

UNIVERSIDADE FEDERAL DO MARANHÃO  
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA  
CURSO DE CIÊNCIA DA COMPUTAÇÃO

**JOÃO OTÁVIO BANDEIRA DINIZ**

**ANÁLISE TEMPORAL DE LESÃO EM MAMOGRAFIAS DIGITALIZADAS USANDO  
ÍNDICES DE SIMILARIDADE**

São Luís  
2015

**JOÃO OTÁVIO BANDEIRA DINIZ**

**ANÁLISE TEMPORAL DE LESÃO EM MAMOGRAFIAS DIGITALIZADAS USANDO  
ÍNDICES DE SIMILARIDADE**

Monografia apresentada ao curso de Ciência da Computação da Universidade Federal do Maranhão como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Aristófanés Corrêa Silva

São Luís  
2015

Diniz, João Otávio Bandeira.

Análise temporal de lesão em mamografias digitalizadas usando índices de similaridade/ João Otávio Bandeira Diniz. – São Luís, 2015.

60 f.

Impresso por computador (fotocópia).

Orientador: Aristófanês Correa Silva.

Monografia (Graduação) – Universidade Federal do Maranhão, Curso de Ciência da Computação, 2015.

1. Processamento de imagens - Lesão em Mamografia. 2. Análise Temporal. 3. Índices de Similaridade. I. Título.

CDU 004.384

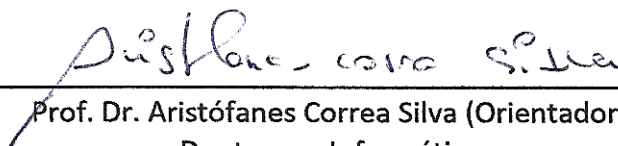
**JOÃO OTÁVIO BANDEIRA DINIZ**

**ANÁLISE TEMPORAL DE LESÃO EM MAMOGRAFIAS DIGITALIZADAS USANDO  
ÍNDICES DE SIMILARIDADE**

Monografia apresentada ao curso de Ciência da Computação da Universidade Federal do Maranhão como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação.

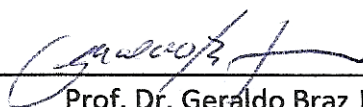
Aprovada em: 16/01/2015

BANCA EXAMINADORA



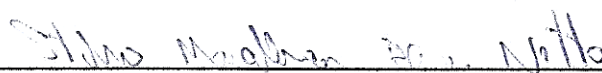
---

Prof. Dr. Aristófanis Correa Silva (Orientador)  
Doutor em Informática  
Universidade Federal do Maranhão



---

Prof. Dr. Geraldo Braz Júnior  
Doutor em Engenharia Elétrica  
Universidade Federal do Maranhão



---

Prof. Ms. Stelmo Magalhães Barros Netto  
Mestre em Engenharia Elétrica  
Universidade Federal do Maranhão

À minha família, em especial meu avô  
Bertoldo e minha avó Enuze.

## AGRADECIMENTOS

Primeiramente a Deus que me permitiu todas as alegrias, conquistas e que me deu forças para lutar sem desistir ou recuar.

Minha família pelo apoio durante todos os anos da minha vida, especialmente meu irmão, Pedro, por sempre me auxiliar durante minha graduação.

Ao amigo, professor e orientador Aristófanés, pela orientação, paciência e conselhos, nessa monografia e em outros trabalhos realizados no decorrer da minha vida acadêmica.

Minha namorada, pelo apoio, força e incentivo.

Meus professores, pelo conhecimento que me proporcionaram. Em especial aos professores: Anselmo, Geraldo e João Dallyson, por sempre se manterem dispostos a auxiliar de alguma forma.

Meus amigos do NCA que estiveram sempre aptos a ajudar, em especial, Caio Belfort, Caio Eduardo, Giovanni e Jefferson que estiveram comigo desde o primeiro período do curso.

A todos que contribuíram direta ou indiretamente para a conclusão deste trabalho.

*“Saber interpor-se constantemente entre si próprio e as coisas é o mais alto grau da sabedoria e prudência.” (Fernando Pessoa)*

## RESUMO

O processamento de imagens mamográficas tem contribuído para a descoberta e o diagnóstico de lesões mamárias, sendo uma importante ferramenta, pois reduz o grau de incerteza do diagnóstico, provendo uma fonte adicional de informação ao especialista. Uma forma de verificar a malignidade de uma lesão é a análise temporal de suas medidas de geometria (forma) e textura (tecido). Entre medidas geométricas comumente utilizadas, a taxa de crescimento da lesão, pode ser utilizada como um dos métodos mais precisos não invasivos de determinação da malignidade. No entanto, somada a outras medidas de textura obtidas no decorrer do tempo, obtém-se informações valiosas sobre o seu comportamento, de forma que o médico pode usar essas medidas na tomada de decisões. Estabelecendo uma correspondência entre lesões encontradas em mamografias em tempos distintos, é possível analisar as mudanças no tecido. Devido a uma série de fatores anatômicos e de aquisição de imagens, uma mesma lesão pode não estar na mesma localização em exames diferentes. Para corrigir esse problema, registros de imagem rígidos e deformáveis mostram-se eficientes. Uma vez estabelecida essa correspondência, é possível analisar as mudanças na textura da lesão através de índices de similaridade. Nesse sentido, será apresentado um método para verificar se lesões encontradas em exames diferentes correspondem a mesma lesão através da aplicação de registros de imagens. Para tanto, um método de análise de mudança de textura foi construído para apresentar o quão similar ou não são essas lesões. Assim o objetivo dessa monografia é apresentar métodos para a análise quantitativa de mudanças de textura na lesão no decorrer do tempo.

Palavras-chave: Processamento de imagens. Lesão em Mamografia. Análise Temporal. Índices de Similaridade



## ABSTRACT

The processing of mammographic images has contributed to the discovery and diagnosis of breast lesions and is an important tool because it reduces the degree of diagnostic uncertainty, providing an additional source of information to the expert. One way to check the malignancy of a lesion is the temporal analysis of its geometry measures (shape) and texture (tissue). Among commonly used geometric measures, the growth rate of the lesion can be used as one of the accurate noninvasive determination of malignancy. However, together with other texture measurements obtained over time, you get valuable information about their behavior, so that the doctor can use these measures in decision-making. Establishing a link between lesions found in mammograms at different times, you could analyze the changes in the tissue. Due to a series of anatomical factors and image acquisition, the same injury cannot be in the same location in different tests. To correct this problem, rigid and deformable image registers show to be efficient. Once established this correspondence, you can analyze the changes in the lesion texture through similarity indices. In this sense, will be presented a method to verify that lesions in different exams match the same injury by applying images registers. Therefore, a method of texture analysis is done by changing how to display similar or not the lesion. So, the goal of this monograph is to provide methods for the quantitative analysis of texture changes in the lesion over time.

Keywords: Image processing. Lesion in Mammography. Temporal Analysis. Similarity Indexes

## LISTA DE FIGURAS

Figura 1: Exemplo de uma mamografia com seus principais elementos. Fonte: Sampaio (2008).....	21
Figura 2: Exames mamográficos. (a) Imagem MLO; (b) Imagem CC. Fonte: Sampaio (2008).....	22
Figura 3: Processo de um registro de imagens. Fonte: (Parraga, 2008).....	26
Figura 4: Fluxo da metodologia. ....	33
Figura 5: Aquisição de imagem (a) Imagem com marcação do especialista (b) Imagem rotacionada e marcação alterada (c) Imagem com ruído na lesão .....	34
Figura 6: Pré-processamento de imagem (a) Remoção de borda (b) Remoção do fundo (c) Crescimento de Região (d) Remoção de objetos indesejados .....	36
Figura 7: Registro de Imagens (a) Imagem tempo t (b) Imagem no tempo t+1 (c) Imagem no tempo t+1 registrada.....	37
Figura 8: Correspondência entre lesões .....	38
Figura 9: Imagem A_1123_1.LEFT_CC da DDSM (a) Marcação do especialista (b) Marcação Reduzida (c) Marcação Aumentada (d) Ruídos na Lesão .....	41
Figura 10: Imagem A_1242_1.LEFT_MLO da DDSM (a) Marcação do especialista (b) Marcação Reduzida (c) Marcação Aumentada (d) Ruídos na Lesão .....	43
Figura 11: Imagem A_1000_1.LEFT_CC da DDSM (a) Marcação do especialista (b) Marcação Reduzida (c) Marcação Aumentada (d) Ruídos na Lesão .....	45
Figura 12: Imagem A_1001_1.LEFT_CC da DDSM (a) Marcação do especialista (b) Marcação Reduzida (c) Marcação Aumentada (d) Ruídos na Lesão .....	47
Figura 13: Imagem A_0038_1.RIGHT_CC da DDSM (a) Marcação do especialista (b) Marcação Reduzida (c) Marcação Aumentada (d) Ruídos na Lesão .....	49
Figura 14: Imagem A_1170_1.RIGHT_MLO da DDSM (a) Marcação do especialista (b) Marcação Reduzida (c) Marcação Aumentada (d) Ruídos na Lesão .....	51
Figura 15: Imagem A_1004_1.LEFT_CC da DDSM (a) Marcação do especialista (b) Marcação Reduzida (c) Marcação Aumentada (d) Ruídos na Lesão .....	53
Figura 16: Imagem A_1003_1.LEFT_CC da DDSM (a) Marcação do especialista (b) Marcação Reduzida (c) Marcação Aumentada (d) Ruídos na Lesão .....	55

## LISTA DE TABELAS

Tabela 1: Índices de similaridade. Adaptado de (Meyer, 2002).....	31
Tabela 2: Teste na imagem A_1123_1.LEFT_CC. Resultado dos índices comparados a marcação original e a marcação reduzida.....	41
Tabela 3: Teste na imagem A_1123_1.LEFT_CC. Resultado dos índices comparados a marcação original e marcação aumentada.....	42
Tabela 4: Teste na imagem A_1123_1.LEFT_CC. Resultado dos índices comparados a marcação original e marcação com ruídos.....	42
Tabela 5: Teste na imagem A_1242_1.LEFT_MLO. Resultado dos índices comparados a marcação original e a marcação reduzida.....	43
Tabela 6: Teste na imagem A_1242_1.LEFT_MLO. Resultado dos índices comparados a marcação original e a marcação aumentada.....	44
Tabela 7: Teste na imagem A_1242_1.LEFT_MLO. Resultado dos índices comparados a marcação original e marcação com ruídos.....	44
Tabela 8: Teste na imagem A_1000_1.LEFT_CC. Resultado dos índices comparados a marcação original e a marcação reduzida.....	45
Tabela 9: Teste na imagem A_1000_1.LEFT_CC. Resultado dos índices comparados a marcação original e a marcação aumentada.....	46
Tabela 10: Teste na imagem A_1000_1.LEFT_CC. Resultado dos índices comparados a marcação original e marcação com ruídos.....	46
Tabela 11: Teste na imagem A_1001_1.LEFT_CC. Resultado dos índices comparados a marcação original e a marcação reduzida.....	47
Tabela 12: Teste na imagem A_1001_1.LEFT_CC. Resultado dos índices comparados a marcação original e a marcação aumentada.....	48
Tabela 13: Teste na imagem A_1001_1.LEFT_CC. Resultado dos índices comparados a marcação original e marcação com ruídos.....	48
Tabela 14: Teste na imagem A_0038_1.RIGHT_CC. Resultado dos índices comparados a marcação original e a marcação reduzida.....	49
Tabela 15: Teste na imagem A_0038_1.RIGHT_CC. Resultado dos índices comparados a marcação original e a marcação aumentada.....	50
Tabela 16: Teste na imagem A_0038_1.RIGHT_CC. Resultado dos índices comparados a marcação original e marcação com ruídos.....	50
Tabela 17: Teste na imagem A_1170_1.RIGHT_MLO. Resultado dos índices comparados a marcação original e a marcação reduzida.....	51
Tabela 18: Teste na imagem A_1170_1.RIGHT_MLO. Resultado dos índices comparados a marcação original e a marcação aumentada.....	52
Tabela 19: Teste na imagem A_1170_1.RIGHT_MLO. Resultado dos índices comparados a marcação original e marcação com ruídos.....	52
Tabela 20: Teste na imagem A_1004_1.LEFT_CC. Resultado dos índices comparados a marcação original e a marcação reduzida.....	53
Tabela 21: Teste na imagem A_1004_1.LEFT_CC. Resultado dos índices comparados a marcação original e a marcação aumentada.....	54
Tabela 22: Teste na imagem A_1004_1.LEFT_CC. Resultado dos índices comparados a marcação original e marcação com ruídos.....	54

Tabela 23: Teste na imagem A_1003_1.LEFT_CC. Resultado dos índices comparados a marcação original e marcação diminuída .....	55
Tabela 24: Teste na imagem A_ A_1003_1.LEFT_CC. Resultado dos índices comparados a marcação original e marcação aumentada .....	56
Tabela 25: Teste na imagem A_1003_1.LEFT_CC. Resultado dos índices comparados a marcação original e marcação com ruídos .....	56

## SUMÁRIO

1. INTRODUÇÃO.....	15
1.1. Motivação.....	17
1.2. Objetivos.....	18
1.3. Trabalhos Relacionados.....	18
1.4. Organização do Trabalho.....	20
2. FUNDAMENTAÇÃO TEÓRICA.....	21
2.1. A Mamografia.....	21
2.2. Processamento de Imagens.....	23
2.2.1. Representação Digital de uma Imagem e <i>Pixel</i> .....	24
2.2.2. Crescimento de Região.....	24
2.3. Registro de Imagens.....	25
2.3.1. Registro Rígido.....	27
2.3.2. Registro Deformável <i>B-spline</i> .....	27
2.4. Índices de Similaridade.....	29
2.5. <i>Insight Segmentation and Registration Toolkit (ITK)</i> .....	31
3. MATERIAIS E MÉTODOS.....	33
3.1. Aquisição de Imagens.....	33
3.2. Pré-processamento das Imagens.....	35
3.3. Registro das Imagens.....	36
3.4. Correspondência entre Lesões.....	37
3.5. Análise Temporal de Textura.....	38
4. TESTES E RESULTADOS.....	40
4.1. Mamas Densas.....	40
4.1.1. Lesões Benignas.....	40
4.1.2. Lesões Malignas.....	44
4.2. Mamas Não-Densas.....	48
4.2.1. Lesões Benignas.....	49
4.2.2. Lesões Malignas.....	52
5. CONCLUSÃO.....	58
5.1. Avaliação do Trabalho.....	58
5.2. Trabalhos Futuros.....	59



## 1. INTRODUÇÃO

A incidência do câncer de mama aumenta 1% ao ano. Segundo o Instituto Nacional de Câncer (INCA), a previsão para 2014 é de 57.120 novos casos desse tipo de câncer e com um risco estimado de 56,09 casos a cada 100 mil mulheres (INCA, 2014).

O câncer de mama de acordo com o INCA (2014) é o tipo de câncer que mais acomete as mulheres em todo. A idade é o principal fator de risco para o câncer de mama. As taxas de incidência aumentam rapidamente até os 50 anos e, posteriormente, esse aumento ocorre de forma mais lenta. Contudo, outros fatores de riscos já estão bem estabelecidos como: aqueles que são relacionados à vida produtiva da mulher (idade da primeira gestação acima do 30 anos, anticoncepcionais orais, menopausa tardia e terapia de reposição hormonal), histórico familiar de câncer de mama e alta densidade do tecido mamário (razão entre o tecido glandular e o tecido adiposo da mama). Além desses, a exposição à radiação ionizante, mesmo em baixas doses, também é considerada um fator de risco, particularmente durante a puberdade, segundo mostram alguns estudos.

A prevenção primária dessa doença ainda não é totalmente possível em decorrência da variação de fatores e das características genéticas que estão envolvidas na sua etiologia. Uma nova característica de rastreamento factível para países com dificuldades orçamentárias tem sido estudada, e, até o momento, a mamografia, para mulheres com idade entre 50 e 69 anos e o exame clínico das mamas anualmente a partir dos 40 anos, é recomendada como método efetivo para a detecção precoce. A amamentação, a prática de atividade física e a alimentação saudável com a manutenção do peso corporal estão associadas a um menor risco de desenvolver esse distúrbio (INCA 2014).

Apesar de ser considerado um câncer de bom prognóstico quando diagnosticado e tratado nas fases iniciais, as taxas de mortalidade por essa doença continuam elevadas, provavelmente porque ela é diagnosticada em estágios avançados. A sobrevida média após cinco anos na população de países desenvolvidos tem apresentado um pequeno aumento, cerca de 85%. Entretanto, em países em desenvolvimento, a sobrevida é próxima a 60%. No Maranhão, esse câncer tem uma taxa de 13,97% para cada 100 mil mulheres (INCA 2014).

As formas mais eficazes para detecção precoce do câncer de mama são o exame clínico da mama e a mamografia. O Exame Clínico das Mamas (ECM) quando realizado por um especialista, pode detectar tumor de até um centímetro, se superficial. Segundo o INCA (2014), o ECM deve contemplar os seguintes passos: inspeção estática e dinâmica, palpação das axilas e palpação da mama com a paciente em decúbito dorsal.

A eficiência do exame é proporcional ao grau de habilidade e experiência do profissional para detectar qualquer anormalidade nas mamas examinadas. Ele deve ser realizado periodicamente e o médico indicará a necessidade de mamografia.

A mamografia é a radiografia da mama que permite a detecção precoce do câncer por ser capaz de mostrar lesões em fase inicial, na faixa de milímetros. É realizada em um aparelho de raios-X apropriado, chamado mamógrafo. Nele, a mama é comprimida de forma a fornecer melhores imagens e, portanto, melhor capacidade de diagnóstico. Em um exame, as duas mamas são analisadas separadamente e, de cada mama, são geradas duas imagens. A primeira imagem é conhecida como Crânio Caudal (CC), que é ortogonal ao eixo vertical da paciente, e a segunda é a Médio Lateral Oblíqua (MLO), que é feita em um ângulo oblíquo em relação ao eixo vertical da paciente.

Estudos sobre a efetividade da mamografia sempre utilizam o exame clínico como exame adicional, o que torna difícil distinguir a sensibilidade do método como estratégia isolada de rastreamento (INCA, 2014). No entanto, a sua precisão depende de diversos fatores, como o tamanho e a localização da lesão, a densidade do tecido mamário e a qualidade dos recursos técnicos utilizados. Além disso, a tarefa de interpretar cuidadosamente um grande número de casos demanda tempo e um nível de atenção muito grande por parte do especialista.

O câncer de mama, como todo tipo de câncer, dependendo se a detecção é precoce ou não, pode ter o diagnóstico limitado. Quando o diagnóstico de uma lesão é limitado, cabe ao médico especialista optar ou não pela remoção cirúrgica. Em alguns casos de diagnóstico indefinido ou, principalmente, em casos de fortes suspeitas de benignidade, a lesão passa a ser acompanhada para ver seu comportamento no decorrer do tempo. Nesse caso, são feitos exames periódicos e são feitas comparações das informações desses exames para chegar ou se aproximar de um diagnóstico.



O acompanhamento de uma doença no decorrer do tempo, através da comparação de exames periódicos, é chamado de análise temporal. A taxa de crescimento da lesão, por exemplo, pode ser utilizada como um dos métodos mais precisos não invasivos de determinação da malignidade. Somada a ela, outras medidas de forma (geométricas) e de tecido (textura) extraídas no decorrer do tempo podem oferecer informações importantes sobre o comportamento da lesão. Essas medidas podem ser usadas pelo médico na tomada de decisões mais coerentes. Por essas razões, confirma-se a importância da análise temporal de lesões mamárias. Esta análise é feita através da comparação de informações extraídas das imagens de exames.

Esta monografia apresenta uma metodologia de análise temporal de lesões mamárias baseada em textura (tecido) utilizando mamografias computadorizadas. Primeiramente, a metodologia busca estabelecer uma correspondência entre a localização de lesões em exames diferentes. Essa correspondência é obtida utilizando, principalmente, técnicas de pré-processamento e registro de imagens. Uma vez estabelecida essa correspondência, são definidas as regiões de interesse das lesões. Essas regiões são, por fim, utilizadas na análise temporal, através da aplicação de índices de similaridade.

## **1.1. Motivação**

O diagnóstico precoce desta doença é um fator de extrema importância para redução das taxas de mortalidade. Porém, existem casos que o diagnóstico é indeterminado, então, se opta por acompanhar a evolução da lesão, esse acompanhamento, geralmente, é eficaz para caracterização da lesão.

Na análise temporal, é comumente usada a forma e o crescimento da lesão, porém somente essas medidas podem não ser suficientes para o diagnóstico. Assim sendo, a mudança na textura da lesão pode indicar a sua natureza em casos que houve pouca mudança na geometria.

O médico, em alguns casos, tem dificuldade em avaliar por si só a textura e, por isso, técnicas que possam auxiliar a validação de mudanças ou não na textura podem fazer toda a diferença na precisão do diagnóstico.

Por essas razões, é necessário o desenvolvimento de técnicas que auxiliem aos médico e especialistas no acompanhamento de lesões e, assim, evitar diagnósticos confusos ou imprecisos.

## **1.2. Objetivos**

Desenvolver um estudo dos índices de similaridade para a análise de mudanças em lesões mamárias no decorrer do tempo. Nesse processo, serão usadas imagens de mamografias digitalizadas de pacientes simulando momentos diferentes.

Esse trabalho não tem como foco o diagnóstico da lesão. Ao invés disso, busca-se oferecer ao médico informações úteis sobre mudanças na textura da lesão no decorrer do tempo de avaliação do paciente. Espera-se que essas informações possam não só auxiliar o médico especialista no diagnóstico como também na decisão do melhor tratamento com base no comportamento da lesão ou como o tratamento está se comportando a partir das mudanças na textura.

Para que seja concretizado o objetivo geral, alguns objetivos específicos devem ser alcançados:

- Estudar, desenvolver e aplicar técnicas de registro de imagens que permita verificar se duas lesões em tempos distintos são de fato a mesma lesão;
- Oferecer ao médico mais subsídios para decisão de diagnóstico;
- Usar índices de similaridade para analisar as mudanças no tecido(textura).

## **1.3. Trabalhos Relacionados**

Alguns trabalhos na literatura propõem formas de correspondência de lesões no tempo, e a análise temporal de lesões.

Lubomir *et al* (2001) desenvolve um sistema de classificação de massas em benigno e maligno baseado em informações adquiridas em intervalo de tempo. Utiliza de característica de comprimento da lesão, característica de textura e características morfológicas para fazer uma previsão do diagnóstico da lesão. Neste trabalho foi feita validação da metodologia em uma base de 140 pares de imagens temporais.

Saskia *et al* (2003) cita que a utilização de registro de mamografias é muito importante para a exibição de casos em estações de visualização digital e também para encontrar regiões correspondentes em análise temporal. Foi feito teste de quatro tipos de registro – *Nipple Alignment*, *Center of Mass Alignment*, *Mutual Information* e *Warping*. Também foi feita comparação entre eles, é citado que o Registro *Mutual Information* superou os outros em alinhamento com base na localização do mamilo.

Sheila *et al* (2005) utiliza técnicas de registro para desenvolver um software para encontrar correspondência de massas em mamografias no decorrer do tempo. Primeiro foi aplicado um registro nas imagens em comparação, depois define-se uma área susceptível de conter a mesma lesão, e por fim utiliza medidas quantitativas para verificar o local que as massas coincidem. Este trabalho obteve o total de acerto de 82% em uma base de 389 casos de massas temporais.

Gu *et al* (2011) compara algoritmos rígidos e deformáveis em eficiência e tempo computacional em imagens. O principal objetivo de seu estudo é verificar quais registros eram melhores para o casamento de nódulos pulmonares em imagens de tomografias computadorizadas. Confirmou-se, em testes, que os registros rígidos têm menos custo computacional, mas em precisão, os melhores são os deformáveis e destaca que os mais utilizados na literatura atual são o *B-spline*, *Demon* e variantes desses.

Diniz (2014) faz uma análise temporal de nódulos e massas pulmonares em tomografias computadorizadas. O principal objetivo deste trabalho, é desenvolver uma metodologia que analise as mudanças do mesmo nódulo ou massa pulmonar no decorrer do tempo, informando medidas quantitativas baseadas nos índices de similaridade. Foi feito uso de técnicas de registro para achar a correspondência entre os nódulos e massas em exames diferentes.

Robert *et al* (2014) fala sobre a detecção de casos anormais em mamografias temporais usando registro de características. Avaliou-se a deformação das características extraídas a partir dos resultados do registro das imagens, a fim de detectar casos anormais de massas. Sua metodologia utiliza pré-processamento de imagem, registro das imagens temporais e classificação das massas baseado em características extraídas da lesão registrada.

Os trabalhos mostram a eficiência dos registros em imagens, principalmente para correspondência das lesões no tempo e para auxiliar na precisão da avaliação médica. Também

mostram que avaliar as informações adquiridas no tempo pode ser de grande ajuda para avaliação da malignidade de lesões. Assim, mostra-se importante a metodologia proposta, tanto para analisar e validar a correspondência da lesão no tempo como para fornecer medidas quantitativas que auxiliem no diagnóstico do paciente em acompanhamento.

#### **1.4. Organização do Trabalho**

Esta monografia apresenta a seguinte organização:

No Capítulo 2, Fundamentação Teórica, segue informações importantes para o contexto e entendimento do trabalho, tais como Processamento de Imagens e Registro de Imagens.

No Capítulo 3, Materiais e Métodos, será explicada a metodologia utilizada como ponto de partida para o desenvolvimento desse trabalho.

No Capítulo 4, Resultados, mostra-se os resultados alcançados na aplicação da metodologia.

No Capítulo 5, Conclusão, apresenta-se a conclusão do trabalho. Nela está contida uma retrospectiva do que foi falado como todo e uma avaliação dos resultados obtidos e os próximos passos a serem tomados.

## 2. FUNDAMENTAÇÃO TEÓRICA

Nesta seção, será abordado os conceitos necessários para entender o funcionamento da metodologia proposta.

### 2.1. A Mamografia

O objetivo da mamografia é produzir imagens de alta resolução das estruturas internas da mama, a fim de permitir a detecção do câncer de mama. Devido ao fato de que as diferenças de contraste entre tecidos doentes e normais são muito pequenas, esse exame requer um equipamento capaz de realçar tais diferenças e fornecer uma resolução de alto contraste.

A Figura 1 apresenta uma mamografia contendo seus principais elementos constituintes.

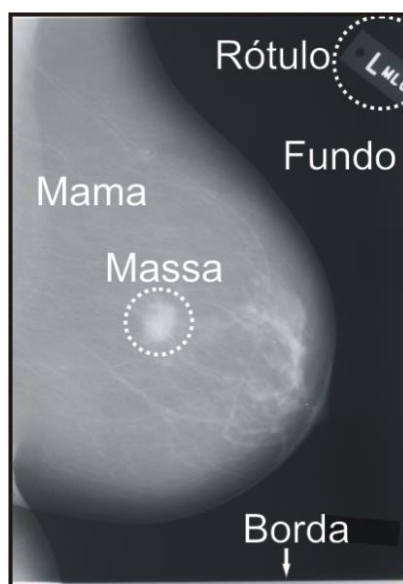
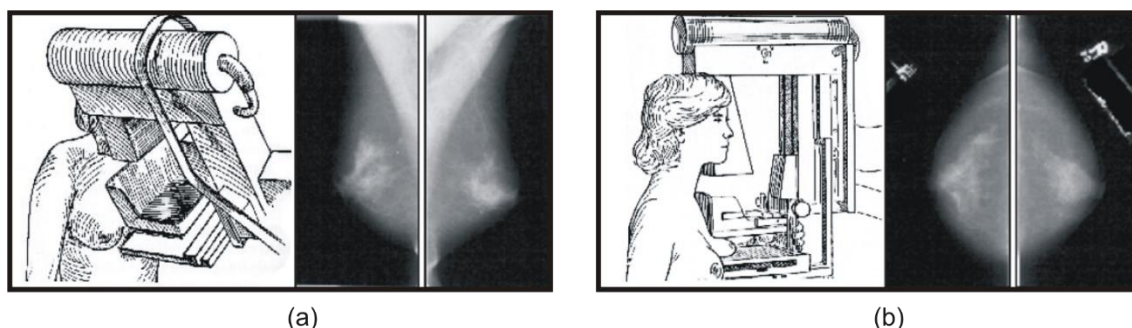


Figura 1: Exemplo de uma mamografia com seus principais elementos. Fonte: Sampaio (2008).

A mamografia deve ser feita em aparelho de raios X específico, chamado mamógrafo. Nele, a mama é comprimida de forma a fornecer melhores imagens e, portanto, melhor capacidade de diagnóstico. Até que sejam planejados sistemas que possam mecanicamente posicionar a mama, são necessários técnicos altamente especializados para posicionar a paciente a fim de obter uma imagem otimizada. A compressão é necessária para evitar a subexposição da base e a superexposição dos tecidos anteriores da mama, mais finos.

Normalmente a mamografia é bilateral, ou seja, é feita uma radiografia de cada mama. Além disso, em um exame de mamografia, duas projeções de cada mama são indispensáveis: uma visão médio-lateral oblíqua (MLO) e uma crânio-caudal (CC), conforme a Figura 2.



**Figura 2: Exames mamográficos. (a) Imagem MLO; (b) Imagem CC. Fonte: Sampaio (2008).**

A projeção MLO é a mais útil, pois permite a visualização do alto da axila para baixo. O termo oblíquo se refere ao plano de compressão da mama. Se a mama não estiver bem posicionada, pode-se deixar de detectar anormalidades. O músculo peitoral deve ser visível, estendendo-se obliquamente até a metade superior da imagem. Além disso, deve ser muito largo no alto e ir se afinando à medida que cruza a parte superior da mama. Estudos sugerem que a mama é representada de maneira ótima quando o músculo peitoral é visível até o eixo do mamilo (KOPANS, 2000).

A projeção CC é a segunda projeção que deve ser obtida rotineiramente. O principal objetivo dessa projeção é obter uma visão da região póstero-medial da mama, complementando a projeção MLO.

A mamografia é um exame de alta sensibilidade. No entanto, sua sensibilidade está diretamente relacionada à idade da mulher, sendo muito menor nas mulheres jovens, que apresentam um tecido mamário bastante denso. Tipicamente, mulheres mais jovens apresentam mamas com maior quantidade de tecido glandular, o que torna esses órgãos mais densos e firmes. Ao se aproximar da menopausa, o tecido glandular vai se atrofiando e sendo substituído progressivamente por tecido gorduroso, até se constituir, quase que exclusivamente, de gordura e resquícios de tecido glandular na fase da pós-menopausa. Essas mudanças de características promovem uma nítida diferença entre as densidades radiológicas das mamas da mulher jovem e da mulher na pós-menopausa, configurando uma dificuldade a mais para o especialista (HEATH, et al. 1998).

Conforme Wang e Karayiannis (1998) existe uma grande dificuldade em se trabalhar com imagens de mamografias, as mesmas apresentam baixo contraste. Isto dificulta a interpretação dos resultados por parte do especialista. Estudos de Bird *et al* (1992) e Kerlikowske (2000) mostram que a mamografia é suscetível a uma alta taxa de falsos positivos como também falsos negativos, causando uma alta proporção de mulheres sem câncer que sofrem uma nova avaliação clínica, enfrentam uma biópsia ou perdem o melhor intervalo de tempo para o tratamento do câncer. Por isso, diversas técnicas têm sido desenvolvidas para ajudar a tornar a mamografia um recurso mais preciso e eficiente. Uma das principais contribuições nessa área são os sistemas de Detecção e Diagnóstico Auxiliado por Computador (CAD / CADx: *Computer-Aided Detection / Diagnosis*).

## **2.2. Processamento de Imagens**

O interesse em técnicas de processamento de imagens digitais surgiu, principalmente, da necessidade de melhorar a qualidade das imagens e fornecer outros subsídios que facilitem a interpretação humana. Ao longo das duas últimas décadas, a área de processamento de imagens digitais experimentou um rápido crescimento, expandindo a cada dia o domínio de aplicações e soluções possíveis. Alguns exemplos são: a análise de recursos naturais e meteorologia por meio de imagens de satélites; análise de imagens biomédicas; aplicações em automação industrial envolvendo o uso de sensores visuais em robôs, etc.

O processamento de imagens é a atividade de tratar imagens com a finalidade de melhorá-las ou apenas resolver algum problema de modo a permitir que o usuário possa ter acesso a algum tipo de informação que, a princípio, ele não tinha (ALBUQUERQUE e ALBUQUERQUE, 2000).

Para tanto vale ressaltar uns processos que didaticamente chama-se de aquisição de imagens, neste processo utiliza-se algum mecanismo para gerar as imagens que se deseja processar. As imagens podem ser obtidas tanto através de equipamentos de captura como câmeras, scanners e radares quanto através de simulações por computador. No caso da mamografia, especificamente, os filmes impressos tradicionais podem ser digitalizados por scanners especializados. Neste trabalho, pela dificuldade de acesso de uma base de dados pública e/ou privada com a finalidade da metodologia que será empregada, foi utilizada, para teste, a base de mamografias digitalizadas a partir de imagens radiográficas DDSM (Digital Database for Screening Mamography) disponível na internet (HEATH, et al. 1998).

O processamento tem a finalidade de aumentar a qualidade da imagem, eliminar objetos indesejáveis, tornando mais fácil a sua identificação. Esta etapa é importante porque aumenta a eficiência das etapas posteriores. Nela podemos utilizar diminuição de ruído, realce de contraste, filtros morfológicos, crescimento de região, dentre outras.

### **2.2.1. Representação Digital de uma Imagem e *Pixel*.**

Uma imagem pode ser definida como uma função bidimensional,  $f(x, y)$  onde  $x$  e  $y$  são coordenadas espaciais (plano), e a amplitude de  $f$  em qualquer par coordenadas  $(x, y)$  é chamada de *intensidade* ou *nível de cinza* da imagem nesse ponto. Quando  $x, y$  e os valores de intensidade de  $f$  são quantidades finitas e discretas, chama-se *imagem digital*. O campo do *processamento de imagens digitais* se refere ao processamento de imagens digitais por um computador digital. Observe que uma imagem digital é composta de um número finito de elementos, cada um com valor e localização específicos. Esses elementos são chamados de *elementos pictóricos*, *elementos de imagem*, *pels* e *pixels*. *Pixel* é o termo mais utilizado para representar os elementos de uma imagem digital.

### **2.2.2. Crescimento de Região**

O algoritmo de crescimento de regiões é um método de segmentação simples. Consiste em agregar conjuntos de pixels vizinhos em regiões maiores. O processamento parte de um elemento inicial, denominado semente, o qual pode ser tanto um único *pixel* como um conjunto de *pixels*, e realiza o crescimento da vizinhança agregando os *pixels* próximos que possuam atributos similares aos da semente. O processo continua até que se atinja uma condição de parada pré-estabelecida, como, por exemplo, um determinado nível de cinza ou uma distância específica (PAL e PAL. 1993).

Na prática, no entanto, algumas dificuldades, razoavelmente complexas, devem ser levadas em conta durante a definição do padrão de crescimento para que resultados aceitáveis sejam obtidos, como, por exemplo, a seleção da semente, o estabelecimento das condições de semelhança e a determinação das condições de parada. Essas dificuldades, em geral, exigem que se tenha certo conhecimento sobre a imagem que se deseja segmentar.

Neste trabalho o algoritmo de crescimento de regiões foi utilizado na etapa de pré-processamento, para remover o fundo da imagem e os rótulos de marcação.



### 2.3. Registro de Imagens

O registro de imagens tem sido de grande importância em diversas áreas, em especial na área médica. Ele tem sido empregado em segmentação, diagnóstico e também na análise temporal de doenças.

A literatura emprega diversas definições ao registro de imagens. Brown (1992) define o registro de imagens como o processo de transformação de diferentes conjuntos de dados para um mesmo sistema de coordenadas. Os dados podem ser várias fotografias, dados de diferentes sensores, de diferentes épocas, ou a partir de diferentes pontos de vista.

Crum (2004) por sua vez, explica o registro de imagens como um processo para a determinação da correspondência entre as características de imagens recolhidas em momentos diferentes ou utilizando diferentes métodos de imagem. As correspondências podem ser usadas para alterar a aparência - por rotação, tradução, alongamento etc. - de uma imagem, de forma que mais se assemelhe a outra a fim de que o par possa ser diretamente comparado, combinado ou analisado. O uso mais intuitivo de registro é para corrigir posições diferentes de pacientes entre as varreduras. Registro de imagem não tem uma finalidade específica agregar valor às imagens, por exemplo, permitindo que imagens estruturais (TC, RM, ultra-som) e funcional (PET, SPECT, ressonância magnética funcional (fMRI)) possam ser vistas e analisadas no mesmo sistema de coordenadas, e facilita novas utilizações de imagens, por exemplo, para monitorar e quantificar a progressão da doença ao longo do tempo em indivíduo ou para construir modelos estatísticos de variação estrutural em uma população.

Historicamente, os registros dividem-se em dois grupos: registros rígidos e registros deformáveis (ou não rígidos). Os registros rígidos consideram que a imagem móvel é um objeto que precisa ser rotacionado e/ou transladado para que haja uma correspondência satisfatória com a imagem fixa. Ou seja, esse tipo de registro usa transformação global. Os registros deformáveis, por outro lado, obtêm essa correspondência através de algumas deformações localizadas. Esse por sua vez usa transformação local.

O registro de imagens é composto pelos seguintes elementos: duas imagens, sendo uma fixa e um móvel, uma medida de similaridade, uma função de transformação, de um otimizador e um interpolador.

A medida de similaridade é uma medida numérica responsável por dizer o quanto a imagem móvel está similar a fixa. A função de transformação é a função que mapeia a imagem móvel na imagem fixa. Essa função de transformação é calculada pelo otimizador, otimizando certa medida de similaridade. Dessa forma, o registro pode ser visto como um problema de otimização, no qual um otimizador estima a função transformação que melhor mapeia a imagem fixa da imagem móvel, de acordo com um valor obtido pela medida de similaridade. Esse processo pode ser visto na Figura 3.

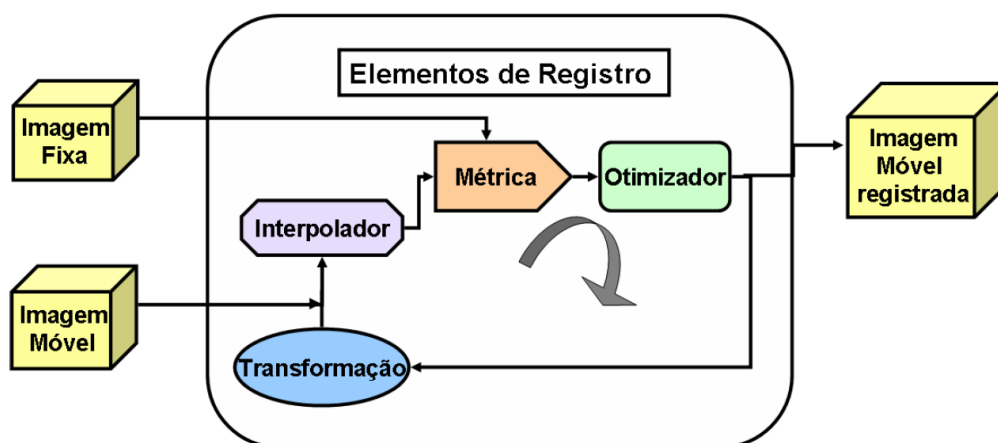


Figura 3: Processo de um registro de imagens. Fonte: (Parraga, 2008).

O registro de imagens médicas, por sua vez, é uma técnica de processamento de imagens que pode ser definido como processo de alinhamento de imagens para que características correspondentes sejam facilmente relacionadas (HAJNAL, 2001), sendo que, estas imagens podem ser adquiridas por sensores diferentes ou em instantes diferentes.

O registro é um processo essencial em várias áreas de aplicação ligadas a imagens médicas. Uma utilização bastante comum é a realização de comparação e análise de imagens de pacientes adquiridas em momentos diferentes (análise temporal). Nesses casos, o registro auxilia o especialista no acompanhamento da evolução da patologia do paciente. Por exemplo, acompanhar o crescimento de um tumor ou verificar imagens adquiridas antes e após a utilização de contraste.

Outras aplicações muito comuns são o procedimento cirúrgico guiado por imagem (GOSHTASBY, 2005), no qual o registro de imagens pré-operatórias e imagens obtidas durante a cirurgia é de fundamental importância para a localização da patologia e a junção de tipos diferentes de exames permitindo agrupar informações complementares em uma mesma imagem.

### 2.3.1. Registro Rígido

O registro rígido é o tipo de registro mais simples. A sua aplicação resume-se apenas a aplicações de operações de rotação e translação, de forma que a imagem móvel não é deformada, e sim posta no mesmo sistema de coordenadas da imagem fixa.

O registro rígido, por si só, não é o mais adequado para a maioria dos problemas de desalinhamento de imagens. Isso ocorre porque ele não consegue corrigir erros locais que apenas seriam resolvidos por meio de deformações. No entanto, ele é comumente usado como pré-processamento de registros deformáveis, possibilitando que os mesmos sejam executados com maior eficiência e menor custo computacional. Isso acontece porque, uma vez que a imagem móvel está no mesmo sistema de coordenadas da fixa, as deformações passam a ser mais locais e menos frequentes.

Mais formalmente, a transformação de coordenadas de imagem móvel,  $M(x)$ , nas coordenadas da imagem fixa  $F(x)$  é dada pela Equação 1.

$$x' = T(x) = R_x(\Theta_x)R_y(\Theta_y)R_z(\Theta_z)x + t \quad (1)$$

onde  $R_x$ ,  $R_y$ , e  $R_z$  são as matrizes de rotação em torno dos eixos  $x$ ,  $y$  e  $z$ , respectivamente, considerando um ângulo  $\Theta$  e  $t$  é o vetor translação. As matrizes de rotação são representadas pelas Equações 2, 3 e 4.

$$R_x(\Theta_x) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \Theta_x & -\sin \Theta_x \\ 0 & \sin \Theta_x & \cos \Theta_x \end{pmatrix} \quad (2)$$

$$R_y(\Theta_y) = \begin{pmatrix} \cos \Theta_y & 0 & \sin \Theta_y \\ 0 & 1 & 0 \\ -\sin \Theta_y & 0 & \cos \Theta_y \end{pmatrix} \quad (3)$$

$$R_z(\Theta_z) = \begin{pmatrix} \cos \Theta_z & -\sin \Theta_z & 0 \\ \sin \Theta_z & \cos \Theta_z & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (4)$$

### 2.3.2. Registro Deformável *B-spline*

O Registro *B-spline* é um dos tipos de registro deformáveis mais utilizados. O seu modelo de transformação é descrito por uma *B-spline* de terceira ordem. É um registro do tipo paramétrico e retorna os parâmetros do modelo representando o campo de deformação.

A transformação de imagem móvel em um ponto qualquer  $x = [x, y, z]/t$  num registro *B-spline* pode ser representada pela Equação 5.

$$T(x, y, z) = \sum_{l=0}^3 \sum_{m=0}^3 \sum_{n=0}^3 \beta_l(u) \beta_m(v) \beta_n(w) \Phi_{i+l, j+m, k+n} \quad (5)$$

onde as funções  $\beta$  são polinômios *B-spline* de terceira ordem, representadas pelas Equações 6,7,8 e 9:

$$\beta_0(t) = (-t^3 + 3t^2 - 3t + 1)/6 \quad (6)$$

$$\beta_1(t) = (3t^3 - 6t^2 + 4)/6 \quad (7)$$

$$\beta_2(t) = (-3t^3 + 3t^2 + 3t + 1)/6 \quad (8)$$

$$\beta_3(t) = t^3/6 \quad (9)$$

O parâmetro  $\Phi_{i,j,k}$  é um conjunto de pontos de controle. Esses pontos são coeficientes que representam o campo de deformação. Eles formam uma malha de pontos de tamanho  $n_x, n_y, n_z$ , que são distribuídos sobre a imagem móvel, com espaçamentos  $\delta_x, \delta_y, \delta_z$ .

Para uma *B-spline* de terceira ordem, quatro funções *B-spline* são necessárias para um determinado ponto  $x$ . Como o registro é tridimensional, leva-se em consideração coordenadas  $x, y$  e  $z$ , o que implica que 43 *B-splines* são importantes para cada ponto. Dessa forma, a transformação *B-spline* é aplicada para cada *voxel* da imagem móvel nas posições em torno dos  $4x4x4$  pontos de controle da malha  $\Phi_{i,j,k}$ . Os índices  $i, j$  e  $k$  representam os índices dos pontos de controle da malha. São eles vistos nas Equações 10, 11 e 12.

$$i = \left\lfloor \frac{x}{\delta_x} \right\rfloor - 1 \quad (10)$$

$$j = \left\lfloor \frac{y}{\delta_y} \right\rfloor - 1 \quad (11)$$

$$k = \left\lfloor \frac{z}{\delta_z} \right\rfloor - 1 \quad (12)$$

onde  $u, v$  e  $w$  são as posições relativas de  $x, y$  e  $z$  dos pontos da malha no espaço 3D. São eles vistos nas Equações 13, 14 e 15.

$$u = \frac{x}{\delta_x} - \left\lfloor \frac{x}{\delta_x} \right\rfloor \quad (13)$$

$$v = \frac{y}{\delta_y} - \left\lfloor \frac{y}{\delta_y} \right\rfloor \quad (14)$$

$$w = \frac{z}{\delta_z} - \left\lfloor \frac{z}{\delta_z} \right\rfloor \quad (15)$$

O registro *B-spline* é um problema de otimização, que busca encontrar os parâmetros  $\Phi_{i,j,k}$  que maximizam a semelhança entre as imagens, ou seja, maximizar a medida de similaridade em busca dos parâmetros que melhor representam a transformação  $T(x, y, z)$ . O processo do registro é representado resumidamente pelas seguintes etapas:

1. Escolher o tamanho da malha de pontos de controle (quantidade de pontos);
2. Inicializar os valores do conjunto de parâmetros  $\Phi_{i,j,k}$ ;
3. Calcular a medida de similaridade para  $\Phi_{i,j,k}$ ;
4. Atualizar o conjunto de parâmetros  $\Phi_{i,j,k}$ ;
5. Aplicar a transformação  $T(x, y, z)$ ;
6. Repetir o passo 3 até algum critério de parada ser satisfeito.

Uma característica importante do registro *B-spline* é que suas deformações são locais, o que impede alterações desnecessárias em outras estruturas. Isso ocorre porque alterações nos pontos de controle afetam somente a vizinhança local da malha. No entanto, algumas consequências sobre a parametrização da malha devem ser levadas em consideração. Uma malha muito grande (ou seja, com muitos pontos de controle) define deformações mais locais na imagem, uma vez que o espaçamento entre os pontos será menor, mas uma malha menor (menos pontos de controle) resulta em deformações mais globais.

Na metodologia, os registros rígidos e *B-spline* são usados em conjunto em uma mamografia no tempo  $t + 1$  (mais atual), para gerarmos uma imagem correspondente a imagem no tempo  $t$  (mais antiga).

## 2.4. Índices de Similaridade

Os índices de similaridade são comumente usados em biologia para fazer comparações entre grupos de espécies. Na biologia, muitas vezes é desejável fazer comparações entre amostras de fauna ou florais tomadas em momentos diferentes, em lugares diferentes, ou por diferentes técnicas. Essas comparações são feitas baseando-se nas características individuais e a abundância de indivíduos das amostras. A necessidade de quantificar essas comparações tornou popular o uso dos índices de similaridade. Apesar da popularização, um grande número de índices, utilizados atualmente, foram criados no século passado (MEYER, 2002).

Alguns destes índices apenas levam em consideração a presença ou ausência de espécies das amostras, enquanto outros incorporam informação sobre a abundância relativa da espécie. O índice preferível em um determinado caso depende das situações e dos tipos de dados disponíveis. Huhta (1979) testou uma longa série de índices de similaridade com dados reais e descobriu que os resultados dependem em grande parte do índice escolhido. Isso sugere que em muitos casos há um perigo de se escolher um índice e este não ser capaz de provar o que se quer demonstrar. Assim, uma avaliação cuidadosa de os vários índices é essencial.

Além dos índices de similaridade, também existem os índices de diversidade. Conceitualmente, a diferença entre os dois é que o primeiro busca medir o grau de semelhança entre grupos e o segundo busca medir o grau de diferença. Matematicamente, um índice de similaridade apresenta alto valor quando há uma forte semelhança entre os grupos, enquanto um índice de diversidade apresenta alto valor quando há baixa semelhança entre os grupos. Apesar de buscarem resultados distintos, geralmente é fácil transformar um índice de similaridade em um de diversidade e vice-versa. Um índice de similaridade  $s$  com domínio  $[0, 1]$ , por exemplo, pode ser transformado em um índice de diversidade  $d$  através da operação complementar  $d = (1 - s)$ .

São muitos os índices de similaridade encontrados na literatura. Nesta monografia será utilizado um número de índices considerados apropriados para o problema, descritos em Meyer (2002).

Considerando dois grupos, chamados grupo 1 e grupo 2, no qual deseja-se medir a similaridade, descreve-se as seguintes variáveis:

- $a$ : número de indivíduos que pertencem tanto ao grupo 1 quanto grupo 2;
- $b$ : número de indivíduos que pertencem grupo 1 e não pertencem ao grupo 2;
- $c$ : número de indivíduos que pertencem grupo 2 e não pertencem ao grupo 1.

Os índices usados são descritos na Tabela 1, juntamente com suas equações e intervalo de domínio.

Tabela 1: Índices de similaridade. Adaptado de (Meyer, 2002).

Índice de similaridade	Equação	Intervalo de Ocorrência
Jaccard	$\frac{a}{a + b + c}$	[0, 1]
Anderberg	$\frac{a}{a + 2(b + c)}$	[0, 1]
Czekanowsky	$\frac{2a}{2a + b + c}$	[0, 1]
Kulczynski 2	$\frac{a}{2} \left( \frac{1}{a + b} + \frac{1}{a + c} \right)$	[0, 1]
Ochiai	$\frac{a}{\sqrt{(a + b)(a + c)}}$	[0, 1]

Na metodologia, os índices de similaridade serão usados na análise temporal de lesões baseado em textura. Através dele será possível definir se há semelhanças ou diferenças nos padrões de textura das lesões no tempo.

## 2.5. *Insight Segmentation and Registration Toolkit (ITK)*

O *Insight Segmentation and Registration Toolkit* (ou ITK) é um *framework* de desenvolvimento multiplataforma de aplicações de código aberto amplamente utilizado para o desenvolvimento de programas de segmentação de imagem e de registro de imagens. Ele também é definido como um kit de ferramentas flexível e multiplataforma, ou seja, funciona muito bem independente do sistema operacional do computador.

O ITK foi desenvolvido com financiamento da *National Library of Medicine* como um recurso aberto de algoritmos para analisar as imagens do *Visible Human Project*.

O ITK é um projeto de código aberto, ou seja, os desenvolvedores de todo o mundo podem usar, depurar, manter e estender o software. ITK usa um modelo de desenvolvimento de software de programação extrema. As principais características da programação extrema são comunicação e testes. A comunicação entre os membros da comunidade ITK é o que ajuda a gerenciar a rápida evolução do software. O teste é o que mantém a estabilidade do software.

Ele fornece algoritmos de segmentação e de registro em duas, três e mais dimensões. Os algoritmos providos por ele são os mais diversos, tanto na área de processamento de imagens

como reconhecimento de padrões. O ITK também é reconhecido por ser utilizado com imagens médicas.

No caso desta monografia, foi utilizado o ITK implementado em C++, embora também seja compatível a outras linguagens com Java, Python e Tcl. Além disso, o ITK tem seu código-fonte aberto e gratuito, e é constantemente utilizado, testado e modificado pela comunidade online em favor de melhorias e detecção de *bugs*, o que garante a sua qualidade, em especial em projetos médicos. O ITK é essencial pela disponibilidade de suas ferramentas de processamento de imagens e reconhecimento de padrões. Os algoritmos utilizados nesse trabalho, como o de registro, foram desenvolvidos com auxílio do ITK.



### 3. MATERIAIS E MÉTODOS

Nesse capítulo, é descrito a metodologia proposta. Detalhes de como as técnicas e recursos apresentados no Capítulo 2 foram empregados nas etapas da metodologia. Para tanto, existem algumas etapas para alcançar a metodologia abordada, como ilustrado na Figura 4.

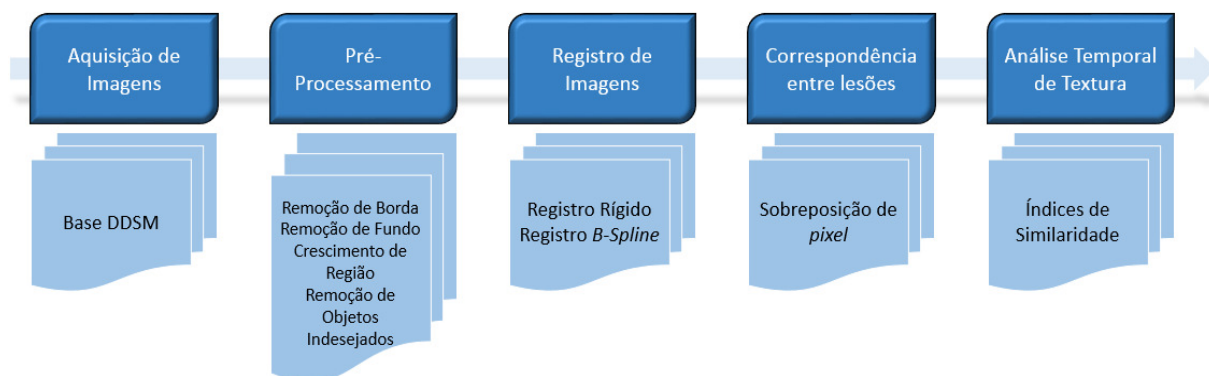


Figura 4: Fluxo da metodologia.

Será descrito, nas próximas seções, cada uma dessas etapas detalhadamente, mostrando as sub-etapas, os problemas encontrados e as pesquisas realizadas durante o trabalho para o alcance final dessa metodologia.

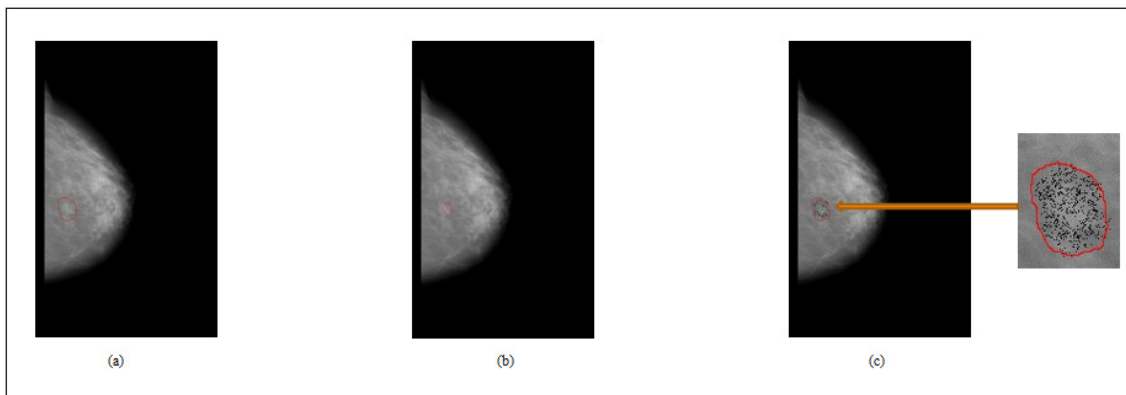
#### 3.1. Aquisição de Imagens

Até o fim deste trabalho, a etapa de maior dificuldade havia sido esta, pelo fato de, até então, não ter sido encontrado nenhuma base mamográfica – seja essa pública ou mesmo privada – de análise temporal, isto é, uma base em que o mesmo paciente tivesse mamografias realizadas no dia  $X$  e outra realizada depois de certo tempo, no dia  $Y$ . Portanto, foi utilizado neste trabalho, no intuito de testes e validações da proposta apresentada, a base de imagens fornecidas pelo DDSM - Digital Database for Screening Mammography (DDSM, 2001).

A base DDSM, é um banco de dados público contendo 2.620 casos, fornecidas gratuitamente na internet (HEATH, et al. 1998), através dos esforços de algumas instituições americanas (Massachusetts General Hospital, Wake Forest University, e Washington University in St. Louis School of Medicine). Cada caso contém duas imagens de cada mama (projeções CC e MLO), além de informações extras sobre o exame (data do estudo, idade do paciente, tipo da patologia, quantidade de anomalias, etc.) e sobre a imagem (nome do arquivo, tipo de filme, data de digitalização, tipo do digitalizador, sequência, *pixels* por linha, bits por

*pixel*, marcação, etc.). Todas as informações contidas no DDSM foram fornecidas por especialistas (HEATH, et al. 1998).

Outra característica dessa base é a forma como essas informações são fornecidas: cada imagem é associada a um arquivo de marcação, contendo informações relevantes sobre o paciente. Sabendo disso, foram feitas alterações nas imagens da base DDSM para simular mudanças de tamanho e textura que podem ocorrer na lesão no tempo, isto é, considerando uma imagem  $t$  já marcada pelo especialista de um paciente  $X$  (Figura 5a), será feita mudanças nesta imagem  $t$  para simular uma mamografia tirada em um tempo diferente. Neste trabalho, a imagem foi rotacionada e o tamanho da marcação alterado (Figura 5b) e gerado ruídos na lesão (Figura 5c) para validar a metodologia proposta. O processo de rotação da imagem, tem como finalidade simular alterações na aquisição da imagem, seja pelo padrão utilizado do exame, seja pelo processo de digitalização. Um exemplo das alterações da base DDSM pode ser visto na Figura 5.



**Figura 5: Aquisição de imagem (a) Imagem com marcação do especialista (b) Imagem rotacionada e marcação alterada (c) Imagem com ruído na lesão**

Após a aquisição da imagem, a mamografia digitalizada precisa ser preparada para a etapa de registro, para assim conseguir a correspondência entre as lesões em tempos distintos. Então, é feito um processo de pré-processamento no intuito de preparar essa imagem para as demais etapas.

### 3.2. Pré-processamento das Imagens

Imagens pertencentes à base de imagens DDSM apresentam ruídos e elementos típicos do exame mamográfico. A região de interesse nesta metodologia é a mama. A mamografia, no entanto, contém outros objetos desnecessários e que podem atrapalhar a análise, tais como o *background*, anotações do especialista, ruídos, dentre outros. O objetivo, então, dessa etapa é eliminar esses componentes indesejados da imagem.

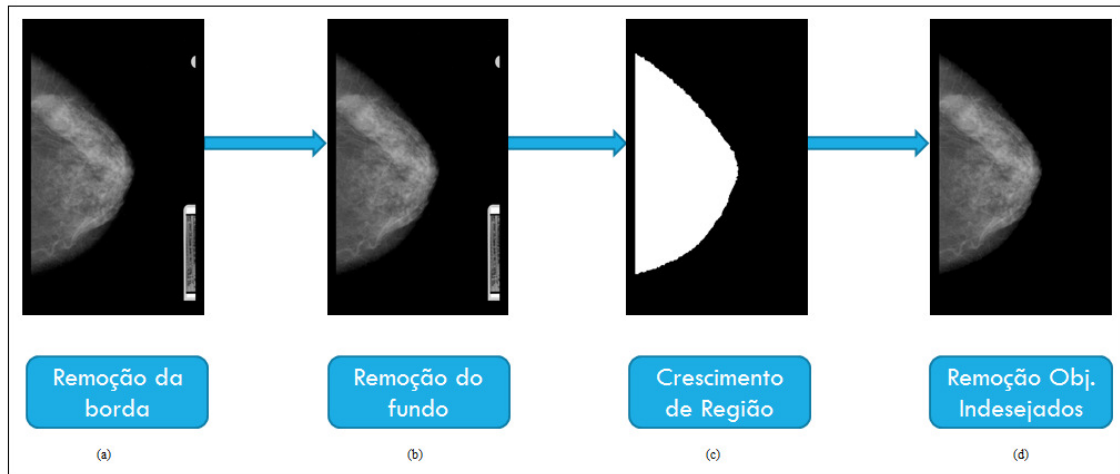
Esse pré-processamento é essencialmente importante para a etapa do registro de imagens. Como já explicado na Seção 2.3, o registro de imagens busca aproximar a forma e localização dos componentes de uma imagem móvel a uma imagem fixa. Quando há componentes indesejados na imagem, essa aproximação pode deformar ou localizar erroneamente a região de interesse para manter a consistência desses outros componentes.

A etapa de pré-processamento consiste primeiramente na redução do tamanho da imagem original. Resultando numa imagem com a altura igual a 1024 *pixels*, reduzindo-se proporcionalmente a largura. Esta redução se justifica para reduzir o tempo de processamento que se tornaria muito elevado quando se trabalha com a imagem no seu tamanho original.

Após a redução de tamanho, todos os pontos a 30 *pixels* de distância das bordas laterais são removidos (Figura 6a). Essa remoção serve para eliminar da imagem, a área entre as bordas do filme da radiografia e os espaços vazios sem o filme.

Então é feita a remoção do fundo (Figura 6b), nesta etapa os valores de *pixel* são divididos em dois grupos de acordo com a intensidade, os de maior intensidade (como os da mama e das marcas de identificação) e os de menor intensidade (como o fundo e os ruídos mais escuros). Com o intuito de remover o fundo o grupo de menor intensidade, é atribuído a ele valores de pixel preto.

Por fim, na intenção de remover os objetos externos, é utilizado o algoritmo de crescimento de região (Figura 6c) que separa da imagem as regiões não conexas, selecionando assim apenas a região de maior área e de nosso interesse. A semente do crescimento de região é posicionada no centro da metade da imagem cujo somatório de valores de *pixel* é mais alto, ou seja, na metade em que se encontra a mama (Figura 6d).



**Figura 6: Pré-processamento de imagem (a) Remoção de borda (b) Remoção do fundo (c) Crescimento de Região (d) Remoção de objetos indesejados**

### 3.3. Registro das Imagens

Assim como a etapa anterior, esta tem um papel importante para a correspondência entre as lesões. A partir desse ponto, a imagem que havia sido simulada no tempo  $t+1$  ou tempo mais atual (mamografia rotacionada e com alteração na marcação ou com ruídos na lesão, pode ser visto na Figura 7b), sofrerá um processo de transformação de diferentes conjuntos de dados para um mesmo sistema de coordenadas, correspondente a imagem no tempo  $t$ , ou tempo mais antigo (Figura 7a). Gerando uma imagem no tempo  $t+1$  registrada (Figura 7c). Esse processo de transformação serve para gerar uma correspondência entre as características recolhidas em tempos diferentes.

Essa etapa é necessária pelo fato de que a lesão pode não estar na mesma localização nos exames de tempos diferentes. Isso pode ocorrer por diversos fatores, tais como: posicionamento diferente, respiração, mudança na estrutura da lesão e diferenças no processo de digitalização das mamografias. Os algoritmos de registros utilizados nessa etapa, são os descritos na Seção 2.3 e foram adaptados de DINIZ (2014).

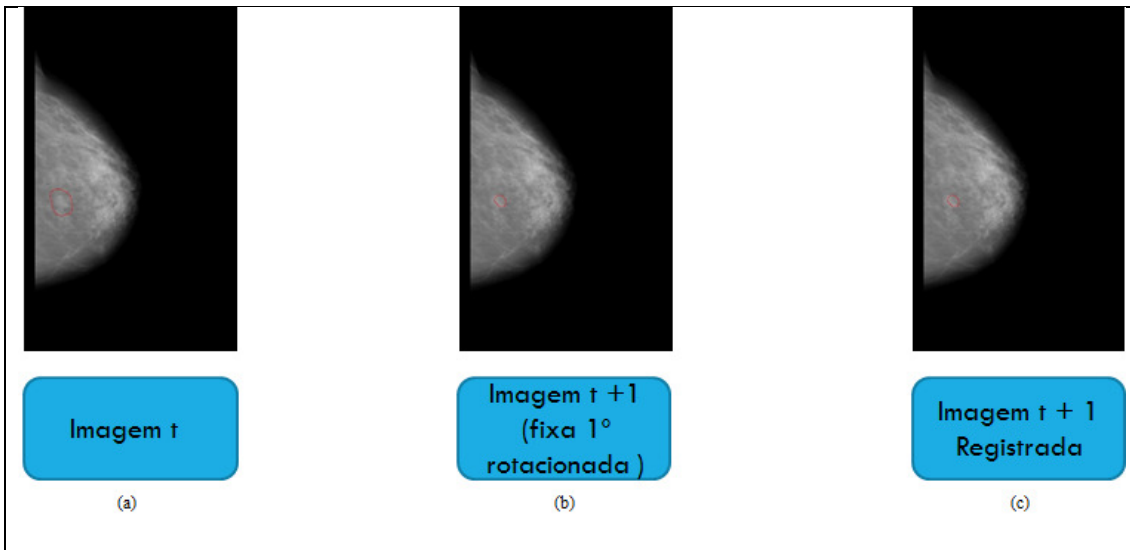


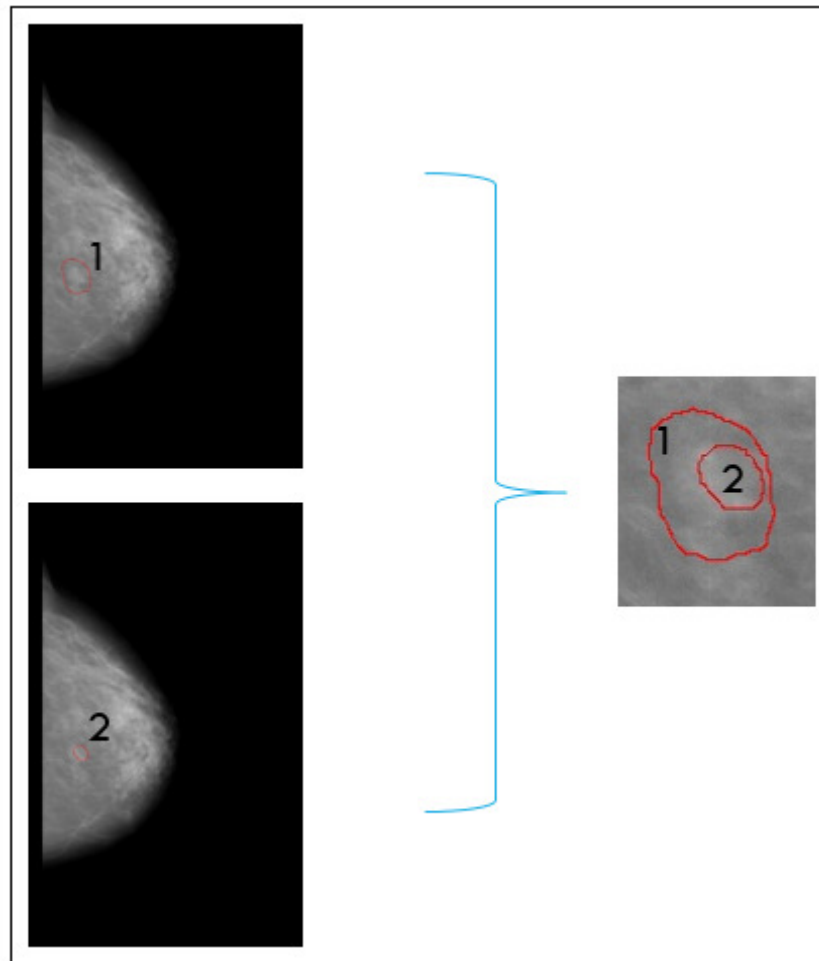
Figura 7: Registro de Imagens (a) Imagem tempo  $t$  (b) Imagem no tempo  $t+1$  (c) Imagem no tempo  $t+1$  registrada

A partir desse ponto, as imagens já estão no mesmo sistema de coordenadas e com correspondência de suas características. Agora precisa-se verificar se a lesão em tempos distintos é de fato a mesma lesão, isto é, se há correspondência entre as lesões.

### 3.4. Correspondência entre Lesões

Esta etapa é responsável por determinar a correspondência entre as lesões localizadas em imagens diferentes, ou seja, saber se a lesão em tempos distintos é realmente a mesma lesão. Basicamente, depois de pré-processada e registrada a imagem no tempo  $t+1$  com a imagem no tempo  $t$ , existem duas imagens do mesmo paciente, com marcações distintas, então surge o impasse de saber se aquela marcação no tempo mais atual é a mesma que a do tempo antigo.

Como visto na etapa de Seção 3.1, as imagens possuem suas marcações, então é necessário determinar se ao menos um *pixel* dentro dessas duas marcações coincidem em localização, ou seja, se as marcações estão sobrepostas espacialmente. Caso isso ocorra, considera-se que as lesões em tempos diferentes são, de fato, a mesma lesão, como mostra a Figura 8.



**Figura 8: Correspondência entre lesões**

É dado muito valor a esta etapa, pois é crucial saber a relação entre as lesões já que o objetivo deste trabalho é fornecer informações temporais da lesão do paciente em tempos distintos para dá um auxílio ao especialista sobre como se comportou essa lesão no tempo subsequente ao primeiro exame. Dessa forma, precisa-se identificar a existência e correspondência desta mesma lesão.

Agora já existe uma correspondência entre as lesões em tempos distintos, portanto, já é possível obter características a partir da análise da textura das lesões.

### **3.5. Análise Temporal de Textura**

Esta etapa é considerada a mais importante desse trabalho. A etapa da análise temporal é responsável por analisar mudanças de tecido (textura) em lesões ocorridas no decorrer do tempo. Ela será aplicada nas duas marcações, uma obtida no exame mais antigo (tempo  $t$ ) e

outra obtida no exame mais atual (tempo  $t+1$ ), isso depois de todas as etapas para comprovar que, de fato, essas duas marcações correspondem a mesma lesão no tempo.

As medidas de mudanças de textura são obtidas através da análise do padrão de textura da lesão que é diretamente relacionada aos seus valores de *pixel*. Nessa etapa, são calculados índices de similaridade (Seção 2.4). Os índices de similaridade utilizados neste trabalho são os descritos em (Meyer, 2002). No trabalho em questão, eles serão utilizados para comparar a similaridade de texturas da lesão em exames diferentes. Considerando as fórmulas dos índices de similaridade da Tabela 1, as variáveis a, b e c serão redefinidas como:

- a: quantidade de *pixels* que pertencem tanto a lesão no exame no tempo  $t$  quanto a lesão no exame no tempo  $t+1$ ;
- b: quantidade de *pixels* que pertencem a lesão no exame no tempo  $t$ , mas não a lesão no exame no tempo  $t+1$ ;
- c: quantidade de *pixels* que pertencem a lesão no exame no tempo  $t+1$ , mas não a lesão no exame no tempo  $t$ ;

O resultado da análise temporal, pode ser usado para auxiliar o especialista a avaliar as mudanças da lesão a partir do comportamento dos valores das equações, dando a ele, um suporte a mais para identificar a malignidade desta lesão ou até se o tratamento aplicado aquele paciente está sendo efetivo, validando ou não tal tratamento.

## **4. TESTES E RESULTADOS**

Neste capítulo é apresentado o resultado obtido com a utilização da metodologia abordada no trabalho para a análise temporal de lesões em mamografias digitalizadas. Os resultados estão apresentados conforme os testes executados na base DDSM, que, assim como descrito na Seção 3.1, foram simulados casos temporais para validação da metodologia proposta.

A base DDSM é dividida em densidade de lesão (mamas densas e não-densas), que por sua vez é subdividida em termo de malignidade de lesão (maligna e benigna), então optou-se por aplicar os testes em um desses tipos de caso. Assim, são apresentados em duas seções de Mamas Densa e Não-Densa, e cada seção apresenta os testes feitos e as anotações sobre cada um.

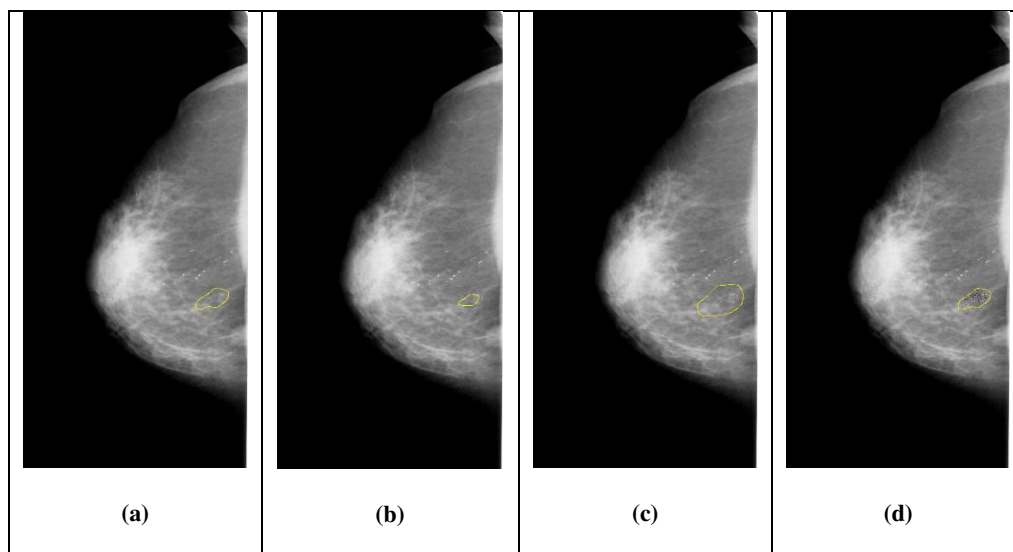
### **4.1. Mamas Densas**

Será apresentado nesta seção os testes realizados em Mamas Densas. Elas são divididas em mamas com lesões benignas e lesões malignas. São feitas simulações de lesões em mamografias digitalizadas de dois pacientes distintos em cada tipo de lesão.

#### **4.1.1. Lesões Benignas**

Foram realizados testes na imagem **A\_1123\_1.LEFT\_CC** da base DDSM, que é uma imagem de uma mama densa com uma lesão benigna. Modificações foram feitas no tamanho da marcação da lesão e ruídos foram gerados para validação dos testes. Como segue a Figura 9.





**Figura 9:** Imagem A\_1123\_1.LEFT\_CC da DDSM (a) Marcação do especialista (b) Marcação Reduzida (c) Marcação Aumentada (d) Ruídos na Lesão

No teste realizado com a marcação do especialista (Figura 9a) e a marcação diminuída (Figura 9b), os índices tendem a ficar baixos como mostra a Tabela 2. Isso acontece pelo fato que, quando as lesões são marcadas, geralmente, o especialista engloba parte da mama dentro da marcação, e ao compará-la com uma região mais interna, que compreende apenas *pixels* da lesão em si, os índices tendem a diminuir.

**Tabela 2:** Teste na imagem A\_1123\_1.LEFT\_CC. Resultado dos índices comparados a marcação original e a marcação reduzida

Índices	Valores
Jaccard	0.491525
Anderberg	0.325843
Czekanowsky	0.659091
Kulczynski 2	0.733333
Ochai	0.695222

Testando a marcação do especialista (Figura 9a) e a marcação aumentada (Figura 9c), percebemos que os índices tendem a se aproximar de 1, como exposto na Tabela 3. Análogo ao teste anterior, ao aumentar a marcação, considera-se parte da mama que não é lesão. Como o especialista ao marcar também engloba parte da mama que não é lesão, logo, na comparação, há um aumento de valores dos índices.

**Tabela 3: Teste na imagem A\_1123\_1.LEFT\_CC. Resultado dos índices comparados a marcação original e marcação aumentada**

Índices	Valores
Jaccard	0.73913
Anderberg	0.586207
Czekanowsky	0.85
Kulczynski 2	0.85094
Ochai	0.850473

Como último teste, foi analisado a marcação real com a mesma marcação, porém com ruídos na lesão. Aplicando a equação dos índices, podemos ver que estes tendem a variar muito quando são colocados *pixels* novos na marcação, sendo estes presentes apenas na Figura 9d (imagem simulando tempo  $t + 1$ ) e ausentes na Figura 9a (simulando tempo  $t$ ). Como pode ser visto na Tabela 4.

**Tabela 4: Teste na imagem A\_1123\_1.LEFT\_CC. Resultado dos índices comparados a marcação original e marcação com ruídos**

Índices	Valores
Jaccard	0.392857
Anderberg	0.24444
Czekanowsky	0.564103
Kulczynski 2	0.564144
Ochai	0.564123

Mais uma série de testes foi aplicado às mamas densas com lesão benignas. Aplicou-se na imagem **A\_1242\_1.LEFT\_MLO** da base DDSM as modificações no tamanho da marcação da lesão, e ruídos foram geradas para validação dos testes. Como segue a Figura 10.

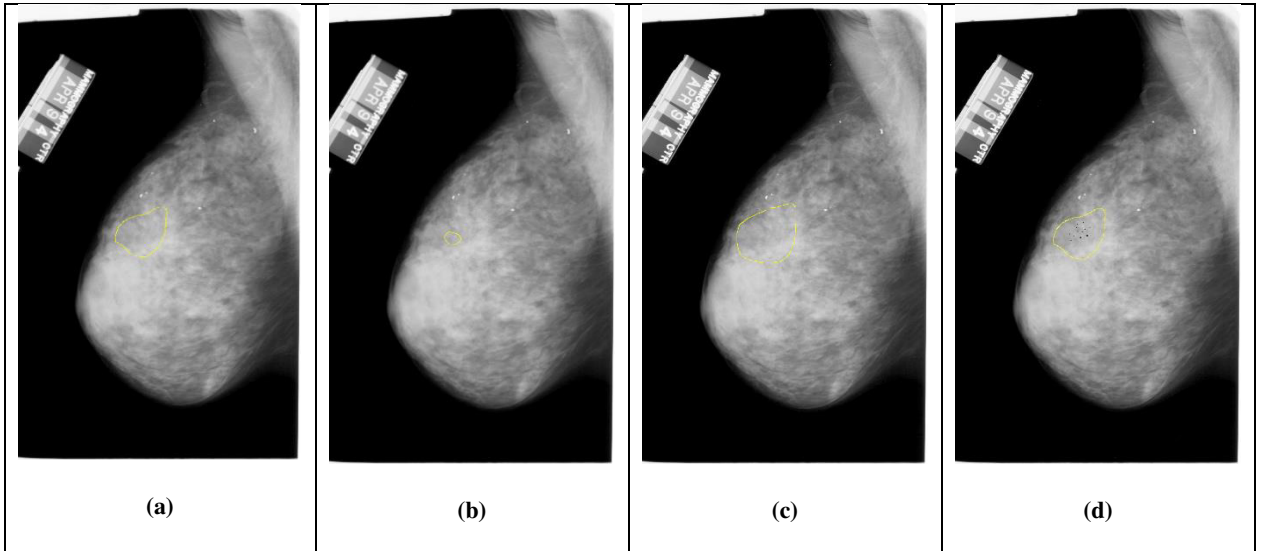


Figura 10: Imagem A\_1242\_1.LEFT\_MLO da DDSM (a) Marcação do especialista (b) Marcação Reduzida (c) Marcação Aumentada (d) Ruídos na Lesão

No teste realizado com a marcação do especialista (Figura 10a) e a marcação diminuída (Figura 10b), os índices tendem a ficar baixos como mostra a Tabela 4. Assim como mostrado no teste realizado na imagem A\_1123\_1.LEFT\_CC, os índices se comportaram semelhante quando houve a redução da marcação.

Tabela 5: Teste na imagem A\_1242\_1.LEFT\_MLO. Resultado dos índices comparados a marcação original e a marcação reduzida

Índices	Valores
Jaccard	0.402597
Anderberg	0.252033
Czekanowsky	0.574074
Kulczynski 2	0.701299
Ochai	0.634506

Considerando a marcação do especialista (Figura 10a) e a marcação aumentada (Figura 10c), percebemos que os índices tendem a ser mais similares, isto é, próximo de 1, como exposto na Tabela 6. Assim, podemos ver que o comportamento do índice quando há um aumento da marcação geralmente mostrará mais similaridade.

**Tabela 6: Teste na imagem A\_1242\_1.LEFT\_MLO. Resultado dos índices comparados a marcação original e a marcação aumentada**

Índices	Valores
Jaccard	0.721154
Anderberg	0.56391
Czekanowsky	0.837989
Kulczynski 2	0.85466
Ochai	0.837989

Por fim, foi analisado a marcação real (Figura 10a) com a mesma marcação, porém com ruídos na lesão (Figura 10d). Aplicando a equação dos índices, que varia considerando ausência e presença de *pixel* dentro das marcações, foram gerados poucos ruídos, logo os valores dos índices não variaram tanto, já que apenas poucos *pixels* são diferentes da marcação do especialista. Esses resultados podem ser observados na Tabela 7.

**Tabela 7: Teste na imagem A\_1242\_1.LEFT\_MLO. Resultado dos índices comparados a marcação original e marcação com ruídos**

Índices	Valores
Jaccard	0.949367
Anderberg	0.903614
Czekanowsky	0.974026
Kulczynski 2	0.974026
Ochai	0.974026

#### 4.1.2. Lesões Malignas

Também foram aplicados testes nas imagens de mamas densas com lesões malignas. A mamografia digitalizada A\_1000\_1.LEFT\_CC da base DDSM, foi testada com modificações no tamanho da marcação da lesão e ruídos também foram gerados para validação dos testes. Como apresentado na Figura 11.

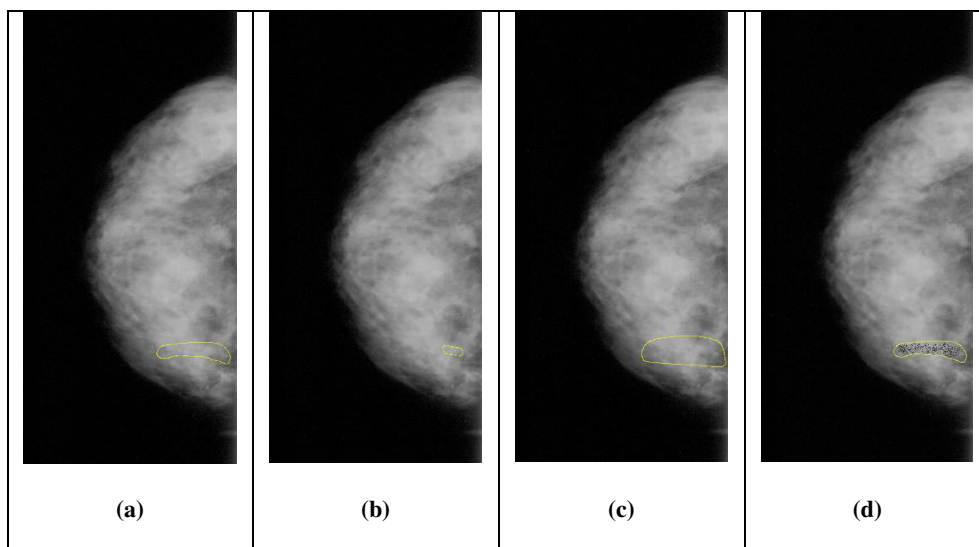


Figura 11: Imagem A\_1000\_1.LEFT\_CC da DDSM (a) Marcação do especialista (b) Marcação Reduzida (c) Marcação Aumentada (d) Ruídos na Lesão

Analogamente aos testes realizados nas mamas densas e benignas, comparando a marcação do especialista (Figura 11a) e a marcação diminuída (Figura 11b), os índices tenderam novamente a ficar baixos, isto é, pouco similares, como mostra a Tabela 8. Novamente ocorre pelo fato que, na marcação reduzida, pegou-se *pixels*, provavelmente, pertencentes apenas a lesão enquanto a marcação do especialista abrange um pouco da mama.

Tabela 8: Teste na imagem A\_1000\_1.LEFT\_CC. Resultado dos índices comparados a marcação original e a marcação reduzida

Índices	Valores
Jaccard	0.523256
Anderberg	0.354331
Czekanowsky	0.687023
Kulczynski 2	0.687023
Ochai	0.723364

Em um segundo teste realizado com esta mama, foi considerado a marcação do especialista (Figura 11a) e uma marcação maior (Figura 11c), podemos ver que, assim como nos testes realizados nas mamas densas e benignas, quando há um aumento da marcação, os índices tendem aumentar o valor, se aproximando de 1. Mais uma vez, isso ocorre pelo fato de que ambas marcações ocupam parte da mama e da lesão em si. A Tabela 9 mostra os resultados do cálculo dos índices.

**Tabela 9: Teste na imagem A\_1000\_1.LEFT\_CC. Resultado dos índices comparados a marcação original e a marcação aumentada**

Índices	Valores
Jaccard	0.876289
Anderberg	0.778917
Czekanowsky	0.934066
Kulczynski 2	0.936894
Ochai	0.935479

Novamente, foram analisados a marcação do especialista (Figura 11a) com a mesma marcação, porém com ruídos na lesão (Figura 11d). Mostrando mais uma vez a eficiência dos índices na comparação de dois grupos para verificação de reincidência de *pixel*, nota-se que os índices variam com o acréscimo de novos *pixels* na lesão. Esses resultados podem ser observados na Tabela 10.

**Tabela 10: Teste na imagem A\_1000\_1.LEFT\_CC. Resultado dos índices comparados a marcação original e marcação com ruídos**

Índices	Valores
Jaccard	0.642105
Anderberg	0.527273
Czekanowsky	0.714286
Kulczynski 2	0.714554
Ochai	0.71442

Mais testes nas imagens de mamas densas com lesões malignas foram feitos. A mamografia digitalizada A\_1001\_1.LEFT\_CC da base DDSM, também foi testada com modificações no tamanho da marcação da lesão e ruídos gerados para validação dos testes. Como segue a Figura 12.

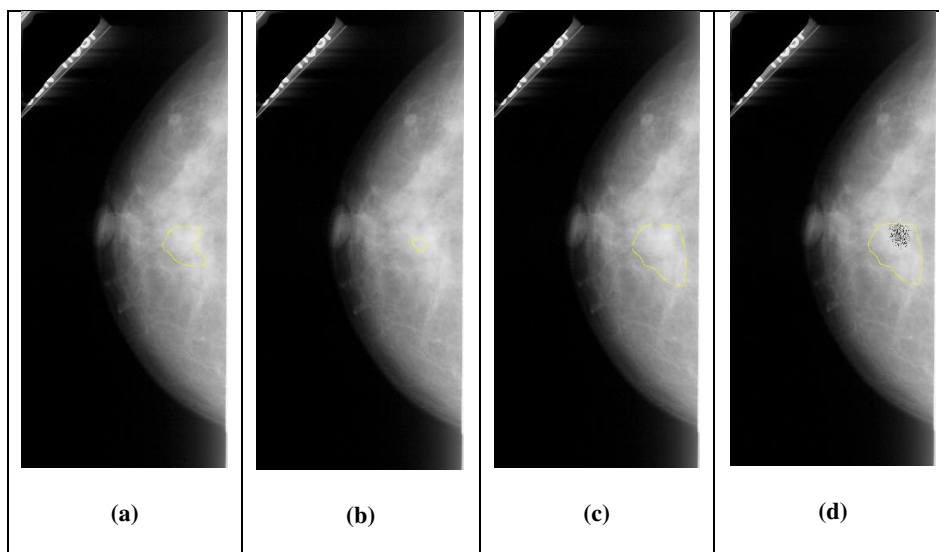


Figura 12: Imagem A\_1001\_1.LEFT\_CC da DDSM (a) Marcação do especialista (b) Marcação Reduzida (c) Marcação Aumentada (d) Ruídos na Lesão

O comportamento dos índices testados com a marcação do especialista (Figura 12a) e a marcação diminuída (Figura 12b) continuaram obedecendo o padrão. Os índices tendem a ficar baixos como mostra a Tabela 11, pelo mesmo fato da região de marcação do especialista abranger parte da mama e, ao reduzir a marcação, a mesma engloba *pixels* bem mais internos da lesão propriamente dita.

Tabela 11: Teste na imagem A\_1001\_1.LEFT\_CC. Resultado dos índices comparados a marcação original e a marcação reduzida

Índices	Valores
Jaccard	0.363636
Anderberg	0.222222
Czekanowsky	0.533333
Kulczynski 2	0.681818
Ochai	0.603023

Considerando a marcação do especialista (Figura 12a) e a marcação aumentada (Figura 12c), percebemos que os índices tendem, novamente, a ser mais similares, isto é, próximo de 1, como exposto na Tabela 12. Assim, podemos ver que o comportamento do índice quando há um aumento da marcação, também obedece o padrão de aumentar a similaridade. Mais uma vez, isso ocorre pelo fato de que está sendo comparado *pixels* da mama e lesão, com mais *pixels* da mama e a mesma lesão.

**Tabela 12: Teste na imagem A\_1001\_1.LEFT\_CC. Resultado dos índices comparados a marcação original e a marcação aumentada**

Índices	Valores
Jaccard	0.72973
Anderberg	0.574468
Czekanowsky	0.84375
Kulczynski 2	0.844575
Ochai	0.844162

Finalizando os testes em mamas densas, foram analisados a marcação do especialista (Figura 12a) com a mesma marcação ruidosa (Figura 12d). Mostrando mais uma vez a eficiência dos índices na comparação de dois grupos para verificação de reincidência de *pixel*, nesse teste em particular, com pouco ruído inserido na marcação, nota-se que não houve muita variação dos índices. Esses resultados podem ser observados na Tabela 13.

**Tabela 13: Teste na imagem A\_1001\_1.LEFT\_CC. Resultado dos índices comparados a marcação original e marcação com ruídos**

Índices	Valores
Jaccard	0.878378
Anderberg	0.783133
Czekanowsky	0.935252
Kulczynski 2	0.93763
Ochai	0.93644

#### **4.2. Mamas Não-Densas**

Análogo aos testes realizados em Mamas Densas, será apresentado nesta seção os testes realizados em Mamas Não-densas. Elas são divididas em mamas com lesões benignas e lesões malignas. Serão feitas simulações de lesões em mamografias digitalizadas de dois pacientes distintos em cada tipo de lesão.



#### 4.2.1. Lesões Benignas

Os testes em mamas não-densas com lesões benignas foram aplicados em duas imagens da base DDSM. Primeiramente, mostraremos a mamografia digitalizada A\_0038\_1.LEFT\_CC. Esta também foi testada com modificações no tamanho da marcação da lesão e ruídos gerados para validação dos testes, como mostra a Figura 13.

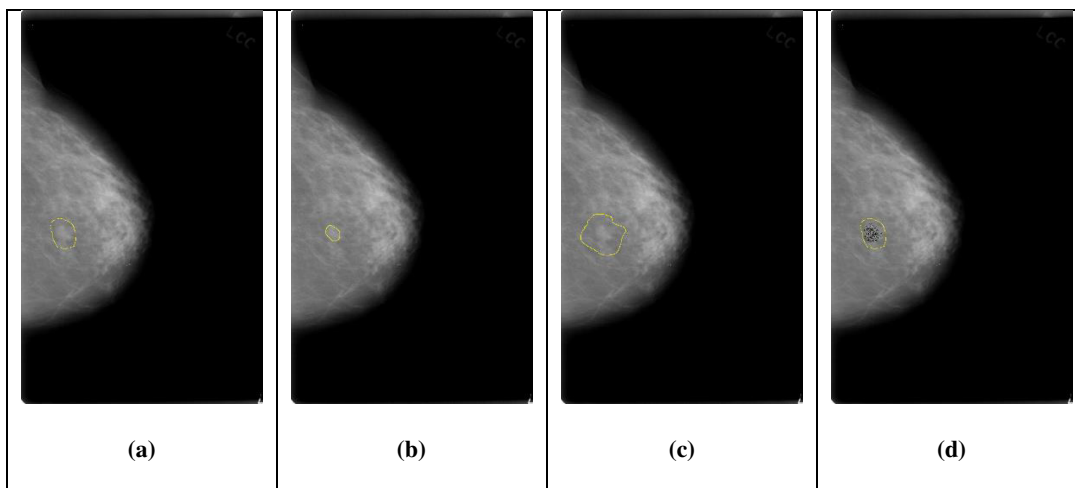


Figura 13: Imagem A\_0038\_1.RIGHT\_CC da DDSM (a) Marcação do especialista (b) Marcação Reduzida (c) Marcação Aumentada (d) Ruídos na Lesão

O comportamento dos índices testados com a marcação do especialista (Figura 13a) e a marcação diminuída (Figura 13b) também continuaram a manter as características das mamas densas. Os índices tendem a ficar baixos como mostra a Tabela 14 e isso ocorre porque a região de marcação do especialista abrange parte da mama, e ao reduzir a marcação, a mesma engloba *pixels* bem mais internos a lesão propriamente dita.

Tabela 14: Teste na imagem A\_0038\_1.RIGHT\_CC. Resultado dos índices comparados a marcação original e a marcação reduzida

Índices	Valores
Jaccard	0.487805
Anderberg	0.322581
Czekanowsky	0.655738
Kulczynski 2	0.657328
Ochai	0.656532

Testando a marcação do especialista (Figura 13a) e a marcação aumentada (Figura 13c), percebemos que os índices tendem, mais uma vez, a ser mais similares, isto é, próximo de 1, como mostra a Tabela 15. Assim, percebe-se que o comportamento dos índices quando há um aumento da marcação também segue as características mostradas nas mamas densas.

**Tabela 15: Teste na imagem A\_0038\_1.RIGHT\_CC. Resultado dos índices comparados a marcação original e a marcação aumentada**

Índices	Valores
Jaccard	0.823529
Anderberg	0.7
Czekanowsky	0.903226
Kulczynski 2	0.907001
Ochai	0.905111

Testes também foram aplicados nas imagens da marcação do especialista (Figura 13a) com a mesma marcação ruidosa (Figura 13d). Os resultados são mostrados na Tabela 16. Os índices também se comportaram como nas mamas densas e tenderam a reduzir a similaridade pelo acréscimo de novos *pixels* não existentes na imagem com marcação real.

**Tabela 16: Teste na imagem A\_0038\_1.RIGHT\_CC. Resultado dos índices comparados a marcação original e marcação com ruídos**

Índices	Valores
Jaccard	0.638889
Anderberg	0.469388
Czekanowsky	0.779661
Kulczynski 2	0.779885
Ochai	0.779773

A segunda imagem a ser testada é a **A\_1170\_1.RIGHT\_MLO**. Também foram feitas variações no tamanho da marcação da lesão e ruídos foram gerados para validar os testes, isso pode ser visto na Figura 14.

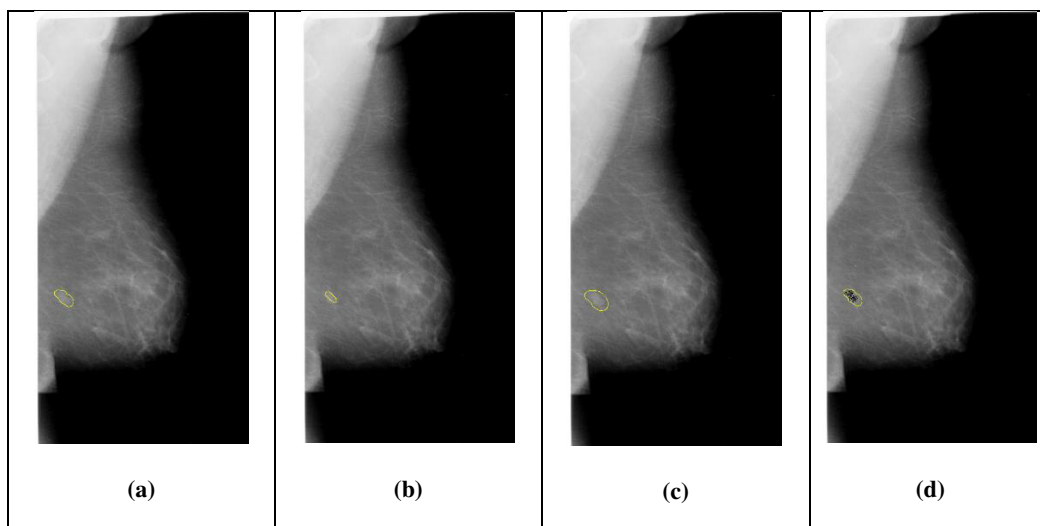


Figura 14: Imagem A\_1170\_1.RIGHT\_MLO da DDSM (a) Marcação do especialista (b) Marcação Reduzida (c) Marcação Aumentada (d) Ruídos na Lesão

Considerando a imagem com a marcação do especialista (Figura 14a) e a marcação diminuída (Figura 14b), mostrou-se pouca similaridade, isto é, próximo de zero como mostra a Tabela 17. A ocorrência disto também é devido ao fato que a região de marcação do especialista abrange parte da mama, e ao reduzir a marcação, a mesma engloba *pixels* bem mais internos a lesão em si.

Tabela 17: Teste na imagem A\_1170\_1.RIGHT\_MLO. Resultado dos índices comparados a marcação original e a marcação reduzida

Índices	Valores
Jaccard	0.309524
Anderberg	0.183099
Czekanowsky	0.472727
Kulczynski 2	0.510714
Ochai	0.491354

Aumentando a marcação (Figura 14c) e comparando-a a marcação real (Figura 14a), mesmo alguns índices não se aproximando de 1, como os Índices de Jaccard e Anderberg, percebe-se que os índices aumentaram seu valor comparado ao teste anterior da Tabela 14. Isso ocorreu pelo motivo de que quando aumenta-se o tamanho da marcação, *pixels* novos - que não estavam nem na lesão em si, nem na marcação (que engloba parte da mama) - estão inseridos dentro da marcação aumentada, e ao calcular as equações de alguns índices a variação ficou próximo de 0,5. Esse resultado é apresentado na Tabela 18.

**Tabela 18: Teste na imagem A\_1170\_1.RIGHT\_MLO. Resultado dos índices comparados a marcação original e a marcação aumentada**

Índices	Valores
Jaccard	0.666667
Anderberg	0.5
Czekanowsky	0.8
Kulczynski 2	0.822857
Ochai	0.811348

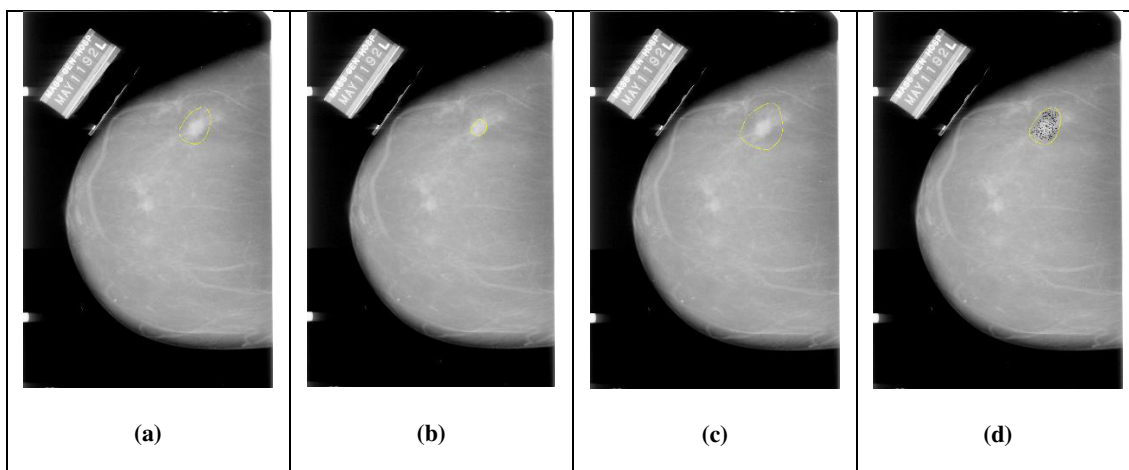
Na comparação entre a marcação real (Figura 14a) e a marcação ruidosa (Figura 14d), como nos demais testes apresentados, o comportamento dos índices se manteve igual. Pela razão que com o acréscimo de valores *pixels* novos, a similaridade diminui, isso é, os índices se aproximam de 0, como é visto na Tabela 19.

**Tabela 19: Teste na imagem A\_1170\_1.RIGHT\_MLO. Resultado dos índices comparados a marcação original e marcação com ruídos**

Índices	Valores
Jaccard	0.513636
Anderberg	0.342623
Czekanowsky	0.660563
Kulczynski 2	0.660714
Ochai	0.660639

#### **4.2.2. Lesões Malignas**

Assim como os testes realizados em mamas não-densas e benignas, será apresentado os testes em mamas não-densas e com lesões malignas. Também foram feitos dois testes em duas imagens da DDSM, usando as mesmas características de diminuição e aumento da marcação, e geração de ruídos para validação de eficiência dos índices. A primeira imagem a ser testada é a **A\_1004\_1.LEFT\_CC**, que é mostrada na Figura 15.



**Figura 15: Imagem A\_1004\_1.LEFT\_CC da DDSM (a) Marcação do especialista (b) Marcação Reduzida (c) Marcação Aumentada (d) Ruídos na Lesão**

Nesse teste, a marcação do especialista (Figura 15a) e a marcação diminuída (Figura 15b), mostrou pouca similaridade, isto é, próximo de zero, como mostra a Tabela 14. Novamente, buscou-se pegar região mais interna à lesão para comparar com a marcação do especialista, justificando os valores de pouca similaridade.

**Tabela 20: Teste na imagem A\_1004\_1.LEFT\_CC. Resultado dos índices comparados a marcação original e a marcação reduzida**

Índices	Valores
Jaccard	0.280702
Anderberg	0.163265
Czekanowsky	0.438356
Kulczynski 2	0.439098
Ochai	0.438727

Em um outro teste, foi analisado a marcação do especialista (Figura 15a) e uma marcação maior que a do especialista (Figura 15c). Os resultados estão na Tabela 21. Mesmo aumentando grande parte da área de marcação, os índices se comportaram dentro do padrão, ou seja, tenderam a aumentar seu valor, já que ambas marcações abrangem a mama e a lesão propriamente dita.

**Tabela 21: Teste na imagem A\_1004\_1.LEFT\_CC. Resultado dos índices comparados a marcação original e a marcação aumentada**

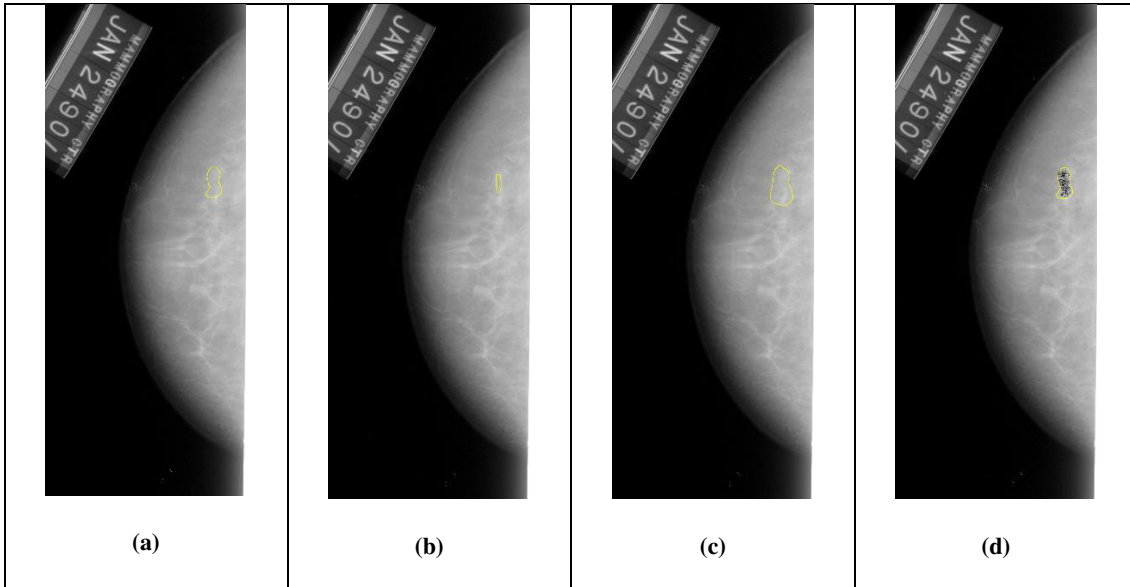
Índices	Valores
Jaccard	0.795455
Anderberg	0.660377
Czekanowsky	0.886076
Kulczynski 2	0.897727
Ochai	0.891883

Considerando a marcação real (Figura 15a) e a marcação ruidosa (Figura 15d), o comportamento dos índices se manteve igual. O acréscimo de valores de *pixels* novos fez com que a similaridade diminuísse, isso é, os índices se aproximam de 0, como é visto na Tabela 22.

**Tabela 22: Teste na imagem A\_1004\_1.LEFT\_CC. Resultado dos índices comparados a marcação original e marcação com ruídos**

Índices	Valores
Jaccard	0.492621
Anderberg	0.372163
Czekanowsky	0.547255
Kulczynski 2	0.548197
Ochai	0.547626

Como último teste da metodologia proposta, foi analisada uma segunda imagem não-densa e com lesão maligna. A imagem A\_1003\_1.LEFT\_CC foi analisada da mesma forma que as anteriores, fazendo redução e aumento da marcação e gerando ruídos. A Figura 16 mostra a imagem e as simulações feitas.



**Figura 16: Imagem A\_1003\_1.LEFT\_CC da DDSM (a) Marcação do especialista (b) Marcação Reduzida (c) Marcação Aumentada (d) Ruídos na Lesão**

Analisando a marcação do especialista (Figura 16a) e a marcação reduzida (Figura 16b), percebemos o mesmo comportamento que até agora foi apresentado. Os índices tendem a ir para 0, isto é, pouca similaridade. Os resultados dos valores dos índices podem ser vistos na Tabela 23.

**Tabela 23: Teste na imagem A\_1003\_1.LEFT\_CC. Resultado dos índices comparados a marcação original e marcação diminuída**

Índices	Valores
Jaccard	0.425532
Anderberg	0.27027
Czekanowsky	0.597015
Kulczynski 2	0.649225
Ochai	0.622573

O padrão de resultados comparando marcação do especialista (Figura 16a) e a marcação aumenta (Figura 16b) se manteve igual aos demais, como pode ser visto na Tabela 24.

**Tabela 24: Teste na imagem A\_A\_1003\_1.LEFT\_CC. Resultado dos índices comparados a marcação original e marcação aumentada**

Índices	Valores
Jaccard	0.76087
Anderberg	0.614035
Czekanowsky	0.864198
Kulczynski 2	0.867503
Ochai	0.865849

Na análise com a marcação do especialista (Figura 16a) e a marcação com ruídos (Figura 16d), também foi notado que os índices se aproximaram de 0. Novamente, isso ocorre pelo fato de novos valores *pixels* que não existem na marcação real serem gerados dentro da marcação. O resultado pode ser visto na Tabela 25.

**Tabela 25: Teste na imagem A\_1003\_1.LEFT\_CC. Resultado dos índices comparados a marcação original e marcação com ruídos**

Índices	Valores
Jaccard	0.325622
Anderberg	0.21326
Czekanowsky	0.478315
Kulczynski 2	0.534325
Ochai	0.526673

O estudo dos índices de similaridade aplicados em mamografias mostrou-se bastante promissor. Analisando cada caso de teste e estabelecendo padrões no comportamento dos índices, estes podem ser de grande auxílio para o especialista.

A variação dos índices nos testes – com aumento, diminuição, e em geração de ruídos – mostra que, em casos em que há mudança, tanto na quantidade quanto na reincidência ou ausência de *pixels*, existe variação nos valores dos índices. Fazendo analogia ao caso real de um paciente em tratamento, em que este tratamento altera o tecido da lesão, a variação dos índices auxiliará o especialista na verificação da eficácia ou não do tratamento.



Em outra situação, considerando um paciente em que não houve diagnóstico exato e optou-se em acompanhar a lesão, no caso em que apenas o comportamento da geometria (forma) não é suficiente para classificação da lesão, analisar a mudança da textura da lesão tem fundamental importância para classificação, ou validação do diagnóstico.

## **5. CONCLUSÃO**

Esta monografia apresentou o estudo de índices de similaridade para serem aplicados na análise temporal de lesões mamárias. Esse estudo mostra-se importante, uma vez que possibilita fazer uma correspondência entre lesões em exames diferentes e fornecer ao especialista informações quantitativas sobre as mudanças de textura utilizando índices de similaridade. Com essa abordagem de informações quantitativas da lesão, será possível um melhor acompanhamento, a fim de determinar o seu comportamento no decorrer do tempo. Dessa forma, pressupõe-se que o médico terá melhor conhecimento sobre as mudanças de textura (tecido) da lesão e possa usá-las em seu favor, dando assim um auxílio importante no que se trata da opção cirúrgica ou não de uma lesão, e até mesmo poder identificar a efetividade de um tratamento que está sendo aplicado aquela lesão no decorrer do tempo.

### **5.1. Avaliação do Trabalho**

Este trabalho ainda encontra-se em processo de desenvolvimento. Visto que testes realizados validam o comportamento dos índices e da sua relevância, porém, como citado, não foi possível a aplicação dele em uma base de casos reais, já que até o presente momento não foi encontrado uma base que tivesse as características desejadas para a aplicação da metodologia proposta.

Atualmente, já há um trabalho para a construção de uma base para atender tais perspectivas. Com uma parceria entre o Programa de Pós-Graduação em Saúde Pública e o Núcleo de Computação Aplicada, ambos da Universidade Federal do Maranhão, está sendo construída uma base de imagens de mamografias com diferentes lesões, posições e com casos para análise temporal. Todavia, até o presente momento ainda não conta com a marcação dos especialistas, mas já está sendo providenciado com profissionais do Programa de Pós-Graduação em Saúde Pública.

O projeto da construção desta base iniciou-se, dentre outras finalidades, para atender o estudo proposto deste trabalho. No entanto, o mesmo ainda não chegou a etapa final. Esta base encontra-se, atualmente, com 89 pacientes, sendo um total de 296 imagens de mamografias digitalizadas. Todas foram digitalizadas utilizando um scanner próprio para digitalização de radiografia (*Scanner MICROTEK MEDI – 6000 plus*).

## **5.2. Trabalhos Futuros**

Como já mencionado, o problema na aquisição de uma base para a análise temporal tem se tornado o maior impasse dentro do trabalho. Para que se obtenha resultados reais que possam melhor validar a metodologia apresentada, como trabalho futuro, será procurado uma base em que possa aplicar a metodologia de análise temporal.

Também, com a conclusão da etapa de marcação da base citada na Seção 5.1, será possível a realização de testes em uma base nacional com marcação de dois especialistas validadas pelo laudo que também está acompanhando cada paciente participante desse projeto.

Na literatura, há muitas citações de índices de similaridade e diversidade, logo, o estudo para a aplicação desses índices também pode ser citado como trabalho futuro a fim de validar a similaridade e/ou diversidade da lesão em mamografias adquiridas em tempos diferentes.

## REFERÊNCIAS

ALBUQUERQUE e ALBUQUERQUE. *Processamento de Imagens: Métodos e Análises*. Centro Brasileiro de Pesquisas Físicas, 2000.

ANGELO. *Sistema de Processamento de Imagens Mamográficas e Auxílio ao Diagnóstico via-Internet. (Tese de Doutorado)*. Universidade de São Paulo. São Carlos, 2007.

BIRD, WALLACE e YANKASKAS. *Analysis of cancers missed at screening mammography*, Radiology, vol. 184, nº 3, p. 613-617, 1992.

BROWN, L. G. *A survey of image registration techniques*. ACM Computing Surveys (CSUR) archive. 24 (4). 325 – 376, 2004.

CRUM, W. R.; HARTKENS T.; HILL, D. L. G. *Non-rigid image registration: theory and practice*. The British Journal of Radiology. 77. 140 – 153, 2004.

DINIZ, P. H. B. *Análise temporal de nódulos e massas pulmonares utilizando índices de similaridade. Dissertação (Mestrado Engenharia Elétrica)*. Universidade Federal do Maranhão, 2014.

GOSHTASBY, A. A. *2-D and 3-D image registration for medical, remote sensing, and industrial applications*, 2007.

HAJNAL, J. V.; HILL, D. L. G.; HAWKES, D. J. *Medical Image Registration*. 1ª Edição. CRC Press, 2001.

HUHTA, V. *Evaluation of different similarity indices as measures of succession in arthropod communities of the forest after clearcutting*. Oecologia (Berl). vol. 41. p. 11-23, 1979.

INCA - Instituto Nacional do Câncer. Disponível em: <<http://www.inca.gov.br/index.asp>>. Última visita em 13/11/2014.

KERLIKOWSKA, CARNEY, GELLER, MANDELSON, TAPLIN, MALVIN, ERNSTER, URBAN, CUTTER, ROSENBERG E BALLARD-BARBASH. *Performance of screening mammography among women with and without a first-degree relative with breast cancer*. Annals of Internal Medicine, vol. 133, nº 11, p. 855–863, 2000.

KOPANS, D.; MOORE, R.; HEATH, M.; BOWYER, K.; KEGELMEYER, P. J. “*The Digital Database for Screening Mammography*”, In: Proceedings of the Fifth International Workshop on Digital Mammography, Medical Physics Publishing, p. 212-218, 2001.

KOPANS. *Imagem da Mama*. 2ª ed. MEDSI Editora Médica e Científica Ltda. Rio de Janeiro, 2000.

LUDOMIR HADJIISKI; BERKMAN SAHINER; HEANG-PING CHAN; NICHOLAS PETRICK; MARK A. HELVIE; METIN GURCAN. *Analysis of temporal changes of mammographic features: Computer-aided classification of malignant and benign breast masses*. American Association of Physicists in Medicine, p. 2309-2317, 2001.

MEYER, A. S. *Comparação de Coeficientes de Similaridade usados em Análises de Agrupamento com Dados de Marcadores Moleculares Dominantes*. (Dissertação de Mestrado). Universidade de São Paulo, 2002

PAL e PAL. *A Review on Image Segmentation Techniques*. *Pattern Recognition*, Vol. 26, p. 1277–1294, 1993.

PARRAGA, A. *Atlas Anatômico da Região da Cabeça e do Pescoço: em Direção à Radioterapia Adaptativa*. (Tese Doutorado em Engenharia Elétrica). Universidade Federal do Rio Grande do Sul, 2008.

ROBERT MARTI; YAGO DÍEZ; ARNAU OLIVER; MERITXELL TORTAJADA; REYER ZWIGGELAAR; XAVIER LLADO. *Detecting Abnormal Mammographic cases in Temporal Studis using Image Registration Features*. Breast Imaging Lecture Notes in Computer Science, vol. 8539, p. 612 – 619, 2014

SAMPAIO, W. B. *Detecção de massas em imagens mamográficas usando redes neurais celulares, funções geoestatísticas e máquinas de vetores de suporte. Dissertação (Mestrado Engenharia Elétrica)*. Universidade Federal do Maranhão, 2009.

SASKIA VAN ENGELAND; PETER SNOEREN; JAN HENDRIKS; NICO KARSSEMEIJER. *A comparison of methods for mammogram registration*. IEEE Transactions on Medical Imaging, vol. 22, nº 11, p. 1436 – 1444, 2003

SEIXAS, F. L. *Diagnóstico Auxiliado por Computador*. Relatório Técnico do Laboratório MídiaCom. Departamento de Engenharia de Telecomunicações da UFF. Niterói. RJ, 2005.

WANG E KARAYIANNIS. *Detection of microcalcifications in digital mammograms using wavelets*, IEEE Transactions on Medical Imaging, vol. 17, nº 4, p. 498–509, 1998.