa etária, idade e genética (OMS, 2013).

Pertencer a um ou mais fatores de risco não necessariamente indica que a paciente tenha câncer de mama, porém sugere que as chances de desenvolvê-lo são maiores (SILVA & RIUL, 2012). Portanto, o acompanhamento médico adequado se faz necessário para que o carcinoma seja detectado tão rápido quanto possível, visto que as chances de cura são muito maiores quando o tratamento é realizado em estágio inicial da doença do que em estágio avançado.

A mamografia é um dos exames capazes de detectar precocemente o câncer de mama, permitindo o diagnóstico quando o nódulo ainda não é perceptível ao toque do especialista (MINISTÉRIO DA SAÚDE, 2004).

## 1.2. Mamografia

A mamografia é uma forma particular de radiografia para o exame da mama capaz de diagnosticar tecidos normais da mama, tumores ou microcalcificações da ordem de 0,1 *mm* de diâmetro (GOES; AMBRÓSIO, 2013). Uma das principais vantagens desse exame é o fato dele ser um método não invasivo, dependendo de pouco tempo para sua execução (HEYWANG-KOBRUNNER *et al.*, 1999).

Apesar de ser uma das melhores técnicas para a detecção do câncer de mama segundo Ball *et al.* (2004), é desejável que o exame tenha uma boa qualidade, apresentando bons níveis de resolução espacial e contraste, para evitar equívocos no diagnóstico (CARLSSON *et al.*, 1986).

Outra característica desse exame é a quantidade de radiação emitida durante a sua realização, podendo representar risco ao paciente, portanto, é recomendável pelos especialistas que o exame seja realizado após exames clínicos suspeitos, ou em mulheres que possuam mais de 40 anos (ACR STANDARDS, 2014) (AMERICAN CANCER SOCIETY, 2004).

### 1.2.1. Classificação BIRADS

Desenvolvido pelo Colégio Americano de Radiologia em conjunto com departamentos de saúde dos Estados Unidos, o índice BIRADS (*Breast Image Reporting and Data System*) é um sistema de classificação com o intuito de padronizar laudos referentes à análise de mamogramas (ANDREADIS *et al.*, 2014), sendo utilizado para classificação de lesões encontradas na mama.

O índice de classificação BIRADS se divide em seis possíveis resultados, variando de 1 a 6, prevendo uma realização de procedimento para cada categoria, INCA (2009). Quando um exame não oferecer recursos necessários ao estabelecimento de conclusões, o mesmo é classificado como BIRADS 0.

No Quadro 1 é apresentada as classificações BIRADS e os seus significados clínicos:

Quadro 1 - Classificações BIRADS e suas descrições

BIRADS	Descrição
0	Exame não conclusivo
1	Normal
2	Achados Benignos
3	Achados provavelmente benignos
4	Achados suspeitos de malignidade
5	Achados altamente suspeitos de malignidade
6	Mamografia cujo diagnóstico é de câncer de mama.

Fonte: (ANDREADIS; NIKITA; SPYROU, 2014)

### 1.2.2. Características da Imagem Mamográfica

A mamografia resulta em uma imagem bidimensional, onde cada pixel da imagem possui um valor de 0 a 1, em caso de representação com números de ponto flutuante, e de 0 a 255, em caso de representação por números inteiros. Cada valor numérico representa uma escala de cinza, onde o valor mais baixo simula a cor mais escura e o valor mais alto a cor mais clara. Essa estrutura corresponde exatamente ao conceito de matriz bidimensional (SANTOS, 2002).

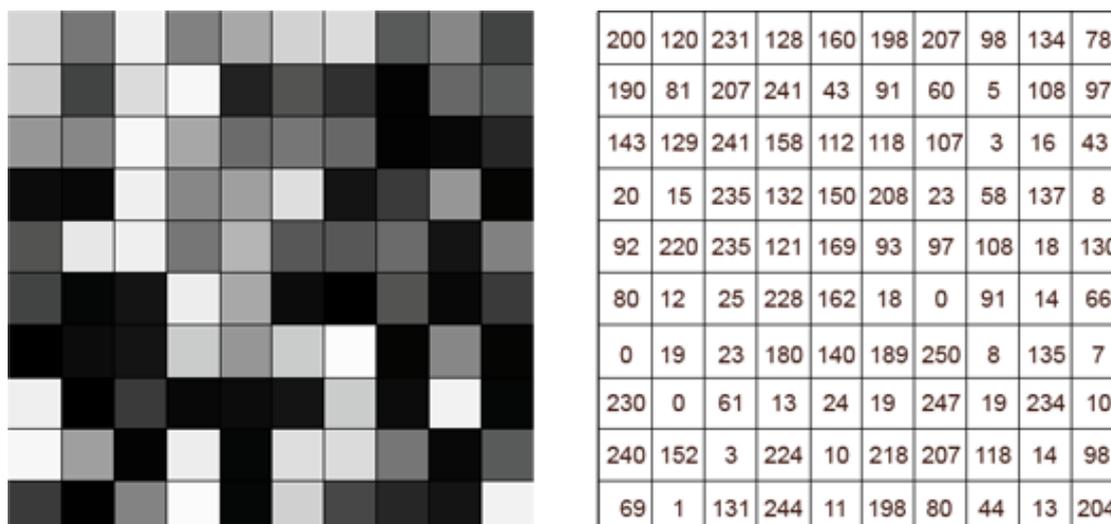


Figura 2 - a) Pixels em tons de cinza; b) Pixels em valores numéricos

Estruturas que possam indicar a presença de alguma anormalidade no exame geralmente apresentam diferença de contraste entre os tecidos envolvidos, apresentando frequências mais altas de tons de cinza em casos de nódulos ou microcalcificações.

### 1.2.3. Qualidade da Imagem do Mamograma

A qualidade de uma mamografia é definida pela capacidade do sistema reproduzir visualmente as características físicas das mamas e representá-las com a maior quantidade de detalhes pertinentes ao diagnóstico, não permitindo distorções e interferências que comprometam a análise. Além disso, é importante que a imagem resultante atenda a medidas satisfatórias de resolução, nitidez, contraste e nível de ruído (INCA, 2007).

Quando essas necessidades de qualidade do exame são atendidas é mais fácil fazer a identificação do câncer de mama pelo especialista. Em contraponto, exames de má qualidade podem causar interpretações equivocadas, gerando diagnósticos imprecisos, ou mesmo falso positivos e falsos negativos (MARTINEZ, 2007).

Uma das formas de tentar minimizar esses equívocos em análise de exames é por meio da utilização de ferramentas computacionais capazes de contribuir para com especialistas na realização do diagnóstico. Esses sistemas também são

conhecidos como sistemas CAD, sigla em inglês para Diagnóstico Auxiliado por Computador.

### 1.3. Fractais

Para analisar e classificar um exame mamográfico, o médico especialista tem de levar em consideração inúmeras particularidades da imagem, tais como a sua densidade, tamanho, presença da lesão nas duas mamas e forma dos contornos (CRISAN *et al.*, 2007).

Dentre as características das lesões uma das mais importantes para a categorização da mamografia é o contorno dos nódulos. Como pode ser observado na Figura 1, lesões com contornos regulares tendem a estar relacionadas a casos benignos, por outro lado contornos irregulares são associados a casos malignos (CRISAN *et al.*, 2007). Em vista disso, o grau de irregularidade do contorno é uma característica importante para ser quantificada, permitindo uma análise baseada também em um parâmetro quantitativo e não apenas pela percepção visual do médico. Conhecidamente através da literatura, uma das formas de mensurar o grau de irregularidade da forma de objetos complexos, é através do cálculo de sua dimensão fractal (SANKAR; THOMAS, 2009) (NGUYEN; RANGAYYAN, 2005).

Fractais são padrões complexos, infinitamente auto-similares em diferentes escalas e níveis de magnitude (MANDELBROT, 1977) (SILVA; SOUZA, 2010). O termo foi utilizado pela primeira vez pelo cientista Benoit Mandelbrot, na década de 70, após observar e interpretar a existência de um padrão de ruído nas transmissões de dados através de linhas telefônicas da IBM.

Não é simples identificar uma definição para o termo fractal, sendo muitas vezes relacionado a qualquer objeto complexo, ao qual as suas partes internas apresentam semelhança ao objeto como um todo, e que não possa ser representado corretamente pela geometria Euclidiana.

Em geral os fractais são padrões criados a partir de processos recursivos, apresentando auto-similaridade, complexidade e dimensão fracionada. Apesar da dificuldade em defini-los, o processo de criar fractais artificiais é relativamente simples. Na Figura 3 podemos observar fractais complexos, sendo criados a partir de processos recursivos simples:

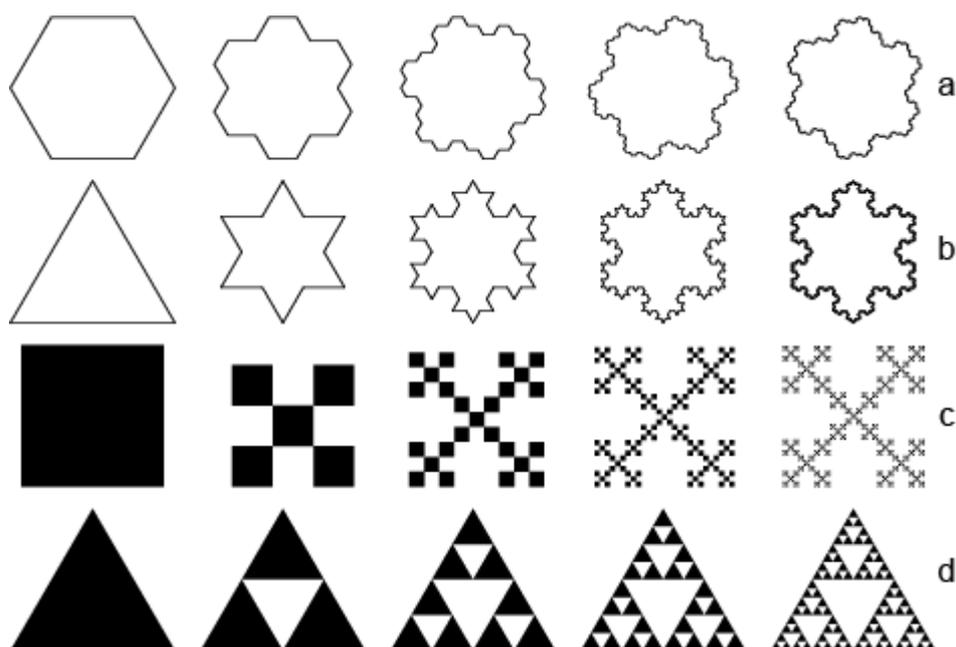


Figura 3 - Processo iterativo para criação de fractais artificiais; a) Ilha de Gosper; b) Floco de neve de Koch; c) Tapete de Sierpinski; d) Triângulo de Sierpinski.

Fonte: <http://mathworld.wolfram.com/Fractal.html>

Na natureza também podemos encontrar vários fenômenos que correspondem ao nível de complexidade dos fractais. Cadeia de montanhas, nuvens, vasos sanguíneos, flocos de neve, raízes de uma árvore, linhas costeiras, relâmpagos, penas de um pavão, neurônios, são todos exemplos de padrões naturais cuja complexidade, pode perfeitamente ser compreendida através do estudo dos fractais, ao invés das formas tradicionais propostas pela geometria Euclidiana (MANDELBROT, 1982).

Na figura 4 podemos observar alguns elementos da natureza que podem ser descritos como fractais:

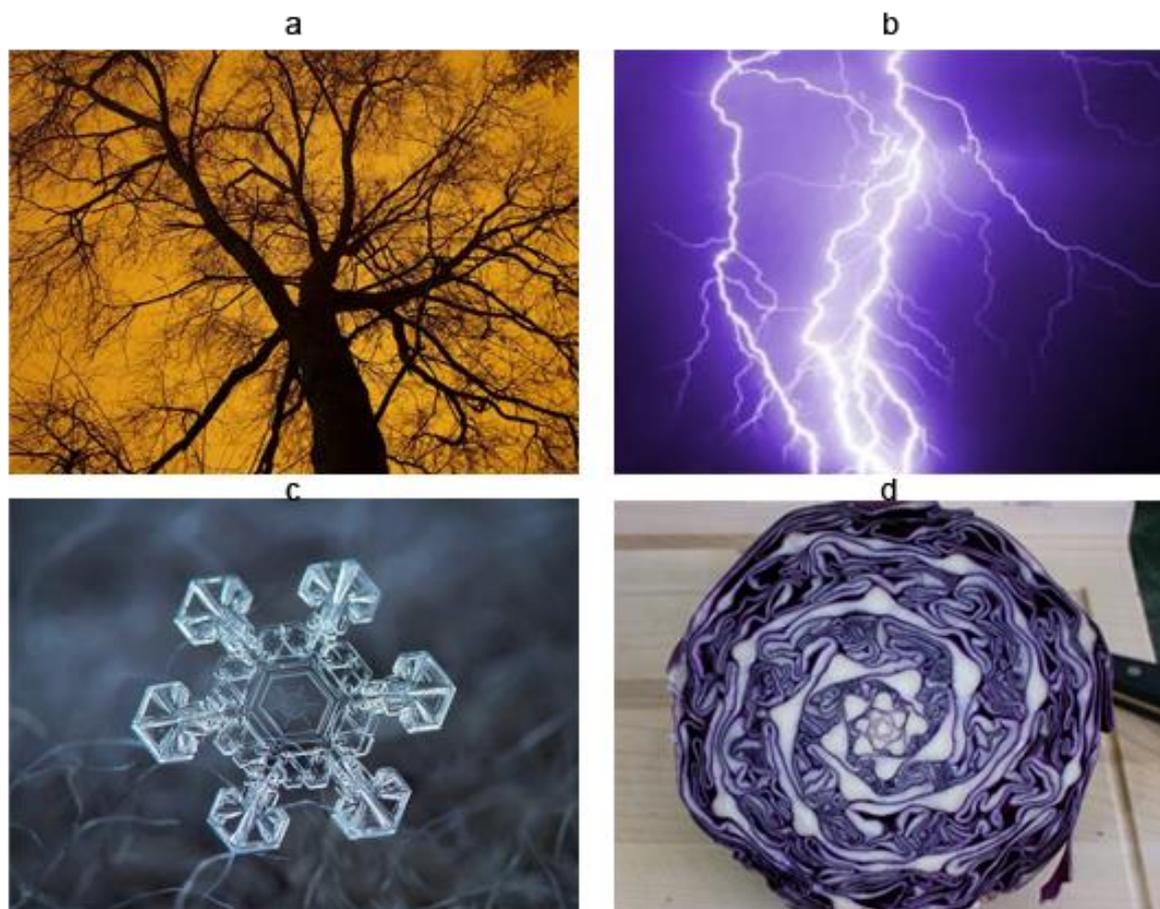


Figura 4 - Elementos naturais que apresentam padrão fractal; a) Galhos de uma árvore; b) Relâmpagos; c) Floco de neve; d) Repolho.

Fonte: <http://news.distractify.com/geek/science/the-most-fascinating-fractals-found-in-nature/>

### 1.3.1. Espaço Fractal

Uma das importantes propriedades dos fractais é que a sua quantificação depende da unidade de medida utilizada para sua mensuração. Quanto menor for a unidade de medida utilizada para sua determinação, maior será a precisão calculada (ANDRIOTTI, 2001). Desse modo, alguns objetos geométricos, como o floco de Koch ou o triângulo de Sierpinski, têm um nível de complexidade tão alto que as medidas clássicas de tamanho, área e volume não podem ser aplicadas.

Um bom exemplo desta complexidade é a curva de Koch (JAMPALA, 1992), obtida por meio de um processo iterativo, partindo-se de um seguimento de reta inicial de qualquer tamanho, dividi-lo em três partes iguais, substituindo a parte central por duas novas partes iguais ao comprimento da parte removida, formando

um ângulo de 60° graus entre elas. Esse processo é repetido nos quatro novos segmentos de retas obtidos na primeira iteração, e depois nos segmentos seguintes, com o número de iterações tendendo ao infinito.

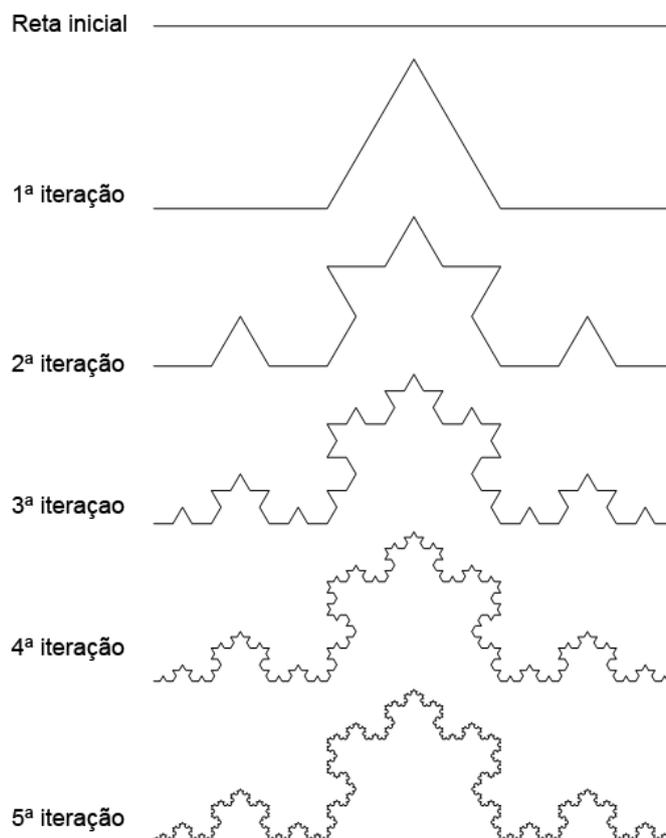


Figura 5 - Processo iterativo para gerar a curva de Koch em 5 iterações.

Fonte: <http://fractalfoundation.org/resources/fractivities/koch-curve/>

A cada iteração o comprimento da curva aumenta na razão de 4/3 em relação ao comprimento inicial da reta. Considerando  $C$  o comprimento da curva e  $n$  o número de iterações, podemos descrever o comprimento da reta através da equação:

$$C(n) = \left(\frac{4}{3}\right)^n \quad (1)$$

Dessa forma, à medida que o número de iterações  $n$  tende ao infinito o comprimento total da curva também tende ao infinito, porém ocupando uma porção definida e finita no plano. Isto posto, a curva não pode ser interpretada adequadamente pela dimensão topológica, pois se a entendermos como uma figura de uma dimensão, não poderíamos endereçar um ponto específico da curva, pois a mesma é infinita. Em contraponto, se a curva for interpretada como uma figura de duas dimensões não seria possível determinar a sua área, pois essa é nula. Então, faz-se necessária a utilização de um espaço ao qual a curva de Koch não seja infinita ou nula, este espaço, cuja dimensão não seja integral, é definido como dimensão fractal (SILVA; SOUZA, 2010).

### 1.3.2. Dimensão Fractal

Na geometria Euclidiana tradicional a dimensão está diretamente associada aos conceitos de comprimento, área e volume, de acordo com a natureza dos objetos (SILVA & SOUZA, 2010). Neste contexto, um ponto possui dimensão zero, a reta possui dimensão um, uma vez que apenas uma coordenada é necessária para endereçar cada ponto sobre a mesma, já o plano possui dimensão dois, pois são necessárias duas coordenadas para descrever um elemento deste conjunto e um cubo possui dimensão três, já que necessitamos de três coordenadas para endereçar um ponto nesse espaço.

Como visto na seção anterior, algumas formas não podem ser definidas pelos conceitos estabelecidos pela geometria clássica. Para encontrar a dimensão de objetos fractais, faz-se necessária a utilização da geometria fractal. É através dessa mudança de paradigma que se torna possível calcular a dimensão fractal de um objeto irregular, valor que representa o quão densamente um conjunto, em diferentes escalas, ocupa o espaço métrico em que encontra-se.

Em 1918, o matemático Felix Hausdorff apresentou uma das abordagens mais utilizadas sobre dimensão fractal de imagens. A dimensão de Hausdorff pode ser obtida dividindo uma reta em  $n$  partes de mesmo comprimento, em que cada segmento de reta possui o comprimento de  $1/n$ . Se aplicarmos esse conceito em um quadrado, e dividirmos os seus lados em  $n$  partes iguais, estaremos dividindo o quadrado como um todo em  $n^2$  quadrados, cada um com lado igual a  $1/n$ . Da

mesma forma, se aplicarmos o mesmo conceito a uma reta, quando dividirmos em  $1/n$  partes, teremos  $n^d$  sub retas (DODSON & KRISTENSEN, 2003). Deste modo, (SILVA & SOUZA, 2010) demonstram em seu estudo que a dimensão na geometria tradicional, corresponde ao valor do expoente de  $n$ ,  $d$ . Portanto, podemos definir a relação para o comprimento inicial da reta  $L$ :

$$N = \left(\frac{L}{n}\right)^d \quad (2)$$

Onde  $N$  será o valor do comprimento de reta na iteração  $n$ . Logo  $d$ , a dimensão de Hausdorff, pode ser definida aplicando o logaritmo em ambos os membros, ou seja:

$$d = \frac{\log N}{\log\left(\frac{L}{n}\right)} \quad (3)$$

No entanto, na maioria das vezes, o cálculo numérico se faz necessário, pois o método analítico da dimensão de Hausdorff para calcular dimensão fractal é aplicável em um número muito restrito de casos. Nessas situações, segundo (FARMER *et al.*, 1983) o método conhecido como *box-counting* pode ser utilizado pois permite obter resultados quase sempre iguais aos resultados obtidos pelo método de Hausdorff.

### 1.3.3. Box-Counting

Existem diversos métodos computacionais para calcular a dimensão fractal de uma imagem, como os apresentados nos estudo de (KLINKENBERG, 1994). Segundo (BALGHONAIM & KELLER, 1998) esses métodos podem ser classificados em três categorias: os métodos de variância, os métodos espectrais e os métodos de contagem de caixa ou *box-counting*.

Originalmente desenvolvido por (VOSS, 1986), devido à sua simplicidade e eficiência, o método de *box-counting* é um dos métodos mais utilizados em vários campos de aplicação para cálculo de dimensão fractal de imagens e sinais (PEITGEN *et al.*, 1992), fornecendo suporte para avaliação de padrões complexos.

O método consiste em criar uma grade de malha de tamanho  $U$  sobre a figura, fazendo a contagem do número  $N$  de caixas e a análise em escala de cada pedaço da figura. A cada etapa a malha é refinada, diminuindo o tamanho das caixas  $U$  e aumentando a quantidade  $N$  de caixas que cobrem a figura. Com base na relação de  $U$  (tamanho das caixas) e  $N$  (números de caixas sobre o contorno da imagem) se estabelece o gráfico logarítmico cuja inclinação determina o valor da dimensão fractal do objeto em questão (ADDISON, 1997) (SILVA; SOUZA, 2010).

Um exemplo da aplicação do método de *box-counting* a uma imagem do contorno do mapa do Brasil pode ser visto na Figura 6:

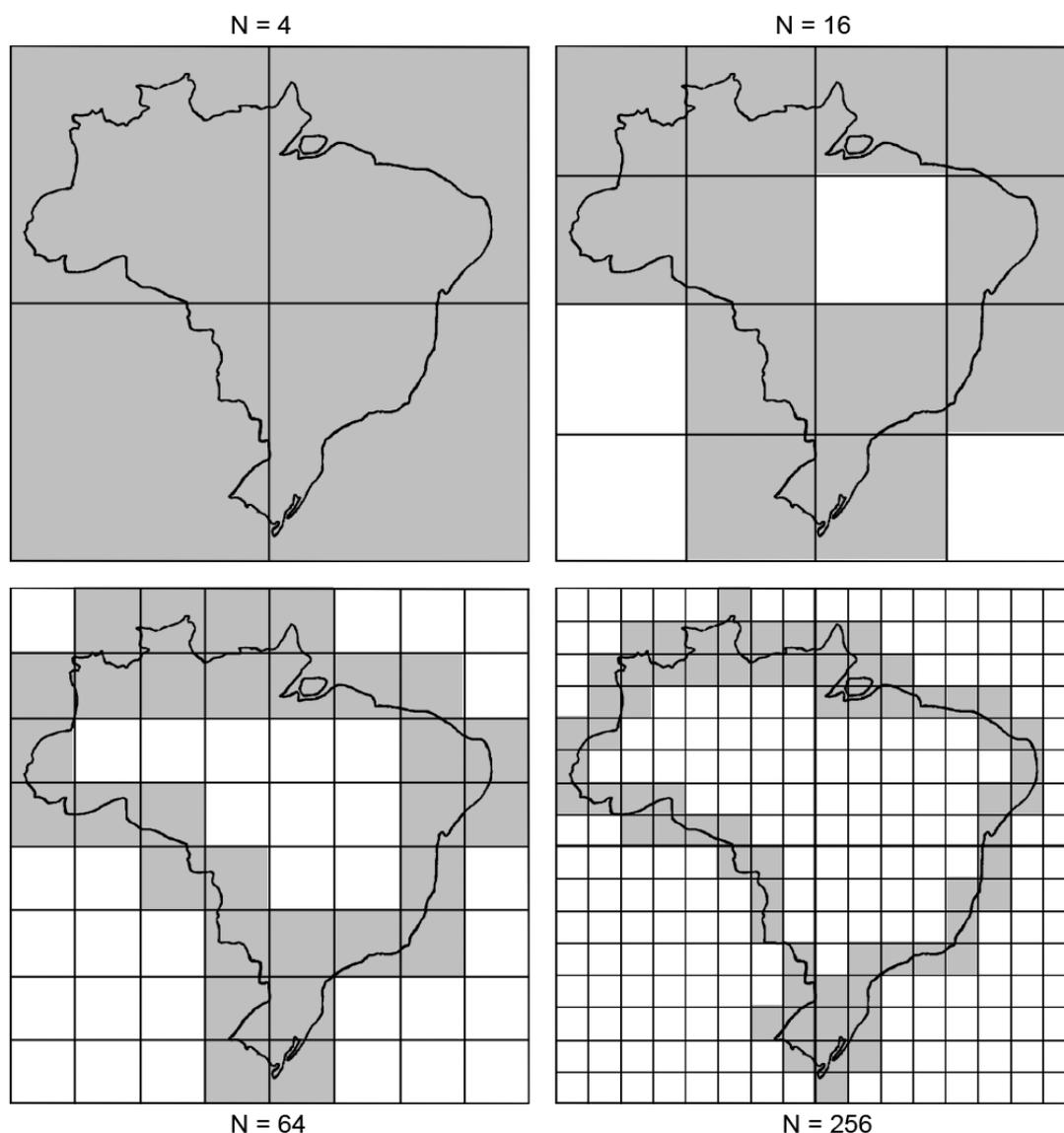


Figura 6 - Aplicação do método de box counting no contorno do mapa do Brasil.

Com base na contagem de  $N$  e  $U$  a cada etapa, podemos criar uma relação entre  $\log(N(U)) \times \log(1/U)$  e definir a dimensão fractal (FD) de *box counting* através do cálculo do coeficiente da reta de iterações demonstrada na Equação 4:

$$FD = \lim_{n \rightarrow \infty} \frac{(\log(N_{n+1}(U)) - \log(N_n(U)))}{\log(\frac{1}{U_{n+1}}) - \log(\frac{1}{U_n})} \quad (4)$$

Onde  $\log(N_{n+1}(U)) - \log(N_n(U))$  é a diferença entre a quantidade de caixas necessárias para cobrir todos os *pixels* do objeto no instante  $n+1$  e no instante  $n$ . Já  $\log(1/(U_{n+1})) - \log(1/U_n)$  é a diferença entre os tamanhos das caixas que cobrem o objeto no instante  $n+1$  e no instante  $n$  (SILVA & SOUZA, 2010).

Segundo (BACKES & MARTINEZ, 2005), o método de *box counting* requer alguns cuidados que devem ser observados para que ofereça um resultado mais preciso para qualquer tipo de forma estudada. Uma das restrições diz respeito ao alinhamento da grade utilizada na estrutura, que deve ser preenchida ao máximo, na tentativa de diminuir a quantidade de caixas desnecessárias do modelo. Além disso, é necessário que durante a próxima etapa, os tamanhos das caixas sejam metade do tamanho da caixa utilizada na etapa anterior.

#### 1.4. Apoio a Diagnóstico por Computador

Sistemas *CADx (Computer-Aided Diagnosis)* têm como objetivo possibilitar ao especialista uma interpretação e análise mais precisos a respeito da classificação e diagnóstico de exames. Esses sistemas trabalham com imagens médicas que nem sempre estão totalmente prontas para análise, necessitando muitas vezes de processamentos digitais para facilitarem sua interpretação (GIGER, 2004).

Esse processamento não tem como finalidade básica acrescentar informações à imagem, porém pode tornar as informações pertinentes ao diagnóstico médico mais acessíveis, fornecendo outros subsídios para a realização do diagnóstico.

Esta área vem crescendo bastante ao longo dos anos, fruto do avanço na produção de dispositivos eletrônicos cada vez mais potentes e com custos menores, possibilitando a investigação nesse campo através de modelos matemáticos e computacionais (ALMEIDA, 1998).

As técnicas de processamento de imagens podem ser efetuadas tanto no domínio da frequência, baseando-se essencialmente em modificações pela transformada de Fourier na imagem, como no domínio espacial, que se caracteriza pelo manuseamento direto dos pixels da imagem.

Esse processamento pode ser dividido em três sub categorias: baixo nível, caracterizada pelas operações de redução de ruído, realce de contraste e aumento de definição; nível intermediário, basicamente envolvendo etapas de segmentação e detecção de bordas; e alto nível, abarcando processos como “a interpretações ou compreensão da imagem, o reconhecimento de padrões, a análise de imagens e a visão assistida” (LIMA, 2009).

Um sistema de processamento de imagens envolve várias etapas. O primeiro passo é a obtenção da imagem através de um dispositivo de digitalização, por exemplo, um mamógrafo. Depois da obtenção da imagem, esta precisa ser representada adequadamente por exemplo através de uma matriz bidimensional. A etapa seguinte é o pré-processamento. Nesta altura, a imagem passa por um processo de filtragem em que o ruído, caso exista, deve ser minimizado, de modo que seja possível discriminar os artefatos presentes na imagem. Nesta fase, obtém-

se uma melhoria no contraste, o que torna mais fácil executar as tarefas de visualização.

Com a imagem filtrada, agora podemos obter informações sobre regiões específicas. Este passo é conhecido como segmentação e consiste em dividir a imagem em diferentes áreas de pixels, tais como objetos de forma independente uma da outra. Na literatura existem diferentes técnicas de segmentação, baseadas em princípios diferentes, como limiarização com base no histograma, detecção de borda, crescimento de regiões, entre outras (GONZALEZ; WOODS, 2011). Esta é uma das etapas primordiais do processamento de imagem e será detalhada na seção 1.4.3.

A etapa seguinte consiste em extrair atributos que permitam analisar os achados na imagem, tais como tamanho, posicionamento, estrutura, formato de borda e textura. Uma vez identificados, os achados devem ser classificados conforme sua semelhança, e finalmente é realizada a etapa de interpretação dos dados obtidos, com base em parâmetros quantitativos.

Uma descrição mais aprofundada destas etapas é mostrada nas seções seguintes.

#### 1.4.1. Aquisição das Imagens

O processo de aquisição de imagens é uma etapa muito importante em qualquer sistema de apoio ao diagnóstico, já que é através dele que a imagem é representada numericamente de forma adequada ao processamento digital. Nessa etapa são definidos fatores como a taxa de resolução e o contraste, que variam de acordo com os equipamentos utilizados no processo.

#### 1.4.2. Pré-processamento

Uma etapa de pré-processamento refere-se ao processo primitivo dos dados brutos da imagem. Esta etapa é realizada antes da segmentação, pois é necessário melhorar ou realçar as imagens originais para obter resultados mais precisos. Para tanto são geralmente realizadas operações que ajustem o contraste, normalizem a gama de valores de intensidade, realcem detalhes e removam artefatos irrelevantes.

Para isso, faz-se uso de operadores de intensidade, operadores locais e operadores morfológicos.

Dos operadores de intensidade fazem parte os operadores de ajuste de contraste, equalização e a expansão do histograma. Já entre as operações locais estão os filtros de redução de ruído da imagem e os filtros de identificação de bordas. Por fim, nas operações morfológicas encontram-se as técnicas de erosão e dilatação e suas combinações, como o filtro de abertura.

#### 1.4.2.1. Operadores de intensidade

Em processamento de imagens existem técnicas de transformação, cujo objetivo é a remoção de aspectos indesejáveis, suavização ou acentuação de aspectos específicos da imagem. Essas operações são realizadas ponto a ponto, uma vez que os níveis de cinza resultantes dependem apenas de seu histograma inicial (GONZALEZ; WOODS, 2011).

A modificação dos níveis de contraste de uma imagem consiste numa transferência de distribuição de intensidade de cinza em cada *pixel*, com a finalidade de aumentar a especificação visual entre os elementos presentes na imagem.

Como o histograma está associado ao contraste da imagem, analisá-lo pode ser favorável para a escolha adequada do mapeamento. O histograma da imagem está diretamente relacionada à distribuição estatística dos níveis de cinza em termos do número de *pixels*.

Na Figura 7 pode ser observada a representação de um histograma de uma imagem em escala de cinza.

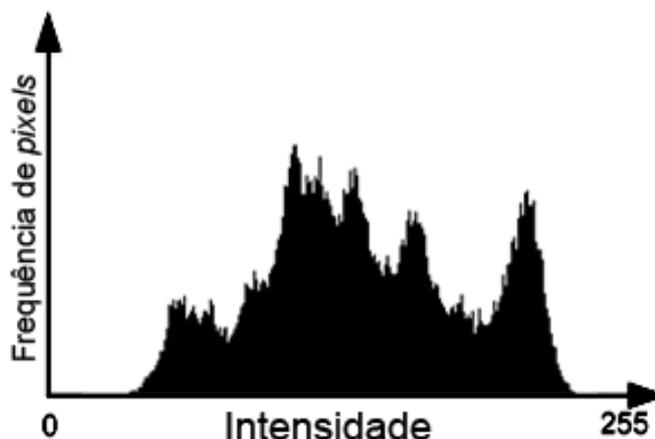


Figura 7 - Histograma de uma imagem em tons de cinza. O eixo x representa as 256 tonalidades de cinza e o eixo y representa a quantidade de pixels que tem essa tonalidade

Esta distribuição pode também ser dada em termos da porcentagem do número total de *pixels* na imagem. Pode ser estabelecida uma analogia entre o histograma de uma imagem e a função densidade de probabilidade, que é um modelo matemático da distribuição de tons de cinza de uma classe de imagens (SPRING, 2006).

A equalização do histograma pode favorecer o contraste da imagem que não apresenta distribuição uniforme dos níveis de cinza, possibilitando a percepção de pequenas variações de tonalidade e aumentando os detalhes visíveis. Já o ajuste de valores de intensidade aumenta o contraste da imagem, pois é possível mapear as alternâncias de tons de cinza do intervalo original para um intervalo apropriado, ampliando o espectro de níveis de cinza da imagem de origem.

Tomando uma determinada imagem representada como uma matriz  $ml \times mc$  ( $ml$  – linhas;  $mc$  – colunas) de intensidades de *pixel*, variando de 0 a  $L - 1$ , onde  $L$  é o número de possíveis valores de intensidade, nesse caso 256. Tomamos  $p$  como a probabilidade de um pixel da imagem possuir determinada intensidade de cinza da seguinte forma:

$$p_n = \frac{(\text{n}^\circ \text{ de pixels com intensidade } n)}{\text{total de pixels da imagem}}, n = 0, 1, \dots, L - 1 \quad (5)$$

Tanto a equalização do histograma, como o mapeamento de intervalo dos níveis de cinza, buscam obter uma melhor distribuição do histograma da imagem, para uma imagem mais uniforme, beneficiando a visualização e permitindo uma melhor análise.

#### 1.4.2.2. Operadores locais

Os operadores locais são conjuntos de técnicas de processamento baseadas na convolução de máscaras, orientadas à vizinhança em que o *pixel* se encontra. Nesse tipo de método o valor final obtido após o processamento depende do valor inicial do *pixel* e de dos *pixels* adjacentes a este (GONZALEZ; WOODS, 2011).

Entre esses métodos encontram-se o filtro de média, onde o valor de cinza do pixel é substituído pelo valor médio dos pixels vizinhos, o filtro Gaussiano, importante para remoção de ruídos com distribuição normal, o filtro Laplaciano, para detectar contornos de regiões com mudança rápida de intensidade.

#### 1.4.2.3. Operadores morfológicos

O princípio básico dos operadores morfológicos consiste em extrair as informações relativas à geometria e à topologia de uma imagem, pela transformação através de outro conjunto completamente definido, conhecido como o elemento estruturante, fundamental para as operações morfológicas básicas (SILVA *et al.*, 2014).

A aplicação dos operadores morfológicos é uma etapa importante para certas operações de processamento de imagem, tais como também o são os processos de realce de contraste, remoção de ruído e segmentação.

Dois operações morfológicas básicas são as operações de dilatação e a de erosão. A dilatação acrescenta *pixels* aos limites dos elementos da imagem, expandindo a sua área total, enquanto a erosão remove tais elementos, reduzindo-a. A quantidade de pixels adicionados ou removidos dos objetos depende do tamanho e da forma do elemento estruturante utilizado para processamento da imagem.

As operações de erosão e de dilatação frequentemente são utilizadas em conjunto. As duas combinações mais comuns destas duas operações originam os filtros morfológicos de abertura e de fechamento (GONZALEZ; WOODS, 2011).

O filtro morfológico de abertura em uma imagem consiste de uma erosão seguida de uma dilatação, usando o mesmo elemento estruturante para ambas as operações. Em contrapartida, a operação morfológica de fechamento de uma imagem é o contrário, ou seja, baseia-se numa dilatação seguida de uma erosão.

Aplicando a operação morfológica de abertura podemos remover pequenos objetos de uma imagem, como ruído ou elementos que não compõem os dados a serem analisados, preservando a forma e o tamanho dos objetos de maior tamanho. Enquanto a operação morfológica de fechamento suprime pequenas aberturas nos objetos da imagem, preenchendo falhas nos contornos e juntando elementos relativamente próximos.

#### 1.4.3. Segmentação

Em processamento de imagens, a segmentação é o processo de transformar uma imagem em uma imagem binária, separando a imagem digital em várias partes - conjuntos de *pixels*, podendo essa imagem pode ser colorida ou em escala de cinza. O objetivo desse processo é simplificar a representação de uma imagem em algo que é mais útil e fácil de analisar (ALMEIDA, 1998). A segmentação de imagens é normalmente usada para identificar áreas limitadas por linhas e curvas dentro da imagem.

No processo de segmentação os *pixels* são agrupados conforme certas características comuns, como por exemplo, a intensidade dos tons de cinza, para facilitar a identificação de aspectos específicos dentro da imagem. Esse trabalho é um dos mais importantes e mais complicados em análise de elementos na imagem.

Existem casos em que a segmentação pode ser resolvida usando técnicas elementares de segmentação. Nestas situações as imagens comumente apresentam objetos com um bom nível de contraste, em um *background* uniforme. Desse modo as técnicas de pré-processamento não seriam primordiais, já que as imagens se encontrariam em condições ideais. Diante disso, podemos utilizar a limiarização, para uma segmentação completa.

Em outros casos, a obtenção de uma segmentação eficaz, onde os objetos da imagem são separados exatamente de acordo com suas características, como identificação de nódulos em uma mamografia, por exemplo, requer um processamento de alto nível. Caso contrário, a segmentação obtida pode não atender às expectativas, identificando áreas que não correspondem ao objeto que se quer analisar, como a identificação de tecido mamário como parte de um nódulo. Nos casos em que as técnicas de limiarização não satisfazem as condições desejadas, alternativas como a técnica de crescimento de regiões podem ser utilizadas.

Os métodos de segmentação podem se fundamentar no conhecimento global das propriedades da imagem, baseando-se no seu histograma, em contornos determinados por mudanças bruscas de intensidade e em regiões, usando uma metodologia de subdivisão da imagem em regiões de acordo com um critério estabelecido previamente.

#### 1.4.3.1. Limiarização (*Thresholding*)

Um dos métodos mais utilizados para segmentação de imagens digitais é conhecido por método de *threshold*. Esta técnica se baseia no histograma da imagem em escala de cinza para realizar a escolha de um limiar apropriado à segmentação.

O princípio da segmentação incide em separar regiões de uma imagem quando esta apresenta duas classes, o *background* e o objeto. Devido ao fato da segmentação retornar uma imagem binária, preta e branca, o processo também pode ser denominado binarização.

Para realizar esse processo é necessário realizar a divisão do histograma, convertendo os *pixels* cujos tons de cinza são maiores ou iguais a um limiar escolhido em brancos e os demais em pretos, como ilustra a Figura 8:

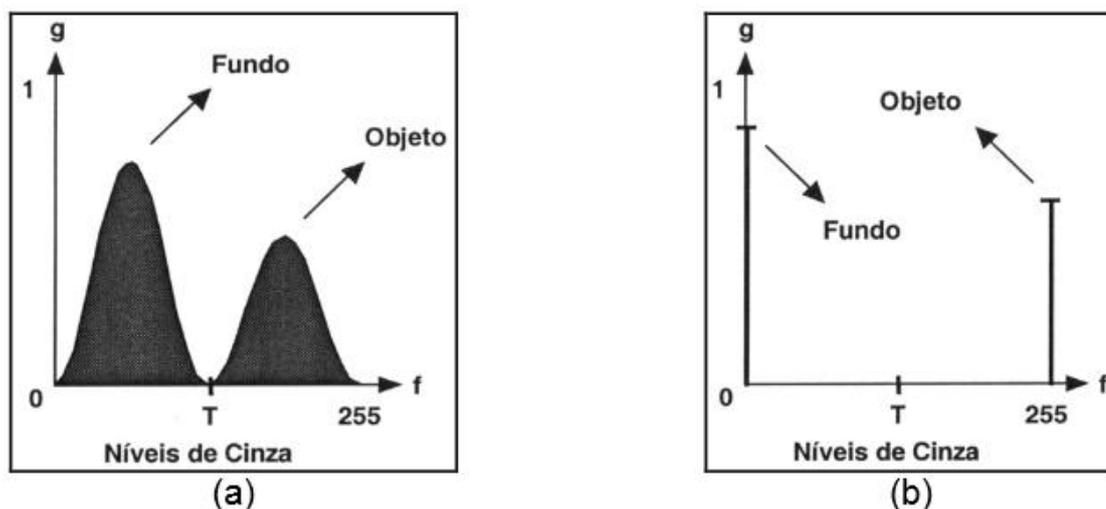


Figura 8 - Limiarização pelo método de *thresholding*. a) Histograma da imagem; b) histograma da imagem binarizada

Matematicamente, a binarização pode ser descrita como uma técnica de processamento de imagens na qual uma imagem de entrada  $f(x, y)$  de  $N$  níveis de cinza produz como saída uma imagem  $g(x, y)$ , chamada de imagem limiarizada, cujo número de níveis de cinza é menor que  $N$ . Frequentemente,  $g(x, y)$  apresenta 2 níveis de cinza, sendo:

$$g(x, y) = \begin{cases} 1 & \text{se } f(x, y) \geq T \\ 0 & \text{se } f(x, y) < T \end{cases} \quad (6)$$

Onde os pixels de cor 1 fazem parte da área do objeto e possuem cor branca, enquanto os pixels de cor 0 fazem parte do background e são pretos.

Quando uma imagem é adequada para o processo de segmentação, seu histograma exibe picos relativos à área do objeto, e picos relativos a seu *background*. Nesses casos, segundo a literatura, podemos definir o limiar entre os picos do histograma, pois os agrupamentos de tons de cinza, são bem acentuados, conforme pode ser observado na Figura 8.

Em casos práticos o histograma pode não ser tão uniforme quanto o esperado, tornando a tarefa de escolher o melhor limiar um passo não trivial, pois

este depende das condições iniciais da imagem original, como o contraste e a iluminação, e também do resultado que se pretende obter após a segmentação.

Depois que o limiar (*threshold*) é definido, todos os *pixels* com tom de cinza maior que limiar é classificado como pertencente ao elemento. Em contrapartida, os *pixels* com valores de cinza inferiores ao limiar são classificados em áreas não pertencentes ao elemento. O resultado desta intervenção é a binarização, ou seja, transformação da imagem original em uma imagem preta e branca.

#### 1.4.3.2. Crescimento de regiões

Outra forma de segmentação é conhecida como crescimento de regiões, onde em vez de dividir a imagem como no método de limiarização, o método começa a segmentação partir de uma semente, posição na imagem, agrupando os *pixels* vizinhos com intensidade de cinza semelhantes à sua, tentando expandir a região até que os pixels adjacentes sejam muito distintos em comparação com os da região para serem classificados como pertencentes à mesma (GONZALEZ; WOODS, 2011). Um teste estatístico é realizado para determinar se este é o caso.

O critério de similaridade é realizado com base em um teste de hipótese estatístico, que verifica a média entre a região e os pixels adjacentes. Em seguida, a imagem é dividida num certo número de sub-regiões e em seguida realiza a união entre eles, de acordo com o limiar de agregação pré-selecionado (MARTINS *et al.*, 2014).

O resultado obtido ao final desse processo de segmentação geralmente é o limite do objeto com seu *background* ou os pontos de sua área interna. Esta representação da imagem contém novas informações a respeito da forma dos elementos que não podiam ser vistas facilmente antes do processo e frequentemente representam melhor o objeto segmentado do que pelo método de *thresholding* (GONZALEZ; WOODS, 2011).

#### 1.4.3.3. Detecção de borda

O método de segmentação, conhecido por detecção de borda, é basicamente uma operação de identificação de mudanças locais expressivas nos níveis da

imagem. É esperado que essas regiões sejam suficientemente homogêneas para que a transição entre elas possa ser determinada baseando-se na descontinuidade de seus níveis de cinza.

Essas mudanças nas intensidades de cinza podem ser descritas através de derivadas parciais, já que uma imagem é formada de coordenadas espaciais, portanto, a ideia predominante das técnicas de detecção de borda é a realização de computação por meio de um operador diferencial.

Devido às características da própria imagem e a falhas no processo de sua aquisição, as bordas podem não ser facilmente identificáveis pelo processo. Quando isso acontece, a borda da imagem é descrita como uma rampa (modelo telhado), ou seja, uma variação gradativa dos tons de cinza entre os *pixels*, que torna mais difícil a delimitação entre as áreas da imagem.

Na Figura 9 pode ser observado um caso onde a mudança entre dois tons de cinza é brusca e um caso onde a mudança é suave.

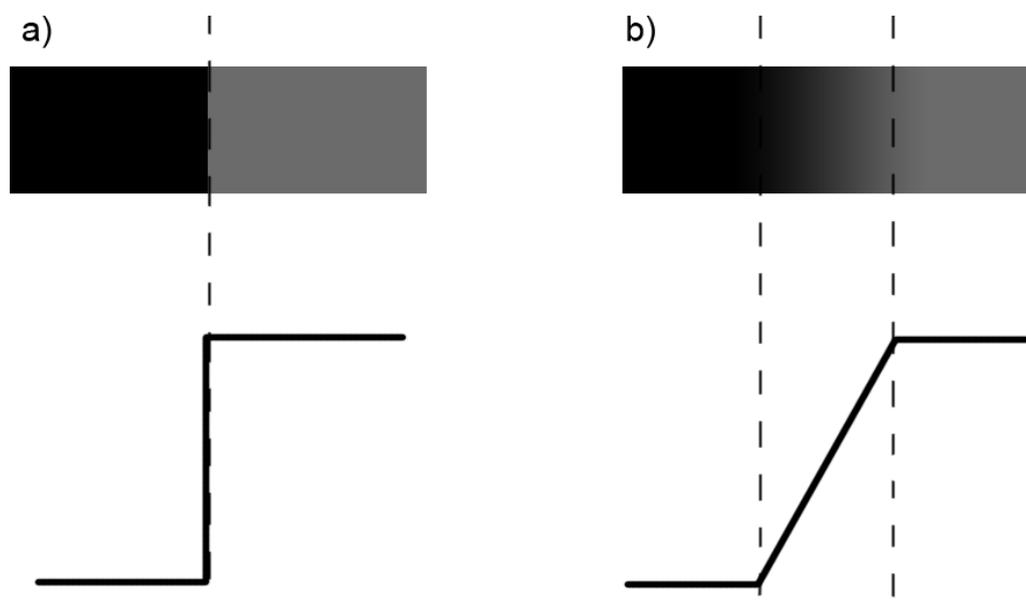


Figura 9 - Imagem com borda bem definida, caracterizada pela mudança brusca entre as tonalidades de cinza; b) Borda mal definida, caracterizada pela transição suave entre os níveis de cinza.

Nos casos em que a borda apresenta modelo de rampa é feita a utilização de gradiente, que nada mais é do que um vetor cuja direção indica o local no qual

existe maior variação nos níveis de cinza. Dessa forma é possível definir com certa precisão mesmo uma borda mal definida.

Por consequência, esse vetor é sempre perpendicular à tangente da borda, fornecendo duas medidas importantes para detecção da borda: a magnitude do vetor gradiente e a direção do vetor gradiente (GONZALEZ; WOODS, 2011).

Como a mudança de intensidade pode ser detectada pela diferença entre os valores de *pixels*, as bordas verticais podem ser identificadas pela diferença horizontal entre pontos, analogamente as bordas horizontais podem ser detectadas pela diferença vertical entre pontos adjacentes da imagem através do gradiente discreto.

A maioria das técnicas de detecção de borda empregam operadores diferenciais de primeira ordem, como os filtros de Roberts, Sobel e Prewitt (GONZALEZ; WOODS, 2011), e de segunda ordem, como o Laplaciano. Esses operadores geralmente fazem uso de algum tipo de suavização da imagem antes da aplicação do operador, para evitar o aparecimento de falsas bordas em decorrências de ruído (ALMEIDA, 1998).

Um dos métodos mais utilizados para detecção de borda é o método de Canny (CANNY, 1986), operador gaussiano de primeira derivada que suaviza os ruídos e localiza as bordas da imagem, baseando-se em três princípios: encontrar apenas bordas que realmente existem, diminuir a diferença entre a borda real e a borda encontrada e encontrar como resposta uma única borda para cada borda real da imagem.

#### 1.4.4. Classificação

Em processamento de imagens digitais a etapa de classificação consiste no emprego de métodos para rotular as imagens analisadas. Esta etapa é o desfecho do processamento de imagem, culminando no resultado esperado após todas as etapas anteriores. De forma geral, a classificação em sistemas CADx leva em considerações os aspectos relacionados aos nódulos ou microcalcificações para estipular a possibilidade de a lesão ser maligna ou benigna. Para tanto, com o passar dos anos, vários métodos computacionais tem sido empregados, com atenção especial às técnicas baseadas na utilização de lógica *fuzzy* (GAVRIELIDES

*et al.*, 2000), inteligência artificial e redes neurais (PAPADOPOULOS *et al.*, 2002) e transformada de *wavelet* (DIEKMANN *et al.*, 2004).

Para ser feita a classificação é necessário utilizar técnicas de apoio a decisão. Vários métodos são desenvolvidos para este fim, desde os que fazem uso de critérios simples, como as funções discriminantes e de cálculo de distância mínima, até o uso de técnicas de aprendizado automático e Inteligência Artificial, como clusterização, redes neurais, algoritmos genéticos e dimensão fractal. A descrição quantitativa destas informações através da extração de características estruturais complementa o sentido da classificação.

Esses e outros métodos são responsáveis pela concepção de modelos computacionais capazes de realizar a identificação de aspectos relacionados a grupos de diagnósticos esperados, como classificação de indivíduo saudável ou indivíduo com patologia. Posteriormente o modelo construído pode classificar novas imagens com base em suas características.

Dentre as características inicialmente mais utilizadas para classificação de lesões estão: calcificação, espicularidade, textura e forma, como descrito em (ACKERMAN; GOSE, 1972). Porém, com o passar do tempo outras características também foram levadas em consideração, permitindo diagnósticos mais precisos, como o sistema construído em redes neurais por (CHISTOYIANNI *et al.*, 2002), para classificação de áreas suspeitas com base na análise de componente independente em regiões de interesse.

O estudo foi realizado na base de dados MIAS, e consistiu em encontrar um conjunto de regiões independentes que geram a região de interesse. Essas regiões possuem características obtidas através dos coeficientes da transformação linear, cujo resultado são passados para o classificador do sistema. Tal sistema obteve eficiência de 88,23% para análise de anormalidades em mamografias.

(YARUSSO *et al.*; 2000) desenvolveram um método computacional para detecção de nódulos mamários, levando em consideração a relação simétrica entre as duas mamas das pacientes, onde quanto maior a diferença entre elas, maiores as chances de massas em potencial. Este método consiste no cálculo da diferença entre a mama esquerda e a mama direita, envolvendo técnica não-linear de subtração bilateral. No estudo realizado, fora, utilizados 154 pares de mamogramas, atingindo uma sensibilidade de 85% aproximadamente.

Um outro trabalho com resultados interessantes foi desenvolvido por (GROHMAN; DHAWAN, 2001), onde descrevem uma rede neural com elementos de lógica *fuzzy* para classificação de tipos de câncer de mama. Nesse trabalho foi introduzido um método para reconhecimento de padrões, que incorpora informações *fuzzy* para melhorar a performance da classificação. Ao todo foram utilizados 191 casos mamográficos, contendo microcalcificações, sendo 128 benignos e 63 malignos. A análise do histograma das imagens forneceu os parâmetros utilizados nas técnicas de *wavelets* que foram aplicadas em regiões contendo lesões. Segundo os autores, os resultados obtidos foram melhores que os resultados obtidos utilizando uma rede *backpropagation*.

#### 1.4.5. Avaliação do Sistema

Toda pesquisa baseada em sistemas *CADx* tem na avaliação de seu desempenho um aspecto de grande importância. Conhecer o nível de confiabilidade do sistema, e qual a sua capacidade de reconhecer corretamente casos positivos e casos negativos de uma doença estudada, é imprescindível para poder avaliar se o sistema proposto é viável ou não (MARTINEZ, 2007).

Essa avaliação pode ser feita através da divisão de dois grupos de médicos especialistas, onde o primeiro grupo realiza o diagnóstico com o sistema *CADx* desenvolvido e o segundo grupo realiza o teste clínico tradicional. Dessa forma é possível estimar se existe um aumento significativo na taxa de detecção correta do primeiro grupo em relação ao segundo, demonstrando a viabilidade de sua utilização. Esse método de avaliação pode ser bastante eficaz, porém experimentos dessa natureza podem estar associados a custos elevados, como descritos em (CHAN *et al.*, 1990), onde estudo semelhante foi realizado para detecção de microcalcificações em mamogramas.

Uma outra forma de avaliar o sistema é descobrindo sua sensibilidade e sua especificidade. O primeiro é a probabilidade de um teste ser positivo, dado que o indivíduo tenha a doença. O segundo é a probabilidade de um teste ser negativo, dado que o indivíduo não tenha a doença. Quanto maiores são as taxas de

sensibilidade e de especificidade de um sistema, maiores as possibilidades de diagnosticar corretamente um exame (PRATI; BATISTA; MONARD, 2008) (ABDELAAL et al., 2010b).

Para conhecer as taxas de sensibilidade e especificidade de um método, é necessário utilizar bases de imagens previamente diagnosticadas, desta forma é possível verificar se o método utilizado é capaz de fornecer um diagnóstico compatível com o laudo médico. Essa verificação se dá por meio da construção de uma matriz que relaciona os exames analisados clinicamente como positivos (indivíduos doentes) e negativos (indivíduos sadios) e os resultados obtidos através do sistema CAD. Essa matriz é conhecida como matriz de confusão (PRATI *et al.*, 2008) e seu conceito pode ser observado no Quadro 2:

Quadro 2 - Matriz de confusão

		Laudo Médico	
		<b>Positivo (Doente)</b>	<b>Negativo (Sadio)</b>
Resultados do Sistema CAD	<b>Positivo</b>	VP	FP
	<b>Negativo</b>	FN	VN
		TD	TS

Nessa matriz, as estruturas VP e VN correspondem respectivamente ao número de doentes cujo teste foi positivo (verdadeiros positivos), e ao número de sadios cujo teste foi negativo (verdadeiros negativos). Em contrapartida, FP e FN correspondem respectivamente ao número de sadios cujo teste foi positivo (falsos positivos), e ao número de doentes cujo teste foi negativo (falsos negativos), enquanto TD corresponde ao total de doentes e TS ao total de exames sadios analisados.

Com base na matriz de confusão, segundo Prati *et al.* (2008) podemos definir a taxa de sensibilidade como sendo o número de verdadeiro positivos VP, divididos pelo total de doentes TD:

$$Sensibilidade = \frac{VP}{TD} \quad (7)$$

Já a especificidade pode ser calculada dividindo o número de verdadeiro negativos VN pelo total de sadios TS:

$$Especificidade = \frac{VN}{TS} \quad (8)$$

Outra medida importante para validação do método é a taxa de falsos positivos TFP, inverso da especificidade, que corresponde à fração de todos os casos negativos que foram incorretamente classificados como positivos (COSTA, 2012). A TFP pode ser calculada da seguinte forma:

$$TFP = \frac{FP}{FP + VN} \quad (9)$$

Para determinar se um exame é de um paciente doente ou sadio é necessário analisar o valor da dimensão fractal do nódulo que possui e determinar um ponto de corte, onde toda dimensão fractal acima desse ponto será considerada pertencente a um nódulo maligno e toda dimensão fractal abaixo de ponto é relacionada a um caso benigno. Tendo isso em vista, a sensibilidade e a especificidade não são valores fixos para este sistema, pois variam de acordo com o ponto de corte escolhido como critério de classificação utilizado. Isso quer dizer que para cada teste de diagnóstico teremos quantidades diferentes de casos malignos e benignos, representados graficamente através de uma curva específica, chamada curva *ROC*.

#### 1.4.5.1. Curvas *ROC*

A análise de curvas *ROC* (*Receiver Operating Characteristic*), na teoria de detecção de sinal, é um método gráfico para avaliação da performance de vários sistemas, entre eles sistemas de apoio a diagnóstico (ZHOU *et al.*, 2008). A curva *ROC* ilustra a relação entre a taxa total de verdadeiros positivos em relação à taxa total de falso positivos. Na curva, a sensibilidade corresponde ao eixo y e a TFP ao eixo x.

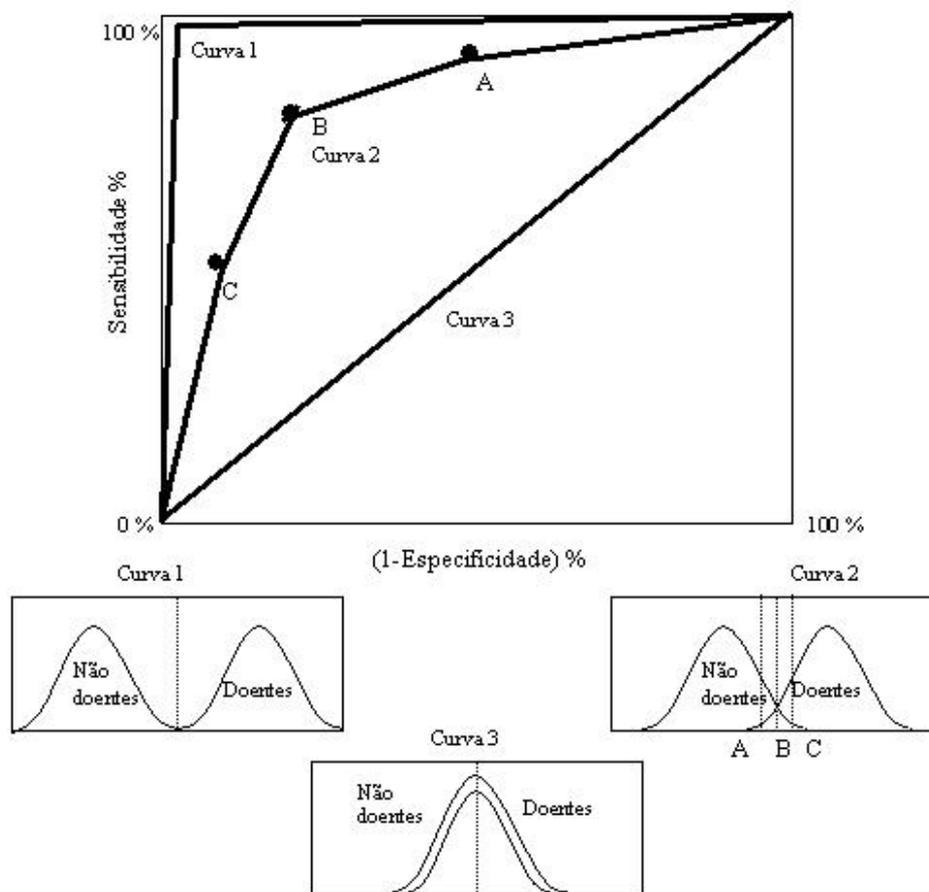


Figura 10 - Curvas ROC de três estudos diferentes  
 Fonte: Faculdade de Medicina da Universidade do Porto, [s.d.]

A Figura 10 apresenta as curvas *ROC* de três sistemas de apoio ao diagnóstico com o mesmo ponto de corte. A Curva 1 representa sistema com altíssima sensibilidade, onde todos VP são diagnosticados sem a presença de FP. A Curva 2 representa um sistema onde ocorrerá um pequeno grupo de FP e FN. A Curva 3 representa um diagnóstico aleatório.

## 2. MATERIAIS E MÉTODOS

Esta seção apresenta a descrição do conjunto de passos seguidos nesse trabalho para obtenção dos resultados. Estes passos se concentram em descrever desde a obtenção das imagens até a classificação das mamografias, passando pelas etapas de aplicação de técnicas de ajuste de contraste e de filtros morfológicos.

O *Matlab* (<http://www.mathworks.com/products/matlab/>) foi utilizado para realizar o processamento das imagens e para comparar os resultados obtidos, bem como implementar os algoritmos de classificação, pois o mesmo oferece um conjunto de funções específicas para esta finalidade.

### 2.1. Obtenção das imagens

As imagens utilizadas nesse trabalho foram extraídas de duas base de dados diferentes, MIAS e BancoWeb.

#### 2.1.1. Base MIAS

A base de dados MIAS (*Mammographic Image Analysis Society*) (SUCKLING et al., 1994) fornece um conjunto de 322 mamografias em formato PGM <sup>1</sup>(Portable Gray Map), com 8bits de contraste (256 níveis de cinza), todas com 1024x1024 pixels de dimensão e resolução espacial de 0,050mm.

Apesar de ser uma base de dados de 1994 e não receber atualizações recentemente, é uma base bem estabelecida e com boa aceitação para fins de pesquisa. As imagens da base MIAS foram utilizadas nesse trabalho pela facilidade de aquisição e processamento, já que todas as imagens possuem o mesmo tamanho, e por apresentarem uma série de características das mamas e dos nódulos, bem como a posição e o raio.

---

<sup>1</sup> O formato de imagens PGM é considerado um meio simples de salvar uma imagem representada em tons de cinza (POSKANZER, 2013).

No Quadro 3 pode ser observado todas as características das mamografias fornecidas pela base MIAS:

Quadro 3 – Conjunto de características a respeito dos exames de mamografias da base MIAS

<b>Características fornecidas da base MIA<sup>2</sup>S</b>	
Identificação do exame	Ex.: mbd012, mdb034;
Aspecto do tecido mamário	F – <i>Fatty</i> G – <i>Fatty-Glandular</i> D – <i>Dense-Glandular</i>
Classificação da anormalidade	CALC - <i>Calcification</i> CIRC – <i>Well-defined/circumscribed masses</i> SPIC – <i>Spiculated masses</i> MISC – <i>Other, ill-defined masses</i> ASYM - <i>Asymetry</i> NORM – <i>Normal</i>
Severidade da anormalidade	<i>Benign/Malignant</i>
Coordenada x do nódulo	1 – 1024px
Coordenada y do nódulo	1 – 1024px
Raio do círculo em que o nódulo está circunscrito	<i>r</i> (1 – 1024px)

Para a realização desse trabalho foram utilizadas ao todo 88 mamografias, sendo 50 de casos benignos e 38 de casos malignos. As demais imagens da base foram descartadas desse estudo por não apresentarem nenhum achado, casos em que a anormalidade foi classificada como *NORM*, ou apresentarem calcificações, classificação *CALC*. Ambos os casos não eram pertinentes para o estudo proposto nesse trabalho.

<sup>2</sup> Para consultar as informações da base MIAS, consultar o endereço: <http://peipa.essex.ac.uk/info/mias.html>

### 2.1.2. Base BancoWeb

BancoWeb (MATHEUS; SCHIABEL, 2011) é uma base de imagens livre para pesquisa, que contém até o presente momento 1703 mamografias cadastradas, sendo atualizada constantemente. As imagens dessa base estão, em sua maioria, em formato TIFF<sup>3</sup> (Tagged Image File Format), possuem 12 bits de contraste e resolução espacial de 0,075 ou 0,150 mm, dependendo da imagem selecionada, pois nem todas as imagens foram obtidas através do mesmo aparelho de mamografia. Possui ainda uma classificação BIRADS (AMERICAN COLLEGE OF RADIOLOGY, 2013), e algumas informações sobre o paciente.

Para esse estudo foram selecionadas aleatoriamente 53 imagens divididas em: 28 imagens BIRADS 2; 13 imagens BIRADS 3; 11 imagens BIRADS 4 e 1 imagem BIRADS 5. Vale ressaltar que não foram selecionadas mais imagens BIRADS 5 e nenhuma imagem BIRADS 6 devido ao fato de não ser possível encontra-las na base BancoWeb.

## 2.2. Método Automático para Classificação de Mamogramas

Esse método consiste em classificar as mamografias com base em sua dimensão fractal. Para isso foi desenvolvido um processo automático, onde só é necessário escolher o exame, sem necessitar de qualquer outra intervenção durante o processo. A técnica foi aplicada tanto para as imagens da base MIAS como para as imagens da base BancoWeb.

### 2.2.1. Identificação da área da mama

Para este processo foi necessário, em muitos casos, detectar a região da mama da paciente e excluir outras áreas não pertinentes ao computo, como tarja de identificação da mama, adesivos, área branca não pertencente a mama, etc.

---

<sup>3</sup> TIFF (acrônimo para Tagged Image File Format) é um formato de arquivo raster para imagens. Foi criado pela Aldus para uso no processo de impressão PostScript, e se transformou no formato padrão dos arquivos gráficos (32-bits) com elevada definição de cores.

Na Figura 11 pode ser observada uma mamografia da base MIAS, em que existe partes não necessárias ao processamento do nódulo, como o adesivo de identificação na parte superior direita.

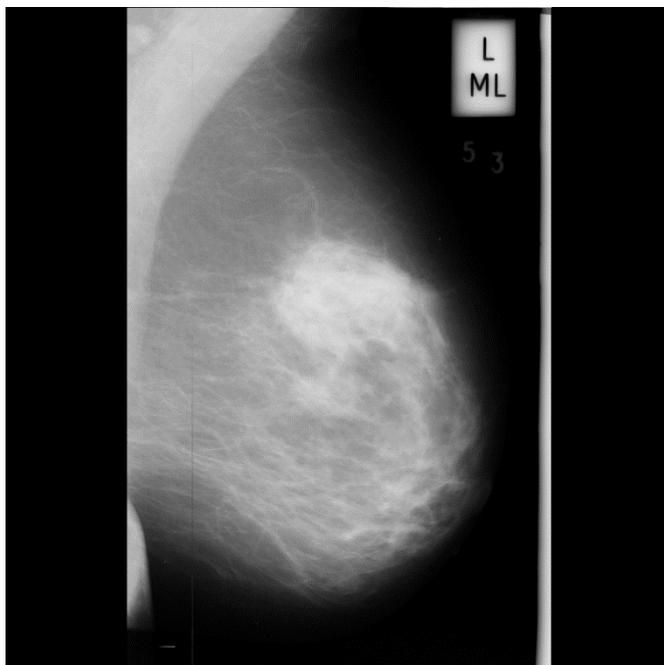


Figura 11 - Mamografia em que existe a necessidade de remover ruídos da imagem, como o identificador no canto superior direito.  
Fonte: Base MIAS

#### 2.2.1.1. Pré-processamento

No processamento digital de imagens, na maioria das vezes não é necessário trabalhar com a totalidade dos dados. Isso acontece quando as informações buscadas se encontram apenas em parte, ou em determinadas características da imagem original. Nesses casos, utilizar todas as informações nela contida pode levar a processamento desnecessário de dados ou mesmo à obtenção de resultados desacertados.

Para evitar esses percalços, foi realizada a etapa de pré-processamento nas imagens, antes de analisá-las de fato. Este processo possibilitou que as imagens resultantes fossem mais adequadas para a aplicação das etapas seguintes de processamento.

### 2.2.1.1.1. Ajuste do contraste

As mamografias utilizadas nesse trabalho apresentam variação entre si em relação aos níveis de tonalidade e intensidade de contraste. Com o intuito de minimizar as variações dessas características nas imagens trabalhadas e de tornar a aplicação do método mais confiável, foi realizado um ajuste de contraste nas imagens que não tinham o histograma bem distribuído. Nesse processo a imagem original foi mapeada com base em sua intensidade dos tons de cinza, conforme pode ser visto na Figura 12.

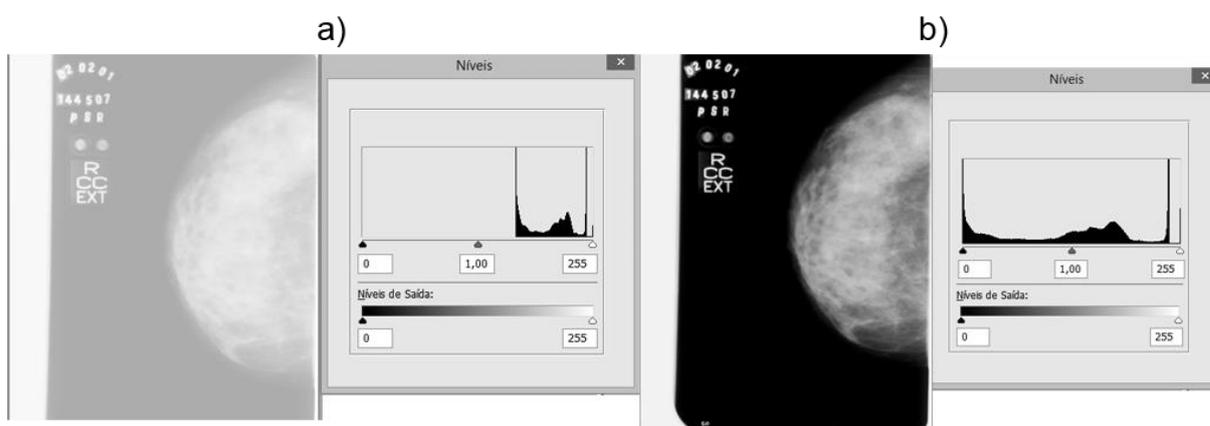


Figura 12 – a) Histograma original - A esquerda podemos observar a imagem e seu histograma com baixo nível de contraste; b) Histograma ajustada - A direita podemos ver a imagens com nível de contraste ajustado e seu histograma bem distribuído.

A mamografia original possui variação de tonalidade que vai de 0 (pixel mais escuro) até 255 (pixel mais claro), já as mamografias com baixo nível de contraste apresentavam valores muito próximos de 255, fazendo com que o histograma ficasse mais “estrito”. Os melhores resultados dessa etapa para as mamografias foram obtidos descartando os pixels com tonalidade abaixo de 241 e mapeando esses valores para as intensidades de 0 a 255, devido às características dos histogramas das imagens analisadas.

### 2.2.1.1.2. Primeira Limiarização – Destaque da Mama

Para identificarmos a mama o primeiro passo foi realizar a limiarização global. Para tanto será necessário escolher um valor de limiar que nos permita separar corretamente, ou para a maioria das imagens, a área da mama do fundo preto da imagem. Para esse estudo foi utilizado um valor de limiar de 0,22, escolhido através de testes de verificação com a base de dados MIAS. A Figura 13 apresenta o resultado da aplicação da binarização:



Figura 13 - Primeira limiarização realizada na imagem para destaque da região da mama. *Threshold* de 0,22.

### 2.2.1.1.3. Aplicação do filtro morfológico de abertura

Após a limiarização aplicou-se a próxima etapa. Essa etapa consiste em identificar e destacar a região da mama das demais regiões da imagem que não fazem parte da análise da dimensão fractal.

Na maioria das mamografias utilizadas como base desse estudo, as imagens possuíam marcações referentes à identificação do paciente. Para evitar processar

esses dados não relevantes ao trabalho, buscou-se uma forma de identificar apenas a área da mama.

Para a identificação da área de interesse foi aplicado um filtro morfológico de abertura na imagem após os procedimentos descritos em 2.2.1 e 2.2.2. Um dos principais entraves nessa etapa do trabalho foi a definição da dimensão e do formato a ser utilizado na aplicação do filtro morfológico de abertura, levando em consideração que as mamas possuíam tamanhos e formas bem diversificadas e a seleção deveria ser feita perdendo o mínimo de informação possível.

Tanto o formato como a dimensão do filtro morfológico de abertura utilizado nesse trabalho, foram definidos através de um conjunto de testes empíricos realizados em um grupo controle de mamografias. O formato de “disco” se mostrou mais adaptável à forma curva das mamas, portanto foi o formato padrão escolhido nessa etapa. O diâmetro escolhido após os testes foi 32, já que valores maiores acabavam desconsiderando informações relevantes da mama e valores menores consideravam informações não pertinentes ao estudo. Na Figura 14 pode ser observado o resultado após aplicação do filtro morfológico:



Figura 14 - Aplicação do filtro morfológico de abertura

#### 2.2.1.1.4. Seleção da região de interesse

A aplicação do filtro morfológico descrito na seção anterior identifica a área onde a mama se encontra no mamograma, descartando as regiões que não interessam. Em contrapartida, essa imagem não apresenta os detalhes dos nódulos da mama, sem os quais não se pode avaliar o exame.

O processo seguinte, descrito nessa seção, apresenta o método para a obtenção da mamografia, permitindo a visualização de todos os detalhes da área de interesse e a exclusão de ruídos externos, presentes na imagem original. Para isso, foi realizada a multiplicação ponto a ponto entre a imagem da mamografia original e a imagem obtida após a aplicação do filtro morfológico de abertura. O resultado deste processo é uma imagem em que a área da mama contém todas as suas características e o resto da imagem apresentam todos os pixels na cor preta, como pode ser observado na Figura 15.

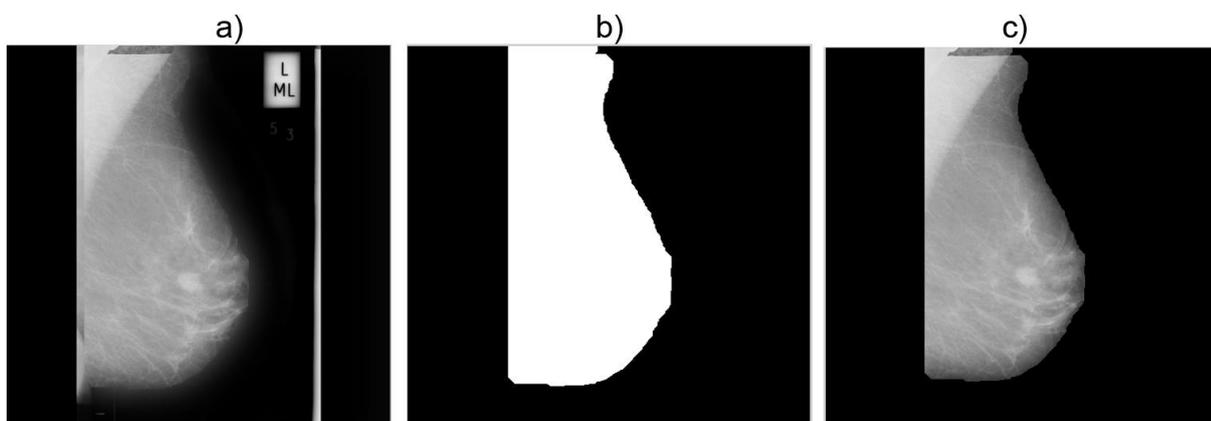


Figura 15 – a) imagem inicial; b) imagem após a passagem do filtro morfológico; c) imagem final. Multiplicação ponto a ponto entre a imagem inicial e a imagem filtrada, resultando na imagem final sem ruídos.

#### 2.2.2. Identificação dos nódulos

Uma vez identificada a região da mama parte-se para a identificação dos nódulos presentes nessa região.

### 2.2.2.1. *Equalização do histograma*

O histograma de uma imagem descreve graficamente a distribuição estatística de suas intensidades de tons de cinza. Se um nível de cinza  $x$  ocorre  $n$  vezes na imagem de  $N$  pixels, podemos estimar a probabilidade de ocorrência deste nível  $x$  na imagem através de sua respectiva frequência.

Nessa etapa do trabalho todas as mamografias foram equalizadas com base em seu respectivo histograma. Esse processo foi realizado para facilitar a diferenciação entre a região da mama e as outras regiões da imagem, complementando o processo de ajuste de luminosidade descrito na etapa anterior, podendo ser observado na Figura 16.

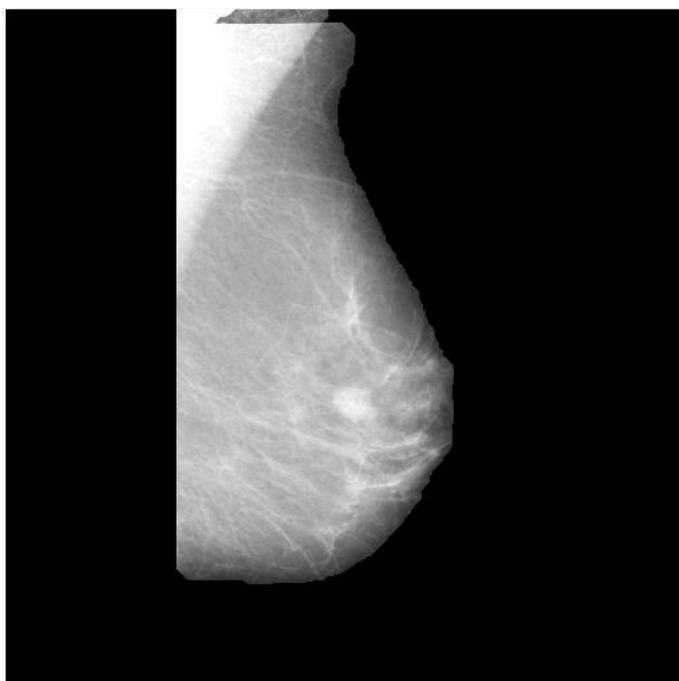


Figura 16 - Imagem equalizada

### 2.2.2.2. *Segunda Limiarização – Identificação dos Nódulos*

Nesta etapa fazemos a segunda limiarização da imagem, dessa vez para identificação dos nódulos presentes na mama. Para esta segunda limiarização, o procedimento utilizado foi o mesmo utilizado na seção 2.2.2. Portanto, foi aplicado

um limiar, obtido de forma empírica, resultando no valor 0,75, que permite realçar os nódulos do mamograma. Para obtenção desse valor de limiar, foram realizados testes com diferentes limiares em várias mamografias selecionada. O resultado desse processo pode ser observado na Figura 17:

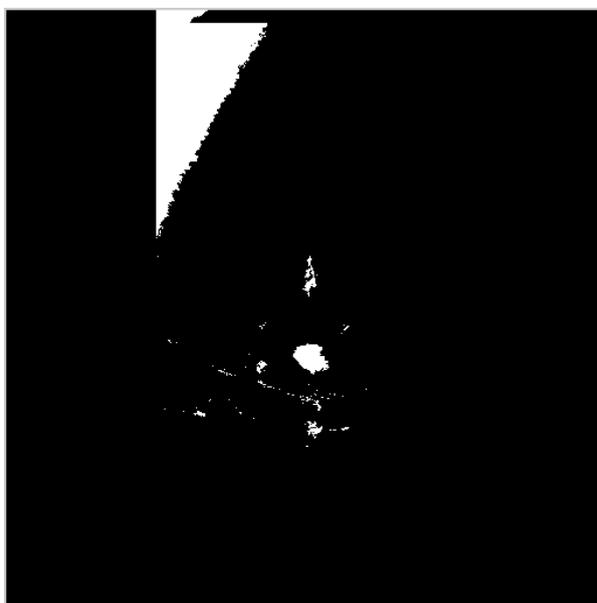


Figura 17 - Limiarização da mamografia para identificação do nódulo. Além do nódulo, parte do músculo peitoral (canto superior esquerdo) e parte do tecido glandular foram selecionados.

### 2.2.2.3. Excluindo Ruídos da Imagem

Um dos principais problemas desse método é a permanência de objetos na imagem que não fazem parte do nódulo. Na Figura 17, verificamos que além do nódulo, foram identificados também o músculo peitoral e pequenas parte de tecidos glandulares da mama.

Se aplicássemos agora o cálculo da dimensão fractal pelo método de *box-counting* seria possível que o valor encontrado diferisse do valor real de dimensão do nódulo, por conta do ruído ainda presente na imagem.

Buscando contornar essa situação, foram desenvolvidas algumas técnicas para minimizar a presença dos ruídos:

- *Remoção de áreas menores que 1256 pixels*: através da base de dados MIAS, conhecemos o menor raio de um nódulo dentre todas as imagens, que é de 20 *pixels*. Calculamos então a área que um círculo cujo raio seria igual a 20 *pixels*. O valor dessa área é de 1256 *pixels*. A imagem então é percorrida à procura de áreas menores que 1256 *pixels*. Com base nesse valor, percorremos a imagem em busca de áreas (regiões de *pixels* adjuntos de mesma intensidade) menores do que a área mínima calculada. Essas áreas são removidas da imagem. Essa etapa objetiva remover pequenos ruídos ainda presentes na imagem após a segmentação;

- *Remoção de áreas maiores que 121922 pixels*: analogamente, podemos utilizar o mesmo raciocínio para remover áreas grandes da imagem, pois é conhecido que o maior raio de um nódulo dentre todas as imagens é de 197 *pixels*. Calculamos então a área que um círculo cujo raio seria igual a 197 *pixels*, aproximadamente 121922 *pixels*. A imagem então é percorrida à procura de áreas maiores que 121922 *pixels*, caso encontre, essas áreas são removidas da imagem. Esta etapa objetiva remover o músculo peitoral da imagem.

Na figura 18 pode ser observada a imagem após a extração dos ruídos:

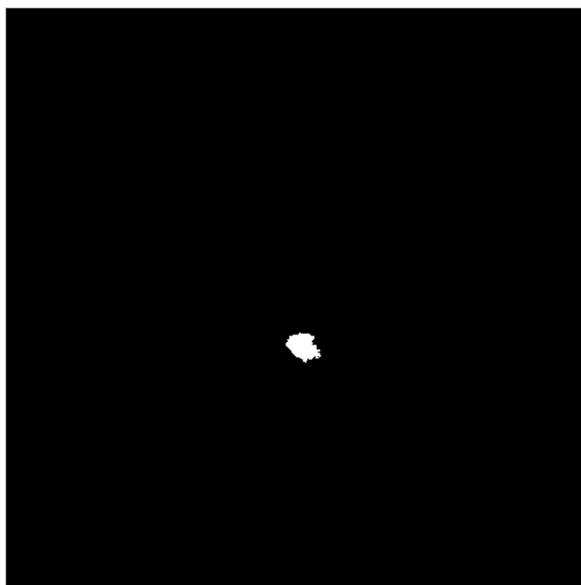


Figura 18 - Imagem após os métodos de remoção de pequenas e grandes áreas

## 2.3. Método Semi-Automático de Classificação de Mamogramas

Nessa seção é apresentado o conjunto de passos necessários para a realização do cálculo da dimensão fractal de um nódulo mamário, em imagens segmentadas por meio do método de crescimento de regiões.

Vale ressaltar que o método de crescimento de regiões foi aplicado apenas às imagens da base MIAS, pois a base de imagens BancoWeb, “apesar de oferecer ferramentas para recorte de imagens e uma base - em crescimento - de regiões de interesse, até esta data não apresenta regiões de interesse demarcadas para as imagens disponíveis.” (MATHEUS, 2010)

### 2.3.1. Pré-processamento

Segmentação por crescimento de regiões consiste em uma técnica de segmentação de imagens, que dado um *pixel* inicial na imagem, busca encontrar outros *pixels* adjacentes ao primeiro, levando em consideração suas tonalidades de cinza.

Neste trabalho a técnica de segmentação de regiões foi utilizada a partir da aplicação desenvolvida em (MARTINS *et al.*, 2014). Para isso foi necessário realizar pré-processamento nas imagens, submetendo-as a uma conversão de tipos, para que seus dados pudessem ser lidos corretamente pela aplicação utilizada. O grupo de imagens da base MIAS foi convertido do tipo *PGM* para o tipo *PNG*, sem perda de informação durante o processo. Essa conversão foi necessária, pois a aplicação supracitada não trabalha com arquivos de imagem do tipo *PGM*.

Outra característica da base de dados MIAS que pode levar a resultados não esperados em algumas análises, é o fato do ponto de origem das imagens começarem a partir do ponto inferior esquerdo, enquanto a maioria dos sistemas propostos trabalha com imagens cuja origem se dá a partir do canto superior esquerdo. Desse modo, as imagens foram espelhadas em função do eixo vertical, para que a origem começasse do ponto superior esquerdo, tal qual a ferramenta testada, como pode ser observada na Figura 19.

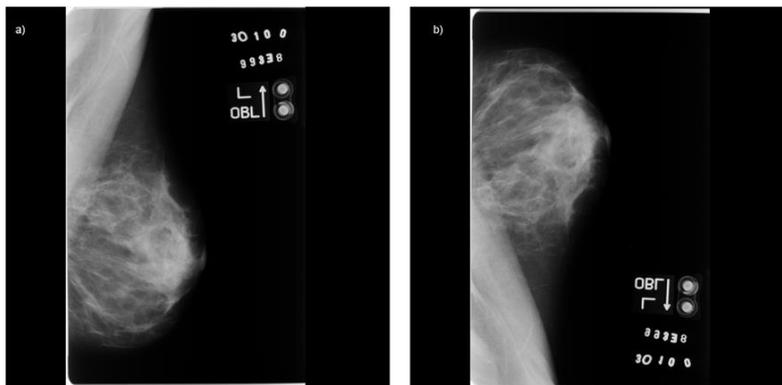


Figura 19- a) Imagem original, em formato PGM; b) Imagem pré-processada, em formato PNG e com eixo y invertido.

### 2.3.2. Escolha da semente

Para identificação da região de interesse pelo método de crescimento de regiões é necessário definir um ponto inicial, a partir do qual a área será definida, esse ponto inicial é conhecido como semente.

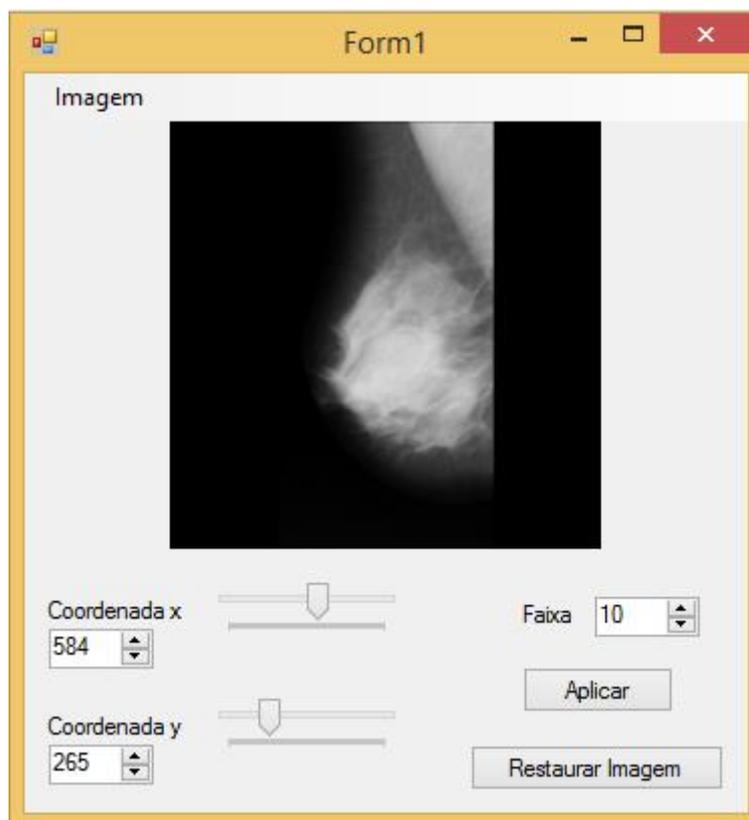


Figura 20 - Escolha das coordenadas x e y que determinam a semente para a segmentação por

crescimento de regiões. Para essa imagem a semente está localizada na posição 582x265.

Conforme pode ser visto na Figura 20, a identificação da semente é feita através do mapeamento das duas coordenadas espaciais da mamografia, as coordenadas  $x$  e  $y$ . Conforme descrito em (SUCKLING et al., 1994), as coordenadas do sistema se originavam no canto inferior esquerdo da imagem, mas foram alteradas para o canto superior esquerdo como visto na seção anterior.

Além disso, todas as imagens dessa base que apresentam a existência de algum achado, seja benigno ou maligno, possuem as coordenadas da semente conhecidas, facilitando o mapeamento através da identificação do ponto médio do nódulo.

Em seguida, foi necessário definir uma verificação dos tons de *pixels* adjacentes em relação à semente. Segundo (MARTINS et al., 2014) essa propriedade é importante para determinar se um *pixel* pertence ou não à uma região, pois leva em consideração o grau de semelhança entre as tonalidades dos *pixels*. Para este estudo foi escolhido empiricamente o valor da intensidade 36 do tom de cinza.

### 2.3.3. Segmentação das mamografias

Com base no valor da faixa de tons de cinza, escolhido na etapa anterior, é necessário realizar a etapa seguinte, destacando o nódulo presente na mamografia. Para isso, é preciso pressionar o botão “Aplicar” para que a região de interesse seja segmentada.

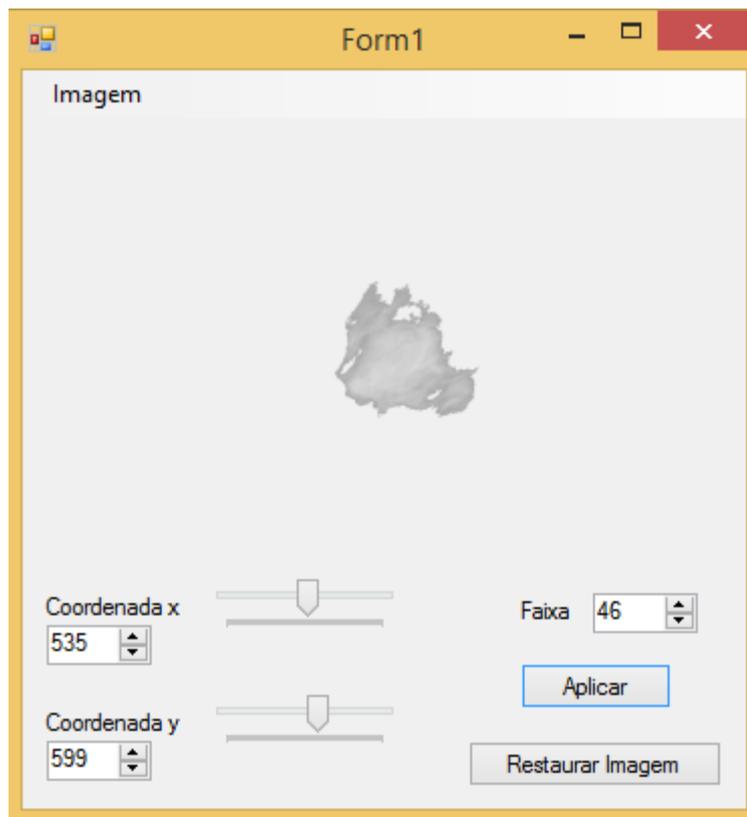


Figura 21 - Segmentação do nódulo por crescimento de regiões

Em seguida essa imagem foi salva em formato PNG, no diretório de imagens processadas, para aplicação do processo de segmentação pelo método de detecção de bordas.

## 2.4. Filtro de Detecção de Bordas

Esta etapa foi efetuada tanto para o método de auxílio à detecção de nódulos como para o método de apoio ao diagnóstico utilizando a segmentação por crescimento de regiões, e consiste em detectar as bordas dos nódulos indicados após o processo de segmentação. A borda deve ser identificada, para realização dos cálculos de dimensão fractal.

Foram feitos experimentos com os filtros detectores de bordas de Roberts, Sobel, Prewitt e Canny. Para esse estudo o que apresentou os melhores resultados foi o detector de Canny.

O método de Canny recebe uma imagem em tons de cinza ou binária, e retorna uma imagem de mesmo tamanho da original, inserindo “1” onde é considerado borda e “0” para o restante da imagem. O resultado da passagem pelo método de Canny é demonstrado na Figura 22.

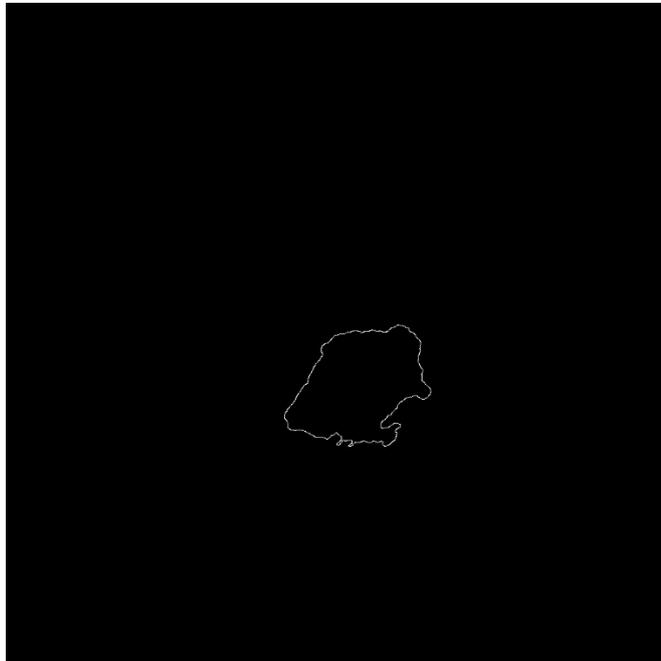


Figura 22 - Detecção de bordas do nódulo utilizando o método de Canny

## 2.5. Aplicação do *Box-Counting*

Como visto na seção 3, a dimensão fractal quantifica a complexidade de um objeto, com base no seu grau de irregularidade, o que é bastante útil para a caracterização de lesões em mamografia. Portanto, a aplicação dessa técnica através do método de *box-counting* foi empregada nesse estudo como forma de gerar parâmetros para a classificação dos nódulos mamários analisados.

Após a aplicação do método, foi então gerado o gráfico da relação de  $\text{Log}(N(U)) \times \text{Log}(1/U)$ , conforme mencionado na seção 1.3.3, para todas as mamografias, e através do coeficiente angular desse gráfico obtemos a dimensão fractal.

### 3. RESULTADOS E DISCUSSÕES

Nesta seção são apresentados os resultados obtidos no trabalho, e as discussões dos principais aspectos observados em todo o processo de extração de características e de classificação de nódulos mamários por meio do cálculo de se sua dimensão fractal.

Os resultados foram obtidos através da análise de 142 mamografias, sendo 88 da base de dados MIAS e 54 da base de dados BancoWeb. Na Tabela 1 pode ser observado o conjunto de mamografias com nódulos diagnosticados, as imagens que do tipo *CALC* foram descartadas por não fazerem parte do escopo desse estudo. Já na Tabela 2 pode ser observada a distribuição das mamografias benignas e malignas, de acordo com suas características.

Tabela 1 - Relação de mamografias analisadas, separadas por suas características.

	<i>CIRC</i>	<i>MISC</i>	<i>ASYM</i>	<i>ARCH</i>	<i>SPIC</i>	<i>CALC</i>
Benignos	19	7	6	9	11	13
Malignos	4	8	9	10	8	15

Tabela 2 - Percentagem de benignas e malignas separadas por suas características

	<i>CIRC</i>	<i>MISC</i>	<i>ASYM</i>	<i>ARCH</i>	<i>SPIC</i>
Benignos	36.5%	13.5%	11.5%	17.3%	21.2%
Malignos	10.3%	20.5%	23.1%	25.6%	20.5%

Na Tabela 3 é apresentada a distribuição das 54 imagens da base BancoWeb analisadas nesse estudo, separadas de acordo com a sua classificação BIRADS:

Tabela 3 - Distribuições das mamografias da base BancoWeb com base em sua classificação BIRADS

BIRADS	Quantidade
2	29
3	13
4	11
5	1

As imagens com classificação BIRADS 1 e 6 não foram utilizadas, pois não apresentam existência de nódulos mamários e por não existirem na base BancoWeb, respectivamente.

As 142 imagens foram segmentadas utilizando o método de limiarização por *thresholding*. Conforme apresentado no Capítulo 2, duas etapas de segmentação foram realizadas para cada imagem, a primeira para o destaque da mama e a segunda para destaque dos nódulos presentes no exame.

Para verificar a eficiência do método proposto foram considerados os casos em que os resultados obtidos foram totalmente compatíveis aos diagnósticos médicos, casos em que o resultado foi parcialmente compatível com o diagnóstico médico, ou seja, outras áreas além do nódulo foram identificadas, porém sem perda de informações na área da lesão, e por sua vez os casos em que os resultados diferiram totalmente do diagnóstico médico.

A Tabela 4 a seguir apresenta as três taxas analisadas para as imagens com nódulo benigno da base MIAS:

Tabela 4 - Resultado da segmentação pelo método de *threshold* para nódulos benignos

Nódulos benignos		
Resultado	Qtd de mamografias	Porcentagem
Erro	26	51%
Acerto parcial	18	35%
Acerto	7	14%

Do total de imagens analisadas, verificou-se que 51% dos testes utilizando segmentação por *threshold* não identificaram corretamente os nódulos, caracterizando erro de análise. Foi considerado erro todo o teste que não identificou nenhum nódulo na imagem ou que identificou como nódulo outra área da mama.

Na Figura 23 pode ser observado um caso em que não foi possível identificar o nódulo após o processo.

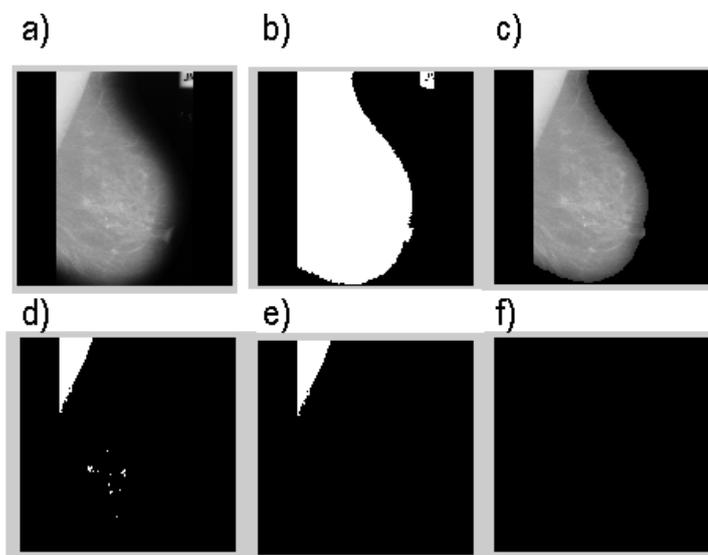


Figura 23- Mamografia mdb204 - Exemplo de teste em que não foi reconhecido nenhum nódulo na mama. a) Mamografia original; b) Primeira segmentação; c) Eliminação de áreas que não pertencem à mama; d) Segunda segmentação; e) Remoção de áreas pequenas; f) Remoção de áreas grandes, resultando na não identificação do nódulo.

Na Figura 24 é apresentado um caso em que a área identificada após o processo não corresponde à área do nódulo.

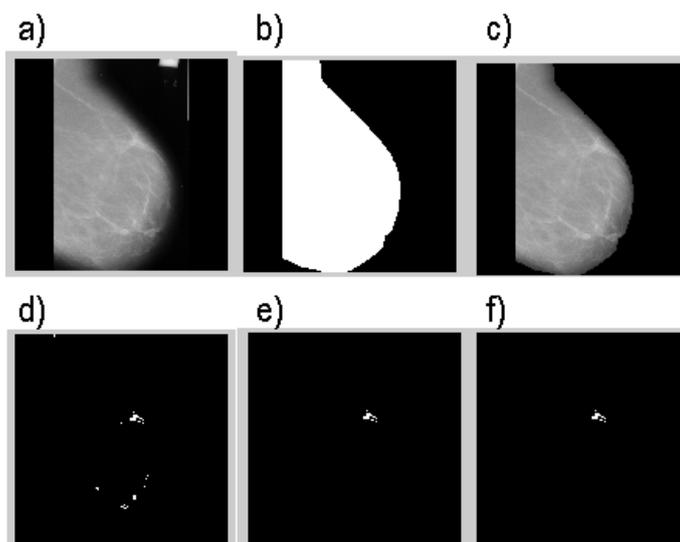


Figura 24 - Mamografia mdb314. Exemplo de teste em que foi detectada uma área diferente da área do nódulo. a) Mamografia original; b) Primeira segmentação; c) Eliminação de áreas que não pertencem à mama; d) Segunda segmentação; e) Remoção de áreas pequenas; f) Remoção de áreas grandes, resultando na identificação de uma área não correspondente ao nódulo.

Para esse teste, também existiram casos em que o nódulo foi identificado, porém, outra área da mama também foi. Das análises de tumores benignos foi obtido um acerto parcial em 35% dos casos analisados.

A Figura 25 apresenta um caso em que a área do nódulo foi identificada, porém outras áreas também o foram.

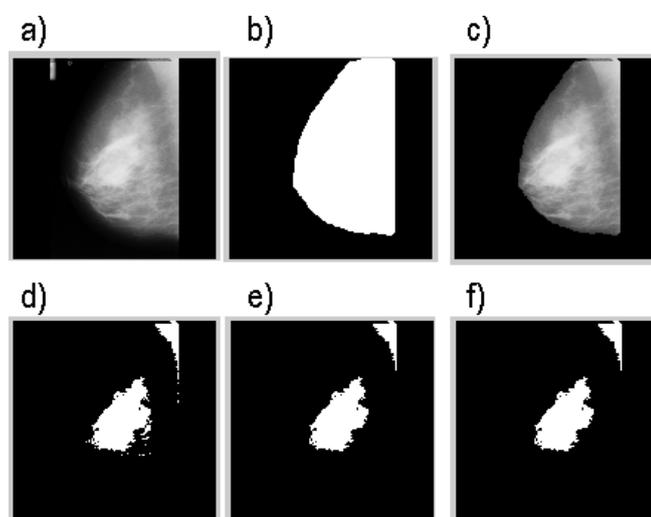


Figura 25 - Mamografia mdb315. Exemplo em que o nódulo mamário foi segmentado corretamente, porém uma área não

correspondente ao nódulo também foi identificada. a) Mamografia original; b) Primeira segmentação; c) Eliminação de áreas que não pertencem à mama; d) Segunda segmentação; e) Remoção de áreas pequenas; f) Remoção de áreas grandes, resultando na identificação do nódulo e de outras regiões.

Por fim, nos casos em que o teste segmentou corretamente os nódulos, de acordo com o especificado na base de dados, foram considerados acertos. Para esse método foi obtido uma taxa de 14% na segmentação de mamografias que apresentavam tumores benignos.

Na Figura 26 pode ser visualizado um caso em que a região correspondente ao nódulo foi corretamente identificada após o processo.

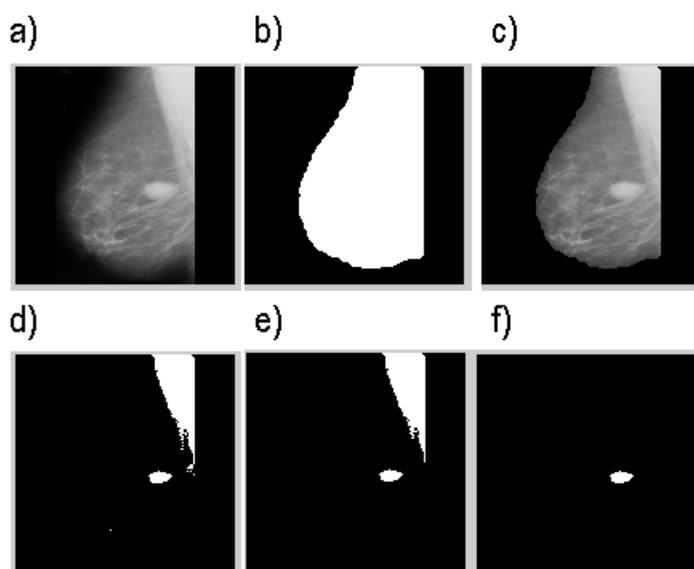


Figura 26 - Mamografia mdb025. Exemplo de segmentação em que apenas o nódulo foi localizado ao final do processo. a) Mamografia original; b) Primeira segmentação; c) Eliminação de áreas que não pertencem à mama; d) Segunda segmentação; e) Remoção de áreas pequenas; f) Remoção de áreas grandes, resultando na identificação do nódulo.

A segmentação pelo método de *threshold* também foi aplicada ao conjunto de imagens classificadas como malignas. Na Tabela 5 podem ser observadas as taxas de erro, acerto parcial e acerto pleno para o teste utilizando este método. Já nas

Figuras 27, 28, 29 podemos visualizar as imagens após o processo em que houve casos de erro, acerto parcial e acerto respectivamente:

Tabela 5 - Resultado da segmentação pelo método de *threshold* para nódulos malignos

Nódulos malignos		
Resultado	Qtd de mamografias	Porcentagem
Erro	14	86.5%
Acerto	5	13.5%

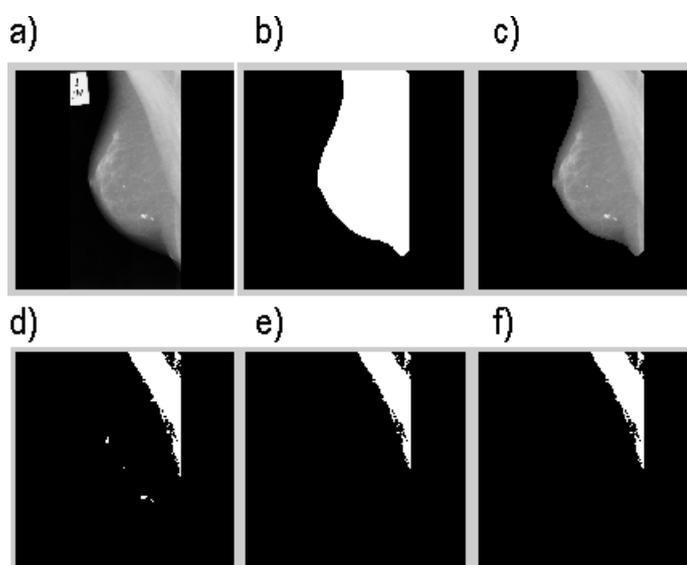


Figura 27 - Mamografia mdb075. Aplicação do teste em mamografia contendo nódulo maligno, onde só foi identificado o músculo peitoral, caracterizando erro. a) Mamografia original; b) Primeira segmentação; c) Eliminação de áreas que não pertencem à mama; d) Segunda segmentação; e) Remoção de áreas pequenas; f) Remoção de áreas grandes, resultando na não identificação do nódulo.

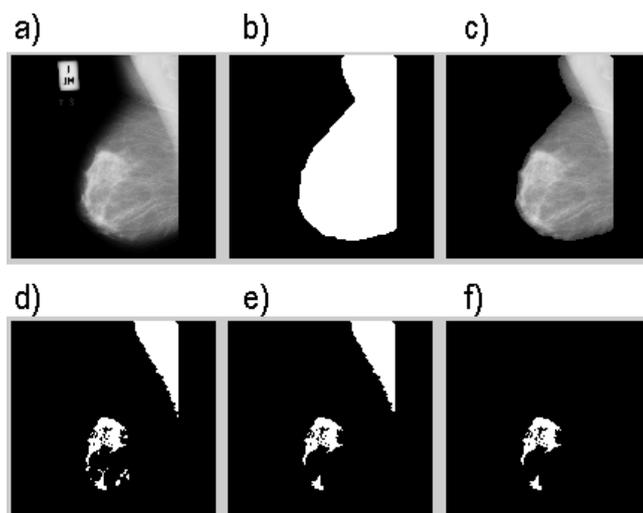


Figura 28 - Mamografia mdb095. Aplicação em que além do nódulo foram identificadas outras áreas da mama. a) Mamografia original; b) Primeira segmentação; c) Eliminação de áreas que não pertencem à mama; d) Segunda segmentação; e) Remoção de áreas pequenas; f) Remoção de áreas grandes, resultando na identificação do nódulo e de outras áreas.

Os resultados obtidos nessa etapa apontam para o fato que mesmo após a etapa de equalização e ajuste da imagem, não foi possível obter-se uma melhora significativa na segmentação dos nódulos por limiarização. Isso ocorreu devido a existência de mamas densas, onde o mamógrafo não consegue extrair todas as características necessárias para identificação da forma do nódulo, e à existência de microcalcificações que embutiram erro durante a segmentação. Além disso, outro fator preponderante para a obtenção de baixos valores de acertos foi o fato de muitas mamografias apresentarem baixo contraste, dificultando a diferenciação dos pixels dos nódulos e dos pixels do resto da mama.

Nesse teste também foram obtidos casos em que o sistema reconheceu o nódulo diagnosticado, porém também reconheceu outra parte da mama.

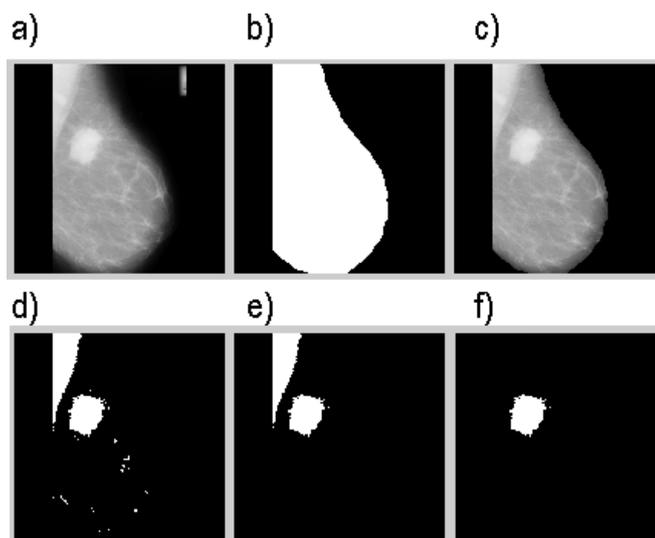


Figura 29 - mamografia mdb184. Aplicação do teste onde o nódulo foi identificado corretamente de acordo com o diagnóstico a) Mamografia original; b) Primeira segmentação; c) Eliminação de áreas que não pertencem à mama; d) Segunda segmentação; e) Remoção de áreas pequenas; f) Remoção de áreas grandes, resultando na identificação correta do nódulo.

A técnica implementada através de limiarização identificou lesões em aproximadamente 54,5% das mamografias analisadas. Nestas imagens, os nódulos estão originalmente mais bem definidos e as imagens possuem melhor definição, comparadas com lesões em outras mamografias.

Nos casos em que o método não funcionou corretamente, as imagens possuem variação nos tons de cinza mais intensa em áreas que não fazem parte dos nódulos, como no músculo peitoral e em regiões de mamas densas.

Além da segmentação por limiarização, também foi utilizado nesse trabalho o método de crescimento de regiões para análise das mesmas imagens. Esse método mostrou-se mais eficiente para segmentação de nódulos mamários que o método de limiarização simples.

Nas Tabelas 6 e 7 podem ser observadas as taxas de erro, acerto parcial e acerto pelo método de crescimento de regiões em imagens contendo nódulos benignos e malignos respectivamente:

Tabela 6 - Resultado da segmentação pelo método de crescimento de regiões para nódulos benignos

Nódulos Benignos		
Resultado	Qtd de mamografias	Percentagem
Erro	18	36.7%
Acerto parcial	11	22.4%
Acerto	20	40.8%

Tabela 7 - Resultado da segmentação pelo método de crescimento de regiões para nódulos malignos

Nódulos Malignos		
Resultado	Qtd de mamografias	Percentagem
Erro	11	50%
Acerto	19	50.0%

Imagens segmentação por crescimento de regiões:

Na Figura 30 pode ser observado um caso em que não foi possível identificar a área do nódulo após o processo.

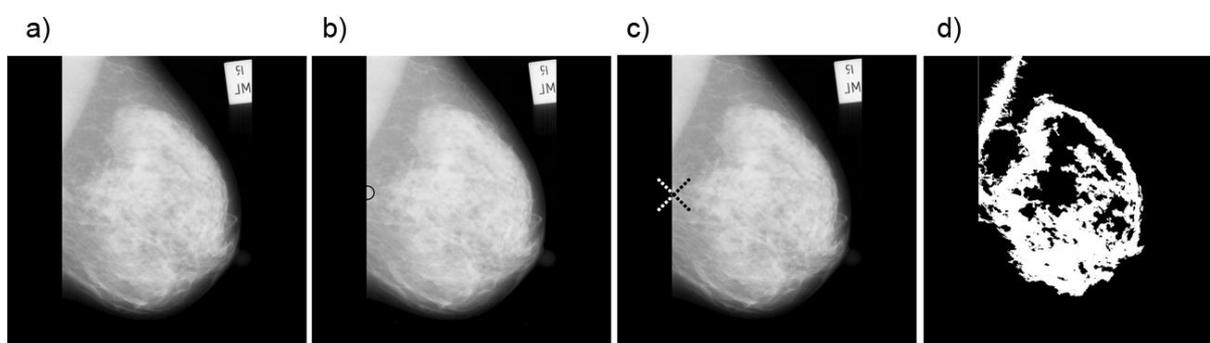


Figura 30 - Mamografia mdb015. Segmentação por crescimento de regiões em que a região segmentada não corresponde a área do nódulo. a) Mamografia original; b) Indicação da região do nódulo; c) Escolha da semente; d) Segmentação

Na Figura 31 pode ser observado um caso em que a área identificada continha o nódulo, porém sem fornecer o seu contorno corretamente.

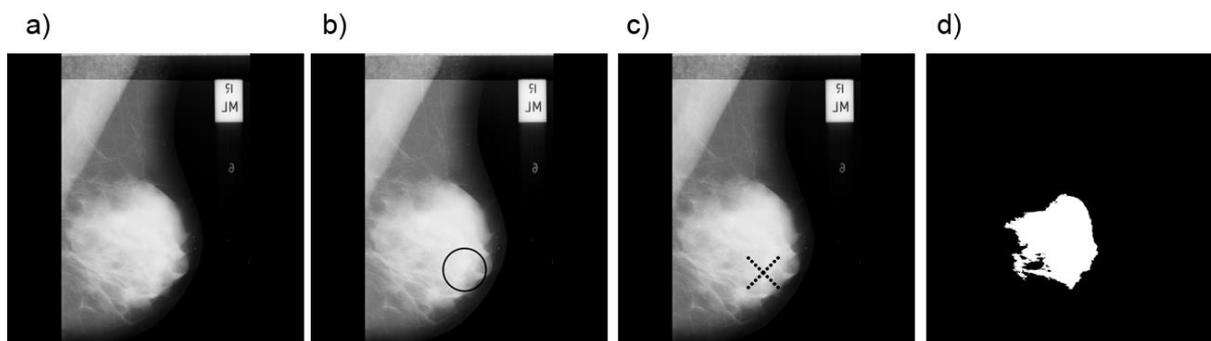


Figura 31 - Mamografia mdb002. Segmentação em que o resultado acertou parcialmente a área ocupada pelo nódulo.

Na Figura 32 pode ser observado um caso em que, após o processo, foi possível identificar corretamente a região do nódulo.

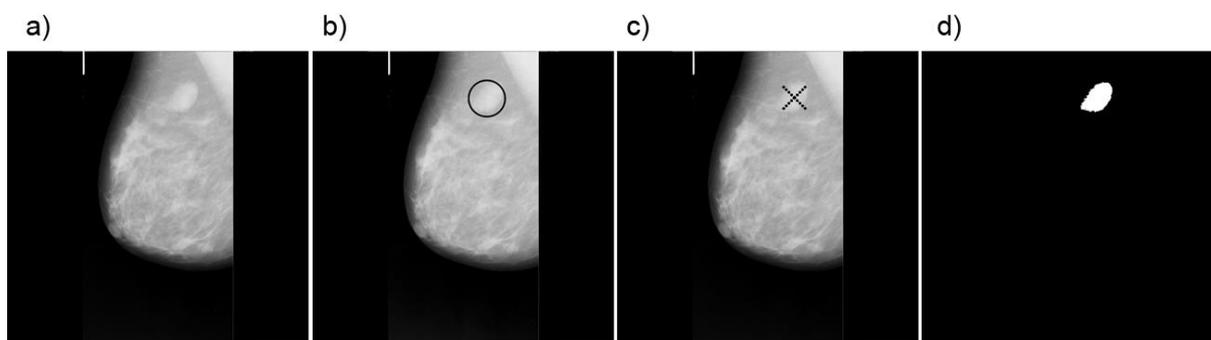


Figura 32 - Mamografia mdb126. Resultado em que a segmentação acerta totalmente na detecção do nódulo.

Nódulos malignos:

A Figura 33 apresenta um caso em que não é possível identificar a região do nódulo.

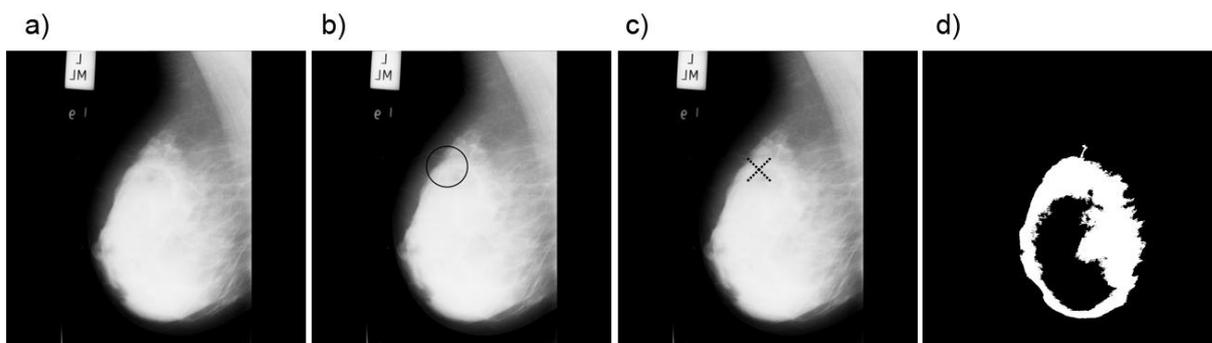


Figura 33 - Mamografia mdb023. Segmentação por crescimento de regiões em que a região segmentada não corresponde a área do nódulo maligno.

Na Figura 34 pode ser observado um caso em que a região selecionada contém o nódulo, porém sem definição correta do contorno do tumor.

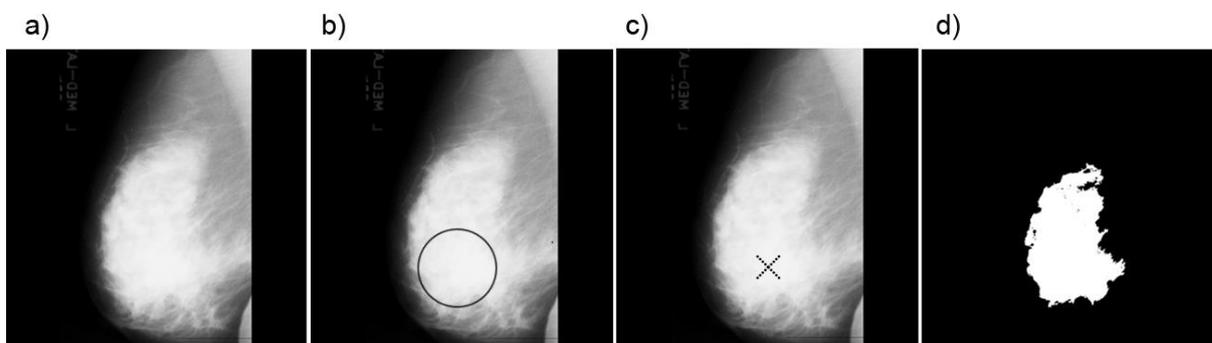


Figura 34 - Mamografia mdb002. Resultado em que a área do nódulo foi selecionada, porém passando um pouco do limite.

Na Figura 35 pode ser observado um caso em que foi possível identificar corretamente a região do nódulo mamário.

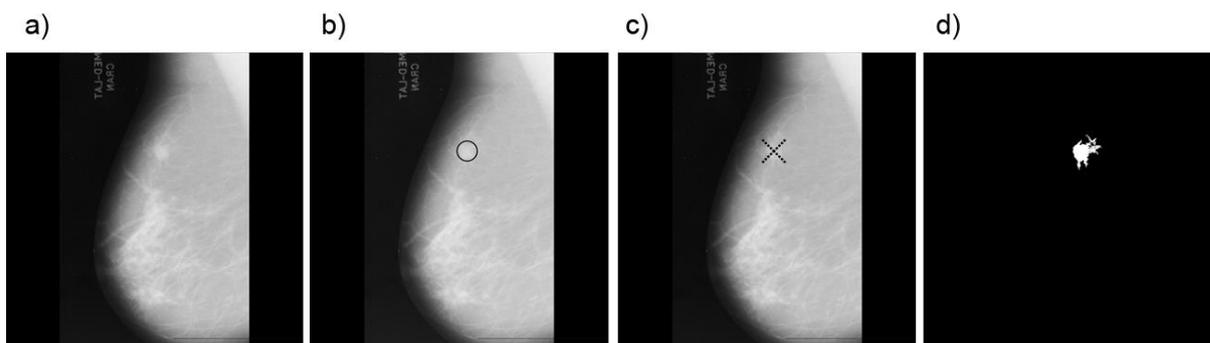


Figura 35 - Mamografia 126. A segmentação ocorreu exatamente na região do nódulo especificado.

O método implementado através da técnica de crescimento de regiões detectou alguma lesão em aproximadamente 66,7% dos mamogramas analisados. Nestas imagens, os nódulos estão originalmente mais enfatizados se comparados a lesões em outras imagens mamográficas.

Nos casos onde o método não funcionou corretamente, aproximadamente 33,3%, os nódulos estão mal definidos e com baixa variação de intensidade de níveis de cinza entre o nódulo e o tecido sadio da mama. Assim, outros objetos de maior intensidade e nitidez são identificados como fazendo parte da lesão.

Após a segmentação para a identificação do nódulo mamário, a imagem passou por uma segmentação baseada em detecção de contornos pelos algoritmos clássicos de Prewitt, Canny, Roberts e Sobel. Devido ao fato do método de Canny possuir maior precisão na detecção de bordas e ser menos suscetível a interferência de ruídos (BIN & YEGANEH, 2012) (ACHARJYA *et al.*, 2012), foi o método utilizado nesse trabalho. Na Figura 36 podem ser constatados esses resultados:

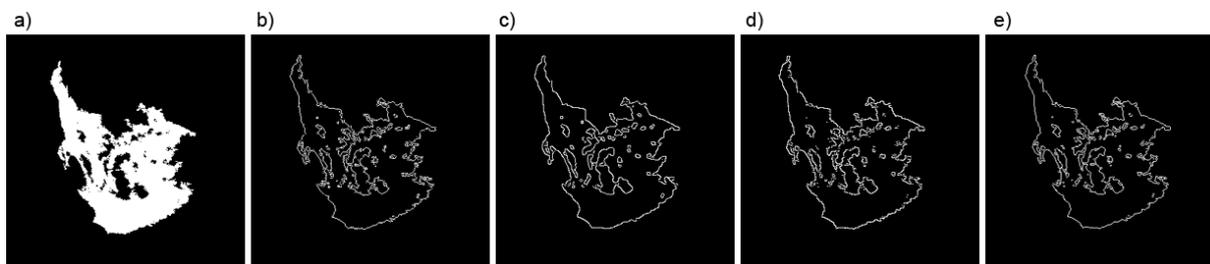


Figura 36 - a) Nódulo maligno; Detecção de bordas pelos métodos de: b) Prewitt; c) Canny; d) Roberts; e) Sobel

O cálculo da dimensão fractal dos nódulos segmentados por detecção de borda mostrou que para os nódulos benignos a maior dimensão calculada foi de 1,226415 e a menor dimensão foi de 0,8154818:

Tabela 8 - Dimensão fractal de nódulos benignos, segmentados pelo método de detecção de bordas de Canny. Maior DF: mdb198; Menor DF: mdb099.

DF - Contorno de Nódulos Benignos							
<i>mdb001</i>	1.104644	<i>mdb063</i>	0.912153	<i>mdb132</i>	0.841785	<i>mdb193</i>	1.040833
<i>mdb002</i>	1.000625	<i>mdb069</i>	0.975989	<i>mdb142</i>	0.877808	<i>mdb195</i>	0.857365
<i>mdb005</i>	0.918982	<i>mdb080</i>	0.939232	<i>mdb145</i>	0.939875	<i>mdb198</i>	1.226415
<i>mdb010</i>	0.85812	<i>mdb081</i>	1.032306	<i>mdb150</i>	1.024079	<i>mdb200</i>	0.961655
<i>mdb012</i>	0.926444	<i>mdb083</i>	0.87879	<i>mdb152</i>	1.057649	<i>mdb204</i>	1.028077
<i>mdb013</i>	0.966534	<i>mdb091</i>	0.900282	<i>mdb160</i>	1.158731	<i>mdb207</i>	1.154834
<i>mdb015</i>	0.871768	<i>mdb097</i>	0.93443	<i>mdb163</i>	1.025267	<i>mdb290</i>	1.027845
<i>mdb017</i>	1.033874	<i>mdb099</i>	0.815482	<i>mdb165</i>	1.029347	<i>mdb312</i>	1.027729
<i>mdb019</i>	0.977808	<i>mdb104</i>	1.039661	<i>mdb167</i>	1.001123	<i>mdb314</i>	1.074735
<i>mdb021</i>	0.898584	<i>mdb107</i>	1.074568	<i>mdb175</i>	0.942417	<i>mdb315</i>	1.076818
<i>mdb025</i>	0.929232	<i>mdb121</i>	1.113314	<i>mdb188</i>	1.123122		
<i>mdb030</i>	0.92854	<i>mdb126</i>	0.850383	<i>mdb190</i>	0.860733		
<i>mdb032</i>	0.907682	<i>mdb127</i>	0.843045	<i>mdb191</i>	0.982814		

Já para os nódulos malignos segmentados por detecção de borda, a maior dimensão calculada foi de 1,2907267 e a menor dimensão foi de 0.7357552:

Tabela 9 - Dimensão fractal de nódulos malignos, segmentados pelo método de detecção de bordas de Canny. Maior DF: mdb148; Menor DF: dmb206;

DF - Contorno de Nódulos Benignos					
<i>mdb023</i>	0.973471	<i>mdb117</i>	1.000983	<i>mdb179</i>	1.053528
<i>mdb028</i>	0.857743	<i>mdb120</i>	1.198584	<i>mdb181</i>	1.219044
<i>mdb058</i>	0.996434	<i>mdb124</i>	1.089482	<i>mdb184</i>	0.938802

<i>mdb072</i>	1.102652	<i>mdb125</i>	1.22125	<i>mdb186</i>	1.145481
<i>mdb075</i>	1.071853	<i>mdb130</i>	1.138694	<i>mdb202</i>	0.845121
<i>mdb090</i>	1.013186	<i>mdb134</i>	0.85157	<i>mdb206</i>	0.735755
<i>mdb092</i>	1.165553	<i>mdb141</i>	1.162799	<i>mdb265</i>	1.000562
<i>mdb095</i>	1.128714	<i>mdb148</i>	1.290727	<i>mdb267</i>	1.05727
<i>mdb102</i>	1.035315	<i>mdb155</i>	0.956796	<i>mdb270</i>	1.017742
<i>mdb105</i>	1.187997	<i>mdb158</i>	0.736632	<i>mdb271</i>	0.910591
<i>mdb110</i>	1.237558	<i>mdb170</i>	1.031402	<i>mdb274</i>	1.085876
<i>mdb111</i>	1.092778	<i>mdb171</i>	1.150184		
<i>mdb115</i>	1.056129	<i>mdb178</i>	1.079685		

As dimensões fractais calculadas pelo método de *box-counting* mostram que os nódulos benignos sofreram menos variação que os nódulos malignos. As Figuras 37 e 38 mostram os dois gráficos de dispersão para os dois grupos de nódulos. Portanto, pode-se concluir que o método, mesmo com um desvio padrão próximos, fornece estimativas de variações distintas para nódulos benignos e malignos.

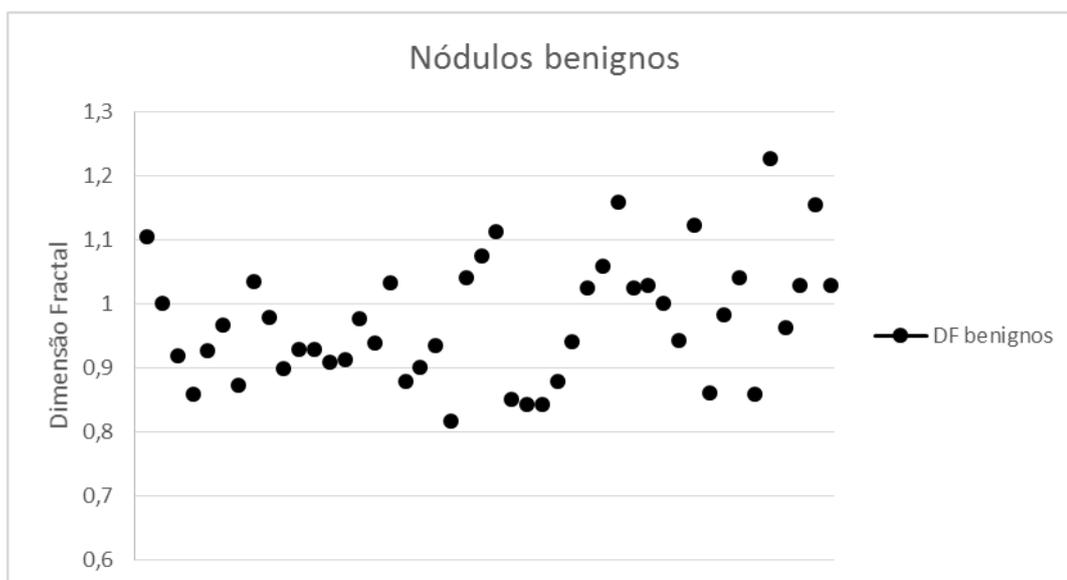


Figura 37 - Dispersão das dimensões fractais de nódulos benignos

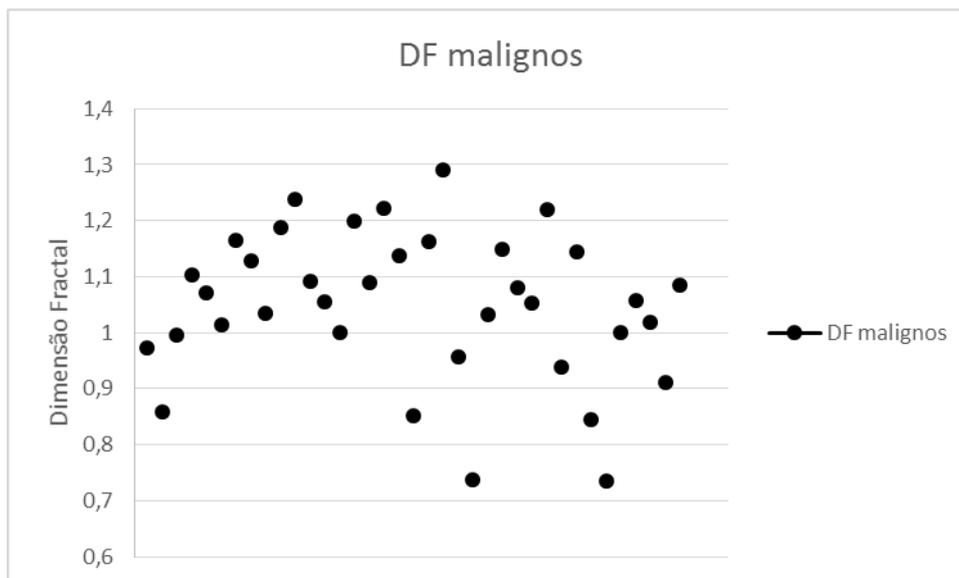


Figura 38 - Dispersão das dimensões fractais de nódulos malignos

Como os nódulos têm valores medidos numa escala numérica, a sensibilidade e a especificidade do teste dependem de onde se coloca o ponto de corte (*cut off*) entre os resultados positivos e negativos, pois a matriz de confusão depende desse ponto para sua definição.

A matriz de confusão para o *cut off* igual a 1, mostra que o método apresenta uma sensibilidade de 0,72972973 e uma especificidade de 0,551020408. Isso quer dizer que quando a probabilidade de o nódulo ser maligno for igual a 100% o teste terá aproximadamente 73% de chance de identificá-lo como tal, e quando a probabilidade do nódulo ser benigno for igual a 100%, o teste terá cerca de 55% de chance de acertar o diagnóstico. Essas probabilidades de sensibilidade podem ser observadas na Tabela 10, para  $P(enf) = 1$  e  $P(enf) = 0$ .

Tabela 10 – Matriz de confusão para limiar igual a 1

		<i>cut off</i> = 1		
		Diagnóstico		
Teste DF		Maligno	Benigno	
	Positivo	27	22	49
	Negativo	10	27	37
		37	49	

Tabela 11 - Análise das probabilidades em relação à Sensibilidade ( $P_+$ ) e Especificidade ( $P_-$ )

$P(enf)$	$P_+$	$P_-$	$VP_+$	$FVP_+$	$VP_-$	$FVP_-$
0	0.44898	0.55102	0	1	1	0
0.1	0.47705	0.52294	0.15296	0.84703	0.94831	0.05168
0.2	0.50513	0.49487	0.28892	0.71107	0.89077	0.10922
0.3	0.53320	0.46679	0.41057	0.58942	0.82630	0.17369
0.4	0.56128	0.43872	0.52004	0.47995	0.75358	0.24641
0.5	0.58935	0.41064	0.61909	0.38090	0.67092	0.32908
0.6	0.61743	0.38257	0.70913	0.29087	0.57612	0.42387
0.7	0.64550	0.35449	0.79133	0.20866	0.46631	0.53368
0.8	0.67358	0.32642	0.86668	0.13331	0.33761	0.66238
0.9	0.70165	0.29834	0.93601	0.06398	0.18469	0.81530
1	0.72973	0.27027	1	0	0	1

Uma forma mais eficiente de demonstrar a relação normalmente antagônica entre a sensibilidade ( $P_+$ ) e a especificidade ( $P_-$ ) dos exames que apresentam resultados contínuos são as curvas ROC. Esse método permite estudar a variação da sensibilidade e especificidade para diferentes probabilidades de  $P(enf)$ .

Na Figura 40 pode ser observada a representação gráfica da curva ROC, que ilustra o desempenho do sistema classificador por *boxcounting* e como o seu limiar de discriminação é variado.

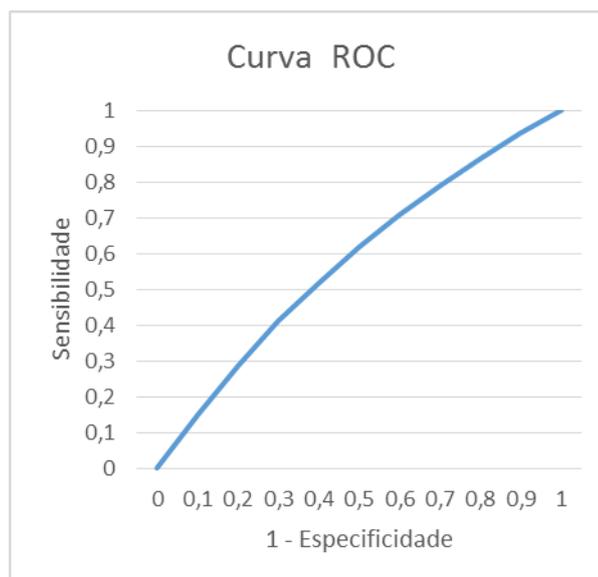


Figura 39 - Gráfico da curva ROC, estabelecendo relação entre Sensibilidade e Especificidade

A área sob a curva é utilizada como medida de eficiência do sistema proposto nesse trabalho, e pode ser interpretada como a capacidade do sistema em prever o resultado. Essa área sob a curva é calculada por:

$$Az = VP * P + VN * N \quad (10)$$

Onde,  $Az$  é a área sub a curva,  $VP$  a porcentagem de resultados verdadeiros positivos,  $P$  a porcentagem dos casos positivos em relação ao total de casos,  $VN$  a porcentagem dos resultados verdadeiros negativos e  $N$  a porcentagem de casos negativos em relação ao total.

Para um  $cut\ off = 1$ , temos:

$$Az = \frac{27}{37} * \frac{37}{86} + \frac{27}{49} * \frac{49}{86} = 0,627906977 \quad (11)$$

Para esse teste, escolhendo um  $cut\ off = 1$ , o sistema apresentou uma eficiência de aproximadamente 62,7%. Se escolhermos uma valor de  $cut\ off = 1,1$

teremos uma sensibilidade = 0,648648649 e uma especificidade = 0,87755102, mostrando que esse ponto de corte diminui um pouco a detecção de nódulos malignos, porém aumenta muito a chance de diagnosticar corretamente nódulos benignos, aumentando no geral o desempenho, como pode ser visto em Figura 40.

Tabela 12 - Matriz de confusão para *cut off* = 1,1.

cut off = 1,1		Diagnóstico		
		Maligno	Benigno	
Análise	Positivo	24	6	30
	Negativo	13	43	56
		37	49	

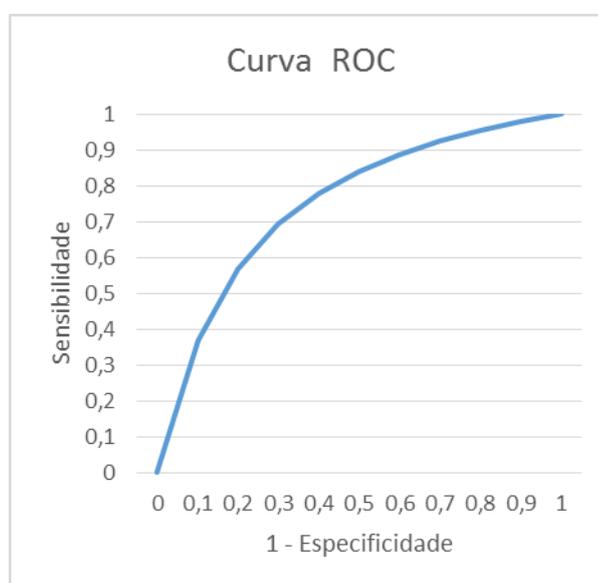


Figura 40 - Curva ROC para valor de *cut off* = 1,1.

Com relação aos métodos utilizados, percebe-se que a dimensão fractal proporcionou o fornecimento de parâmetros necessários para a classificação de nódulos. Isso se deve pelo fato da dimensão fractal apresentar uma maior tolerância a pequenas variações na disposição dos pixels da borda do nódulo.

Para o *cut off* = 1,1, foi obtido um valor de  $Az = 0,779069767$ , melhor resultado entre os valores de *cut off* aplicados, o que, segundo os estudos de (DANAGA *et al.*, 2009) (ABDELAAL *et al.*, 2010a) (JIANG *et al.*, 1999) indicam que o

método proposto nesse trabalho é satisfatório para a classificação dos nódulos mamários. Porém existem ressalvas quando se trata do câncer de mama, devido aos impactos que um diagnóstico errado, seja ele falso positivo ou falso negativo, pode ter sobre o paciente. No entanto a classificação de nódulos mamários pelo método de *box-counting* tem se mostrado bastante promissora como demonstrado nesse trabalho para as bases de imagens estudadas.

A partir dos experimentos realizados através dos métodos propostos, pôde-se constatar que a medida de dimensão fractal possui grande potencial para análise do contorno de nódulos. Os resultados obtidos mostraram-se satisfatórios, apesar da utilização de técnicas de classificação menos elaboradas do que as que utilizam descritores de Fourier ou *wavelet*, ratificando o potencial da análise de borda como característica para classificação de nódulos.

#### 4. CONCLUSÃO

A configuração típica de um sistema *CADx* consiste na extração das informações visuais relevantes de uma imagem na forma de valores, fornecendo subsídios na tomada de decisão durante o diagnóstico médico para análise de doenças como o câncer de mama, cujo número de mortes têm aumentado a cada ano. Com a premissa dos sistemas *CADx* de diminuir o número de biópsias desnecessárias, assim como o número de mortes ocasionados pela doença, nesse trabalho foram aplicadas técnicas que culminaram na quantificação de nódulos, que foram calculados através de *box-counting* e foram utilizados como fonte para classificação do sistema proposto.

Devido a problemas relacionados à diferença existente entre a percepção do nódulo pelo especialista médico e as características automaticamente extraídas de nódulos mamários, um aspecto desafiador dos sistemas *CADx* consiste em obter quantificação da forma da irregularidade, de modo que possa descrevê-lo adequadamente, permitindo sua classificação.

Para chegarmos às conclusões desse trabalho foi feito um levantamento bibliográfico atualizado sobre alguns trabalhos importantes na utilização de dimensão fractal aplicados a sistemas *CADx*, bem como a realização da análise de várias mamografias, de duas bases de dados, possuindo diferentes diagnósticos e características.

Nesse processo foram testados os métodos de segmentação por limiarização simples e crescimento de regiões, realizando uma análise comparativa e validação desses métodos, que apresentaram taxas de acerto satisfatórias para imagens mamográficas de boa qualidade e índices altos de erros para imagens com pouca distinção entre os nódulos e outras regiões da mama.

O método de segmentação por crescimento de regiões aplicado nesse trabalho apresentou uma vantagem significativa em relação a segmentação por limiarização simples, mesmo para os casos em que a mamografia apresentava baixa resolução espacial e baixo nível de contraste.

Os trabalhos e aplicações desenvolvidas nessa pesquisa evidenciaram o fato da dimensão fractal ser uma excelente ferramenta para analisar, reconhecer e

quantificar padrões de contorno dos nódulos de uma imagem, mesmo quando as etapas iniciais de segmentação e detecção não são totalmente precisas.

Este trabalho apresentou uma modelagem utilizando conceito de dimensão fractal para contribuir com o diagnóstico de mamografias. No entanto, vale ressaltar que nenhum sistema é capaz de substituir a análise clínica especializada, e em todos os casos, o diagnóstico final é de responsabilidade do médico e sua equipe.

#### **4.1. Trabalhos Futuros**

Com as técnicas apresentadas nesse trabalho foram obtidos resultados satisfatórios. No entanto, ainda é possível realizar o desenvolvimento de propostas e técnicas como trabalhos futuros:

- Investigação do efeito de realce no domínio da frequência nas mamografias;
- Análise de outras características dos nódulos, como textura e posicionamento;
- Implementação de rotinas para processamento paralelo do cálculo da dimensão fractal.
- Trabalhar com uma equipe de radiologistas para validação dos testes em base de imagens que não possuem diagnóstico;
- Aplicação de outras técnicas de reconhecimento de padrões, como as transformadas *Wavelets*;
- Empregar abordagem multifractal para análise, visto que os nódulos possuem diferentes dimensões em diferentes escalas.

## 5. REFERÊNCIAS

ABDELAAL, M. M. A. et al. **Using data mining for assessing diagnosis of breast cancer**. Computer Science and Information Technology (IMCSIT). **Anais...Cairo, Egypt: IEEE, 2010a**

ABDELAAL, M. M. A. et al. **Using data mining for assessing diagnosis of breast cancer** Proceedings of the International Multiconference on Computer Science and Information Technology. **Anais...IEEE, out. 2010b** Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5679647>>. Acesso em: 7 jan. 2015

ACHARJYA, P. P.; DAS, R.; GHOSHAL, D. Study and Comparison of Different Edge Detectors for Image Segmentation. **Global Journal of Computer Science and Technology - Graphics & Vision**, v. 12, n. 13, p. 5, 2012.

ACKERMAN, L.V.; GOSE, E.E. Breast lesion classification by computer and xeroradiograph. **Cancer**, v. 30, n. 4, p. 1025, 1072.

ACR STANDARDS. **ACR PRACTICE PARAMETER FOR THE PERFORMANCE OF SCREENING AND DIAGNOSTIC MAMMOGRAPHY**. ACR, , 2014. Disponível em: <<http://www.acr.org/~media/3484ca30845348359bad4684779d492d.pdf>>

ADDISON, P. S. Fractals and Chaos. **CRC Press**, p. 256, 1997.

ALMEIDA, A. B. Usando o Computador para Processamento de Imagens Médicas. **Revista Informática Médica**, v. 1, 1998.

AMERICAN CANCER SOCIETY. **Cancer Statistics 2004**. Disponível em: <[www.cancer.org](http://www.cancer.org)>. Acesso em: 8 jul. 2013.

AMERICAN COLLEGE OF RADIOLOGY. American College of Radiology (ACR) Breast Imaging Reporting and Data System Atlas (BIRADS Atlas). 2013.

ANDREADIS, I.; NIKITA, K.; SPYROU, G. **Investigating the performance of a CADx scheme for mammography in specific BIRADS categories** 2014 IEEE International Conference on Imaging Systems and Techniques (IST) Proceedings. **Anais...IEEE, out. 2014** Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6958500>>. Acesso em: 5 jan. 2015

ANDRIOTTI, J. L. S. **Aplicações da Teoria dos Fractais em Geociências**, 2001. Disponível em: <[http://www.cprm.gov.br/publique/media/Art\\_aplic\\_teo\\_fractais\\_Andriotti.pdf](http://www.cprm.gov.br/publique/media/Art_aplic_teo_fractais_Andriotti.pdf)>. Acesso em: 27 nov. 2014

BACKES, A. R.; MARTINEZ, O. B. **Técnicas de Estimativa da Dimensão Fractal: Um Estudo Comparativo**. [s.l.] USP - Universidade São Paulo, São Paulo, 2005.

BALGHONAIM, A. S.; KELLER, J. M. A maximum likelihood estimate for two-variable fractal surfaces. **IEEE Transactions on Image Processing**, p. 1746–1753, 1998.

BALL, J. E.; BUTLER, T. W.; BRUCE, L. M. **Towards Automated Segmentation and Classification of Masses in Digital Mammograms.** 26<sup>a</sup> Annual International Conference of the IEEE Embs. **Anais...**San Francisco: 2004

BIN, L.; YEGANEH, M. S. Comparison for Image Edge Detection Algorithms. **IOSR Journal of Computer Engineering (IOSRJCE)**, v. 2, n. 6, p. 1–4, 2012.

CARLSSON, G. A. et al. Generalized use of contrast degradation and contrast improvement factors in diagnostic radiology. Application to vanishing contrast. **Physics in Medicine and Biology**, p. 737–749, 1986.

CHAN, H. P. et al. Improvement in radiologists' detection of clustered microcalcifications on mammograms. The potential of computer-aided diagnosis. **Investigative radiology**, v. 25, n. 10, p. 1102–10, out. 1990.

CHISTOYIANNI, I; KOUTRAS, A; DERMATAS, G; KOKKINAKIS, G. Computer aided diagnosis of breast cancer in digitalized mammograms. **Computerized Medical Imaging and Graphics**, v.26, p. 309-319, 2002.

COSTA, A. **Mineração de imagens médicas utilizando características de forma.** [s.l.] USP, 2012.

CRISAN, D. A.; DOBRESCU, R.; PLANINSI, P. **Mammographic Lesions Discrimination Based on Fractal Dimension as an Indicator** 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services. **Anais...IEEE**, 2007

DANAGA, A. R. et al. Avaliação do desempenho diagnóstico e do valor de corte para o índice de respiração rápida e superficial na predição do insucesso da extubação. **Jornal Brasileiro de Pneumologia**, v. 35, n. 6, p. 541–547, jun. 2009.

DIEKMANN, F. et al. Evaluation of a wavelet-based computer-assisted detection system for identifying microcalcifications in digital full-field mammography. **Acta Radiologica**, v.45, n. 2, p. 136-141, 2004.

DODSON, M. M.; KRISTENSEN, S. Hausdorff Dimension and Diophantine Approximation. **American Mathematical Society**, 2003.

DOI, K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. **Computerized medical imaging and graphics: the official journal of the Computerized Medical Imaging Society**, v. 31, n. 4-5, p. 198–211, 2007.

FARMER, J. D.; OTT, E.; YORKE, J. A. The dimension of chaotic attractors. **Physica D: Nonlinear Phenomena**, v. 7, n. 1-3, p. 153–180, 1983.

GIGER, M. L. Computerized analysis of images in the detection and diagnosis of breast cancer. **Seminars in ultrasound, CT, and MR**, v. 25, n. 5, p. 411–8, out. 2004.

GOES, E. S. DE; AMBRÓSIO, P. E. **Caracterização de Nódulos em Imagens Mamográficas com Base no Cálculo de Dimensão Fractal** (UESC, Ed.)16° Encontro de Modelagem Computacional. **Anais...**Ilhéus: 2013

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. 3. ed. [s.l.] Pearson Education, 2011. v. 21p. 976

GAVRIELIDES, M.A.; LO, J.Y.; VARGAS-VORACEK, R; FLOYD, C.E. Segmentation of suspicious clustered microcalcifications in mammograms. **Medical Physics**, v. 27, n.1, p. 13-22, 2000.

HEYWANG-KOBRUNNER, S. H. et al. **Mama - Diagnóstico por Imagem**. 1. ed. [s.l.] REVINTER, 1999. p. 412

INCA. MAMOGRAFIA: da prática ao controle - Recomendações para os profissionais de saúde. 2007.

INCA. **O que é o câncer? - INCA**. Disponível em: <[http://www1.inca.gov.br/conteudo\\_view.asp?id=322](http://www1.inca.gov.br/conteudo_view.asp?id=322)>. Acesso em: 2 abr. 2014a.

INCA. **MAMA**. Disponível em: <<http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/mama>>. Acesso em: 13 abr. 2014b.

JAMPALA, S. **Fractals: classification, generation and applications**[1992] Proceedings of the 35th Midwest Symposium on Circuits and Systems. **Anais...IEEE**, 1992Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=271120>>. Acesso em: 27 nov. 2014

JIANG, Y. et al. Improving breast cancer diagnosis with computer-aided diagnosis. **Academic Radiology**, v. 6, n. 1, p. 22–33, jan. 1999.

KLINKENBERG, B. A Review of Methods Used to Determine the Fractal Dimension of Linear Features. **Mathematical Geology**, v. 26, n. 1, p. 23–46, 1994.

KOPANS, D. B. **Imagem da Mama**. 2. ed. Rio de Janeiro: Revinter, 2000. p. 852

LIMA, J. J. P. DE. **Técnicas de Diagnóstico com Raios X. Aspectos Físicos e Biofísicos**. 2<sup>a</sup>. ed. Coimbra: Imprensa da Universidade de Coimbra, 2009. p. 761

MANDELBROT, B. B. **Fractals: Form, Chance and Dimension**. 1st. ed. San Francisco: W. H. Freeman & Company, 1977. p. 365

MANDELBROT, B. B. **The Fractal Geometry of Nature**. 1. ed. Nova York: W. H. Freeman & Company, 1982. p. 468

MARTINEZ, A. C. **Desenvolvimento de Novas Técnicas para Redução de Falsopositivo e Definição Automática de Parâmetros em Esquemas de Diagnósticos Auxiliado por Computador em Mamografia.** [s.l.] Universidade de São Paulo, São Carlos, 2007.

MARTINS, M. N.; GOES, E. S.; AMBROSIO, P. E. **SEGMENTAÇÃO POR CRESCIMENTO DE REGIÕES EM IMAGENS DE MAMOGRAFIA** XXIV Congresso Brasileiro de Engenharia Biomédica - CBEB 2014IV. **Anais...**Uberlândia, Minas Gerais: CBEB, 2014

MATHEUS, B. R. N. **BancoWeb : base de imagens mamográficas para auxílio em avaliações de esquemas CAD.** [s.l.] Universidade de São Paulo, São Carlos, 2010.

MATHEUS, B. R. N.; SCHIABEL, H. Online Mammographic Images Database for Development and Comparison of CAD Schemes. **Journal of Digital Imaging**, v. 24, n. 3, p. 500–506, 2011.

MEDICA, S. B. D. C. **Diagnostico e Tratamento, Volume 3.** 1ª. ed. [s.l.] MANOLE, 2007. p. 1774

MINISTÉRIO DA SAÚDE. **Controle do Câncer de Mama - Documento de Consenso.** Disponível em: <<http://www.inca.gov.br/publicacoes/Consensointegra.pdf>>. Acesso em: 27 nov. 2014.

NGUYEN, T. M.; RANGAYYAN, R. M. **Shape analysis of breast masses in mammograms via the fractal dimension**@7th Annual International Conference of the Engineering in Medicine and Biology Society. **Anais...**IEEE, 2005

NORVARTIS ONCOLOGY. **Câncer de Mama Avançado - Guia para o Escritor Científico.** [s.d.].

OMS. **OMS | Câncer.** Disponível em: <<http://www.who.int/mediacentre/factsheets/fs297/es/>>. Acesso em: 27 nov. 2014.

PAPADOPOULOS, A; FOTIADIS, D.I.; LIKAS, A. An automatic microcalcification detection system based on a hybrid neural network classifier. **Artificial Intelligence in Medicine**, v.25, n.2, p. 149-167, 2002.

PEITGEN, H. O.; JURGENS, H.; SAUPE, D. **Chaos and Fractals: New Frontiers of Science**Springer. **Anais...**Berlin: 1992

PRATI, R. C.; BATISTA, G. E. A. P. A.; MONARD, M. C. Curvas ROC para avaliação de classificadores. **Revista IEEE América Latina**, v. 6, p. 215–222, 2008.

SANKAR, D.; THOMAS, T. **Analysis of mammograms using fractal features.**In World Congress on Nature & Biologically Inspired Computing (NaBIC 2009). **Anais...**IEEE, 2009

SANTOS, V. T. **Segmentação de imagens mamográficas para detecção de nódulos em mamas densas.** [s.l.] USP, 2002.

SCLOWITZ, M. L. et al. Conduas na prevenção secundária do câncer de mama e fatores associados. **Rev Saúde Pública**, p. 340–9, 2005.

SILVA, M. M. DA; SOUZA, W. A. Dimensão Fractal. **REMat - Revista Eletrônica de Matemática**, v. 2, p. 1–15, 2010.

SILVA, P. A. DA; RIUL, S. DA S. **Câncer de mama: fatores de risco e detecção precoce.** Disponível em: <<http://www.scielo.br/pdf/reben/v64n6/v64n6a05.pdf>>. Acesso em: 27 nov. 2014.

SILVA, T. C. et al. **Detecção Automática de Tumores em Mamografias Utilizando técnicas de Processamento Digital de Imagens**XXIV Congresso Brasileiro de Engenharia Biomédica - CBEB 2014. **Anais...2014**Disponível em: <[http://www.canal6.com.br/CBEB2014/artigos/cbeb2014\\_submission\\_402.pdf](http://www.canal6.com.br/CBEB2014/artigos/cbeb2014_submission_402.pdf)>. Acesso em: 28 nov. 2014

SPRING. **Manuais: Tutorial de Geoprocessamento.** Disponível em: <[http://www.dpi.inpe.br/spring/portugues/tutorial/introducao\\_pro.html](http://www.dpi.inpe.br/spring/portugues/tutorial/introducao_pro.html)>. Acesso em: 28 maio. 2014.

SUCKLING, J. et al. **The mammographic Image Analysis Society Digital Mammogram Database**International Congress Series 1069. **Anais...Exerpta Medica**, 1994Disponível em: <<http://peipa.essex.ac.uk/info/mias.html>>. Acesso em: 28 nov. 2014

TABAR, L.; DEAN, P. B. **Teaching Atlas of Mammography.**Georg Thieme Verlag, 1985.

VIEIRA, S. C. et al. **Oncologia Básica.** 1. ed. Teresina, PI: Fundação Quixote, 2012. p. 304

VOSS, R. F. Characterization and Measurement of Random Fractals. **Physyca Scripta**, v. 1986, 1986.

YARUSSO, L.M., NISHIKAWA, R.M., GIGER, M.L., PAPAIOANNOU,J., BAEHR, A.E., VENTA, L.A., NAGEL., R., KUPINSKI, M.A.. Aplication of Computer-Aided Diagnosis to Full-Field Digital Mammography. In 5<sup>o</sup> International Workshop on Digital Mammography. **Proceedings**, 2000.

ZHOU, X.-H.; OBUCHOWSKI, N. A.; MCCLISH, D. K. **Statistical Methods in Diagnostic Medicine.** 2. ed. New York: Copyright © 2002 John Wiley & Sons, Inc., 2011. p. 592