

BRUNO ROBERTO NEPOMUCENO MATHEUS

**BancoWeb: base de imagens mamográficas para  
auxílio em avaliações de esquemas CAD**

Dissertação apresentada a Escola de  
Engenharia de São Carlos da  
Universidade de São Paulo para  
obtenção de título de Mestre em  
Ciências, Programa de Engenharia  
Elétrica

Área de concentração: Processamento  
de Sinais e Instrumentação

Orientador: Prof. Dr. Homero Schiabel

São Carlos

2010

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica preparada pela Seção de Tratamento  
da Informação do Serviço de Biblioteca – EESC/USP

M427b Matheus, Bruno Roberto Nepomuceno  
BancoWeb : base de imagens mamográficas para auxílio  
em avaliações de esquemas CAD / Bruno Roberto Nepomuceno  
Matheus ; orientador Homero Schiabel. -- São Carlos,  
2010.


Dissertação (Mestrado - Programa de Pós-Graduação em  
Engenharia Elétrica e Área de Concentração em  
Processamento de Sinais e Instrumentação) -- Escola de  
Engenharia de São Carlos da Universidade de São Paulo,  
2010.

1. Mamografia. 2. Diagnóstico auxiliado por computador  
CAD. 3. Base de imagens. 4. Avaliação de CAD. I. Título.

**FOLHA DE JULGAMENTO**

Candidato: Bacharel **BRUNO ROBERTO NEPOMUCENO MATHEUS**.

Dissertação defendida e julgada em 22/04/2010 perante a Comissão Julgadora:



Prof. Associada **HOMERO SCHIABEL** - (Orientador)  
(Escola de Engenharia de São Carlos/USP)

APROVADO



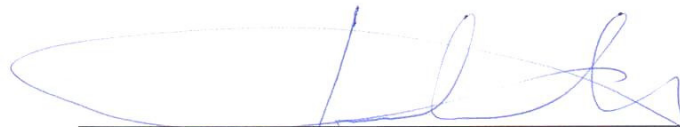
Dr. **MARCO ANTONIO GUTIERREZ**  
(Instituto do Coração/USP)

APROVADO



Prof. Titular **CAETANO TRAINA JUNIOR**  
(Instituto de Ciências Matemáticas e de Computação/USP)

Aprovado



Prof. Titular **GERALDO ROBERTO MARTINS DA COSTA**  
Coordenador do Programa de Pós-Graduação em Engenharia Elétrica e  
Presidente da Comissão de Pós-Graduação



À minha amada Gi,  
por me indicar o caminho, me acalmar e me agüentar.

Aos meus pais,  
por me fazer quem eu sou, me apoiar e sempre me incentivar.

E ao meu irmão,  
por me irritar, me guiar e sempre ser meu exemplo, mesmo quando longe.

Eu não teria chegado aqui sem vocês.



## **Agradecimentos**

Ao Prof. Dr. Homero Schiabel, pelo auxílio, direção e paciência durante a orientação desta pesquisa.

Ao Prof. Dr. Mauricio Cunha Escarpinati, por me encaminhar neste projeto.

Ao Jarbas da Silva Ferro Neto, por me auxiliar com a inserção das imagens, me permitindo finalizar o projeto.

Ao Departamento de Engenharia Elétrica da Escola de Engenharia de São Carlos, pela oportunidade de realização deste mestrado.

Ao Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto, Santa Casa de Misericórdia de São Carlos, Hospital São Paulo e Hospital das Clínicas da Faculdade de Medicina da UNESP, pelas imagens concedidas ao longo dos anos que culminaram na realização deste projeto.

À CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), pela concessão da bolsa de mestrado e auxílio financeiro para realização desta pesquisa.





## Resumo

Matheus, B. R. N., **BancoWeb: base de imagens mamográficas para auxílio em avaliações de esquemas CAD**. 2010. 71 f. Dissertação (Mestrado) - Escola de Engenharia de São Carlos - Universidade de São Paulo, São Carlos, 2010.

Este trabalho teve como objetivo desenvolver uma base de imagens mamográficas *online* com acesso público para desenvolvimento, testes e avaliação comparativa de esquemas computadorizados de auxílio ao diagnóstico (CADs). A base contém imagens de vários hospitais, com grande variedade de laudos, também disponíveis na base assim como informações sobre dados clínicos (não confidenciais) dos pacientes. Uma interface detalhada foi criada para permitir o fácil acesso público, permitindo o uso de ferramentas de busca, recorte, análise estatística e inserção remota de imagens, entre outras. Testes comparativos com bases já existentes e amplamente usadas mostraram que a base desenvolvida tem quantidade e qualidade de imagens comparável ou superior as outras, além de oferecer uma quantidade de ferramentas muito maior.

**Palavras-chave:** Mamografia, Diagnóstico auxiliado por computador (CAD), base de imagens, avaliação de CAD



## Abstract

Matheus, B. R. N., **BancoWeb: Mammographic image database for CAD schemes evaluation**. 2010. 71 f. Dissertação (Mestrado) - Escola de Engenharia de São Carlos - Universidade de São Paulo, São Carlos, 2010.

This work has the objective of developing an online mamographic image database with public access for development, test and evaluation of computer-aided diagnosis (CAD). The database contains images from several hospitals, with great variety of medical reports, also available on the database together with other clinical information (not classified) from the patient. A detailed interface was developed to allow easy public access, allowing the use of tools for search, clipping, statistical analyses, remote insertion and others. Comparative test with other already existing databases shown that the presented database has quantity and quality comparable or superior to the others and offers a greater set of tools for the user.

**Key words:** mammography, computer-aided diagnosis, images database, CAD evaluation.



# Lista de figuras

FIGURA 1: DESEMPENHO DE UM ESQUEMA CAD PARA DETECTAR E CLASSIFICAR MICROCALCIFICAÇÕES COMPARADOS AO DESEMPENHO DOS RADIOLOGISTAS SEM AUXILIO (VYBORNY, GIGER E NISHIKAWA, 2000).....	26
FIGURA 2: EXEMPLO DE TABELA DE MODELO RELACIONAL.....	27
FIGURA 3: ANATOMIA DE UMA DECLARAÇÃO EXEMPLO EM SQL.....	29
FIGURA 4: MODELO RELACIONAL DA BASE.....	44
FIGURA 5: PAGINA INICIAL.....	45
FIGURA 6: CADASTRO DO USUÁRIO.....	46
FIGURA 7: MENUS PARA NÍVEL DE ACESSO "USUÁRIO".....	47
FIGURA 8: MENUS PARA NÍVEL DE ACESSO "PESQUISADOR".....	48
FIGURA 9: MENUS PARA NÍVEL DE ACESSO "ADMINISTRADOR".....	48
FIGURA 10: INSERÇÃO DE DADOS DO HOSPITAL.....	49
FIGURA 11: INSERÇÃO DE DADOS APARELHO DE RAIOS X.....	50
FIGURA 12: INSERÇÃO DE DADOS DO PACIENTE.....	51
FIGURA 13: INSERÇÃO DE DADOS DO SCANNER.....	51
FIGURA 14: INSERÇÃO DE DADOS DO EXAME.....	52
FIGURA 15: INSERÇÃO DE DADOS DA IMAGEM.....	53
FIGURA 16: INSERÇÃO DOS ACHADOS DA IMAGEM.....	54
FIGURA 17: PAINEL DE BUSCA POR (A) DADOS DA IMAGEM; (B) DADOS DO PACIENTE; (C) APARELHO DE RAIOS X; (D) SCANNER.....	55
FIGURA 18: EXEMPLO DE RESULTADOS DA BUSCA.....	56
FIGURA 19: EXEMPLO DE BLOCO DE RESULTADO.....	57
FIGURA 20: EXEMPLO DA OPÇÃO RECORTE (REGIÃO DEMARCADA APENAS PARA ILUSTRAÇÃO).....	58
FIGURA 21: OPÇÃO "MAIS INFORMAÇÕES".....	58
FIGURA 22: INSTRUÇÕES PARA USO DO SISTEMA DE RECORTE.....	59
FIGURA 23: RECORTE SENDO REALIZADO.....	60
FIGURA 24: FERRAMENTA DE PERFIL ESTATÍSTICO.....	61
FIGURA 25: EXEMPLO DO USO DA FERRAMENTA DE PERFIL ESTATÍSTICO.....	61
FIGURA 26: DISTRIBUIÇÃO ESTATÍSTICA DOS ACHADOS.....	68
FIGURA 27: DISTRIBUIÇÃO DOS PACIENTES DE ACORDO COM A IDADE NO MOMENTO DO EXAME.....	69



## Lista de siglas

INCA	Instituto Nacional de Câncer
ECM	Exame clínico de mama
CAD	<i>Computer-aided Diagnosis</i>
LAPIMO	Laboratório de Análise e Processamento de Imagens Médicas e Odontológicas
FDA	U. S. Food and Drug administration
ROC	<i>Receiver Operating Characteristic</i>
SQL	<i>Structured Query Language</i>
SGBD	Sistema de gerenciamento de base de dados
MySQL	Dialeto do SQL
CT	<i>Computer Tomography</i>
MRI	<i>Magnetic Resonance Image</i>
TIFF	<i>Tagged Image File Format</i>
LJPEG	<i>Loosy Joint Photographic Experts Group</i>
RAW	Formato de imagem minimamente processado
MIAS	<i>Mammographic Image Analysis Society</i>
DDSM	<i>Digital Database for Screening Mammography</i>
LLNL/UCSF	<i>Lawrence Livermore National Laboratories / University of California at San Fransisco</i>
CALMa	<i>Computer-Assisted Library for Mammography</i>
HCFMRP-USP	Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto
UNESP	Universidade Estadual Paulista
BI-RADS	<i>Breast Imaging-Reporting and Data System</i>
DICOM	<i>Digital Imaging and Communications in Medicine</i>
CR	<i>Computed Radiography</i>
FFDM	<i>Full Field Digital Mammography</i>





# Sumário

Capítulo 1: Introdução .....	19
1.1 Câncer de mama.....	19
1.2 Detecção e diagnóstico de câncer de mama.....	19
1.3 Sistemas CAD.....	20
1.4 Objetivos .....	21
1.4 Disposição do trabalho.....	22
Capítulo 2: CAD mamográfico .....	23
2.1 Desenvolvimento de esquemas CAD mamográficos.....	23
2.2 Avaliação de CAD.....	24
2.3 Curvas ROC .....	25
Capítulo 3: Banco de dados.....	27
3.1 Definição de banco de dados .....	27
3.2 Modelo relacional .....	27
3.3 Linguagem de consulta.....	28
3.4 SQL.....	29
3.4.1 Oracle .....	30
3.4.2 MySQL .....	31
Capítulo 4: Bases de imagens mamográficas .....	33
4.1 Características de Bancos de imagens médicas .....	33
4.2 Requerimentos de Bases mamográficas .....	33
4.2.1 Conjuntos de imagens.....	34
4.2.2 Interface .....	35
4.3 Bases existentes .....	35
4.3.1 MIAS database .....	35
4.3.2 DDSM.....	36
4.3.3 LLNL/UCSF .....	37
4.3.4 CALMa .....	38
4.4 Considerações finais.....	38
Capítulo 5: Metodologia.....	39
5.1 Histórico do projeto .....	39
5.2 Aquisição das imagens .....	40
5.3 Estruturação do banco .....	43
5.4 Interface e uso do BancoWeb .....	45
5.4.1 Procedimento de cadastro .....	46
5.4.2 Inserção de imagens.....	49
5.4.3 Busca e recuperação de imagens .....	54

5.4.4 Opções e ferramentas .....	55
Capítulo 6: Resultados.....	63
6.1 Acesso, manutenção e atualização .....	63
6.2 Quantidade de imagens .....	63
6.3 Resolução espacial e de contraste .....	65
6.4 Formato dos arquivos de imagem.....	65
6.5 Sistemas de busca .....	66
6.6 Regiões de interesse.....	66
6.7 Outras ferramentas .....	67
Capítulo 7: Discussão e Conclusões .....	71
Referências .....	77

## Capítulo 1: Introdução

### 1.1 Câncer de mama

Segundo estatísticas do INCA (Instituto Nacional de Câncer, <http://www.inca.gov.br>) o câncer de mama é o segundo câncer mais comum no mundo. Em mulheres ele corresponde a 22% dos casos diagnosticados de câncer a cada ano, o que corresponde a uma estimativa para 2010 de 49.240 novos casos apenas no Brasil.

Se corretamente diagnosticado e tratado em seus estágios iniciais o prognóstico é relativamente bom. Infelizmente a taxa de mortalidade ainda é alta, muito provavelmente porque a doença é diagnosticada em seus estágios mais avançados. Na população mundial a taxa de sobrevivência após 5 anos é de 61%.

Normalmente o câncer de mama é encontrado em mulheres de mais do que 35 anos, apesar de raros casos acontecerem antes. A probabilidade aumenta rapidamente com o aumento da idade.

### 1.2 Detecção e diagnóstico de câncer de mama

A detecção do câncer de mama pode ser feita, de modo eficaz, por dois métodos: o exame clínico de mama (ECM) e a mamografia.

O ECM exige apenas um médico ou enfermeira treinados podendo detectar tumores de até 1,0 cm, se superficial. As limitações deste método em geral resultam em detecções tardias. A mamografia é considerada, até o momento, o melhor método para detecção precoce e diagnóstico de câncer de mama, sendo, quando propriamente realizada, capaz de detectar lesões em fases iniciais, com menos de um milímetro.

O exame mamográfico, ao contrário do ECM, exige um equipamento apropriado, chamado mamógrafo, que é um aparelho de raios X com o campo mais restrito e uma maior sensibilidade para detecção de objetos pequenos. Como este exame tem como foco a detecção e avaliação de lesões extremamente pequenas, é muito importante que todo o processo de aquisição da imagem seja feito com grande cuidado e atenção. Isto inclui manutenção e calibração dos mamógrafos e qualificação de técnicos especializados para realização do exame.

Cada exame mamográfico é composto geralmente por 4 imagens (uma crânio-caudal e uma médio lateral, para cada mama). Isto quer dizer que cada exame realizado exige que o radiologista avalie 4 imagens.

Estudos indicam que entre 10 e 15% (PEIXOTO; CANELLA; AZEVEDO, 2007) das mamografias são erroneamente avaliadas, podendo resultar em biópsias desnecessárias, ou até mesmo, atrasar o tratamento do câncer.

### 1.3 Sistemas CAD

Estudos mostram que o uso de uma segunda opinião melhora consideravelmente os resultados (GEORGIAN-SMITH; MOORE; HALPERN, 2007), mas a falta de profissionais treinados impede esta opção de ser utilizada na maioria dos hospitais e clínicas. Uma solução a este problema é a estruturação de um esquema CAD (*Computer-aided Diagnosis*), que corresponde a um programa de auxílio ao diagnóstico, ou seja, um sistema computacional que analisa as informações disponíveis e tenta apresentar um diagnóstico com base nos dados fornecidos pela imagem processada. No caso da mamografia, usualmente o sistema analisa as imagens resultantes do exame e fornece marcações com regiões de interesse para o médico, informando quais achados são prováveis em cada região (BRZAKOVIC; LUO; BRZAKOVIC, 1990; KALLERGI, 2004).

Os esquemas CAD atuais ainda não estão nem perto de se equiparar a um radiologista experiente, mas estudos mostram que um bom esquema CAD no lugar da segunda opinião gera resultados muito próximos a um segundo profissional (GEORGIAN-SMITH; MOORE; HALPERN, 2007).

Para se desenvolver um CAD é necessário que o desenvolvedor tenha, além do conhecimento das estruturas a serem analisadas, acesso a uma significativa quantidade e variedade de imagens com seus laudos confiáveis. Os laudos são necessários para verificar os resultados do programa contra a opinião de um especialista (já que a maioria dos desenvolvedores não é mastologista) e poder facilmente avaliar os resultados obtidos.

A dificuldade surge ao tentar obter uma grande quantidade de imagens mamográficas com laudo. Normalmente, estas imagens, quando não são levadas pelas pacientes, são mantidas em arquivos dos hospitais e clínicas e incluem informações confidenciais. Para acessar tais arquivos, o pesquisador teria que ter contatos em vários hospitais e, no caso de imagens guardadas em filme, autorização para transportar as imagens para digitalização em equipamentos apropriados.

Além da dificuldade de obtenção das imagens de teste durante o desenvolvimento, os esquemas CAD dependem intrinsecamente da qualidade dessas imagens a serem processadas. Características como quantidade de níveis de cinza, resolução espacial, qualidade do mamógrafo, assim como digitalizador utilizado, níveis de ruído, etc. Se apenas uma destas características variar, o resultado de um esquema CAD pode ser alterado consideravelmente (SCHIABEL et al., 2000).

Com o intuito de fornecer imagens mamográficas de qualidade em suficiente quantidade e variedade, bases de imagens começaram a ser construídas, de início localmente, elaboradas por equipes de laboratórios e empresas para desenvolvimento dos seus próprios esquemas de processamento. Posteriormente, alguns centros de pesquisa resolveram tornar suas bases públicas, embora cada base com sua estrutura e seu padrão tanto em relação às características das imagens quanto às ferramentas disponíveis ao desenvolvedor. Estas diferenças dificultam muito a comparação entre diferentes esquemas CAD, pois cada esquema vai ser desenvolvido para ser utilizado com imagens com as características selecionadas pela base escolhida, assim como os testes de validação serão feitos em tal conjunto de imagens.

Para comparar diferentes esquemas CAD entre si seria necessário um conjunto de imagens de alta qualidade, com características muito próximas e livre para uso de todo desenvolvedor.

#### **1.4 Objetivos**

O LAPIMO (LAPIMO, 2009) tem trabalhado nessa linha de pesquisa desde 1997, com o desenvolvimento de uma base de dados/imagens mamográficas para ser usada nos testes de desenvolvimento de programas de processamento voltados à elaboração de um esquema CAD em mamografia (BENATTI, 2003; NUNES et al., 2004; NUNES; SCHIABEL; GOES, 2007). Tendo em vista, ainda, a necessidade e demanda por imagens tentando atender o requisito considerado no parágrafo anterior, tem-se procurado atuar no sentido de tornar disponível um sistema de busca e gerenciamento desses dados de maneira mais ampla, utilizando as ferramentas da Internet (SCHIABEL; ESCAPINARTI; FREITAS, 2006).

Por isso, o objetivo deste trabalho é desenvolver um esquema de gerenciamento da base de dados *on line*, capaz de fornecer um conjunto amplo de imagens mamográficas de qualidade apropriada ao processamento digital. Uma premissa é que o

acesso a esta base deve ser livre para qualquer pesquisador interessado, necessitando apenas de uma conexão com a internet.

Este sistema tem como intuito tornar-se uma base padronizada, de referência para o desenvolvimento e, principalmente, comparação entre diferentes esquemas CAD. Por isso, ela se estrutura de modo a satisfazer todas as exigências usuais de desenvolvedores e todas as características recomendadas na literatura. Além de uma grande preocupação com a qualidade e variedade de imagens, a base apresentada aqui oferece variadas ferramentas para facilitar o estudo das imagens e localização das informações pertinentes a cada caso.

#### **1.4 Disposição do trabalho**

Além deste capítulo de introdução, este trabalho divide-se em 6 outros capítulos divididos como segue abaixo:

**Capítulos 2, 3 e 4:** Apresentam revisões bibliográficas sobre CAD mamográfico, Bancos de dados e Bases de imagens mamográficas.

**Capítulo 5:** apresenta a metodologia do trabalho, incluindo a motivação por trás da escolha das estruturas selecionadas e os detalhes da implementação do mesmo.

**Capítulo 6:** Apresenta os resultados preliminares do projeto e a comparação com as principais bases existentes.

**Capítulo 7:** Apresenta as pendências para o término do trabalho e um cronograma com o que já foi feito semestre a semestre no desenvolvimento deste projeto.

## Capítulo 2: CAD mamográfico

### 2.1 Desenvolvimento de esquemas CAD mamográficos

Em mamografia é extremamente importante localizar e diagnosticar estruturas suspeitas precocemente, pois, como as causas do câncer de mama não são ainda compreendidas, a única forma de reduzir as conseqüências e aumentar a taxa de sobrevivência é o diagnóstico precoce (ANDERSON, 2008).

No entanto as estruturas indicativas de câncer são muito próximas das estruturas naturais de uma mama saudável, em termos de absorção de raios-X e contraste na imagem mamográfica. Este problema se torna particularmente complexo em estruturas muito pequenas, como microcalcificações, que podem ser menores que 0,5 mm. Ao lidar com estruturas tão pequenas, mesmo uma minuciosa avaliação da imagem pode não ser suficiente para encontrá-las (GIGER, 2004).

Para facilitar a detecção de estruturas Winsberg (WINSBERG et al., 1967) utilizou computadores para processar as densidades ópticas de imagens mamográficas e correlacionar diferenças bruscas com áreas suspeitas, produzindo o primeiro sistema de diagnóstico auxiliado por computador em imagens mamográficas.

Apesar de tentativas como a de Winsberg, até o fim da década de 80 os sistemas iniciais de CAD, que tinham como objetivo separar tumores malignos de benignos, sofriam com precisões muito baixas se comparados a observadores humanos experientes (CHAN et al., 1998).

O desenvolvimento durante a década de 90 de sistemas radiológicos mais modernos e técnicas mais avançadas de processamento de imagens tornaram os sistemas de CAD uma importante ferramenta para o (DAVIES; DANCE, 1990) tanto para a detecção quanto para a classificação de microcalcificações (QIAN, 1995; JIAN; NISHIKAWA; PAPAIOANNOU, 1998; CHENG et al., 2003; KALLERGI, 2004), nódulos e tumores (LAI; LI; BISCHOF, 1989; BRZAKOVIC; LUO; BRZAKOVIC, 1990; HADJIISKI; CHAN; SAHINER, 2004).

Em 1998 o R2 Technology's Image Checker®(U. S. FOOD AND DRUG ADMINISTRATION, 1998) foi o primeiro sistema de CAD mamográfico aprovado pelo FDA. Este sistema foi desenvolvido para localizar sinais que indiquem câncer de mama. Massas são marcadas com asteriscos e aglomerados de microcalcificações com triângulos(SAMPAT; MARKEY; BOVIK, 2005).



Em 2002 mais dois esquemas CAD para mamografias foram aprovados: MammoReader® da Intelligent System Software Inc. (ISSI)(U. S. FOOD AND DRUGS ADMINISTRATION, 2002) e SecondLook da CADx Medical Systems (U. S. FOOD AND DRUGS ADMINISTRATION, 2002). Mantendo o sistema de marcar a imagem onde microcalcificações e massas foram detectadas, mas utilizando símbolos diferentes de acordo com a padronização do sistema.

Estes sistemas, assim como outros algoritmos e técnicas não comerciais, têm como intuito auxiliar o radiologista em seu trabalho, funcionando como uma (SAMPAT; MARKEY; BOVIK, 2005) e aumentar a eficiência dos exames mamográficos, diminuindo a quantidade de erros de diagnóstico (incluindo biópsias desnecessárias) e a quantidade de mortes (VYBORNY; GIGER; NISHIKAWA, 2000; GIGER, 2004).

## 2.2 Avaliação de CAD

Para avaliar técnicas de esquemas CAD Vyborny (VYBORNY; GIGER; NISHIKAWA, 2000) recomenda utilizar três fatores:

- **Sensibilidade (S)**: é a porcentagem de detecções positivas do sistema corretamente localizadas (verdadeiros-positivos), ou seja, o número de verdadeiros-positivos (VP) dividido pela soma do número de casos verdadeiro-positivos (VP) mais falso-negativos (FN), conforme especificado na equação (2.1).

$$S = \frac{VP}{VP + FN} \quad (2.1)$$

- **Especificidade (E)**: é a porcentagem de detecções negativas do sistema corretamente acertadas (verdadeiros-negativos), ou seja, o número de verdadeiros-negativos (VN) dividido pela soma do número de casos verdadeiro-negativos (VN) mais falso-positivos (FP), conforme especificado na equação (2.2).

$$E = \frac{VN}{VN + FP} \quad (2.2)$$

- **Natureza da base de dados**: indica o conjunto de imagens utilizado para a análise estatística do sistema. Este fator inclui características físicas da imagem

(resolução espacial, nível de contraste, tamanho da imagem, etc.) e origem das mesmas. Testes indicam que este fator pode causar grandes diferenças nos fatores acima e no resultado final (SCHIABEL et al., 2000).

Esta avaliação pode ser feita selecionando-se um conjunto de imagens com laudos conhecidos e contando a quantidade de verdadeiros-positivos e de falsos-positivos resultantes do sistema. É importante lembrar que estas duas características devem ser avaliadas juntas, ou seja, não adianta um sistema ter sensibilidade de 100%, mas apresentar muitas detecções falsas em imagens benignas.

No entanto, um programa de CAD ao ser calibrado (durante a construção do programa ou pelo usuário final, quando possível) exige que seja feita uma escolha do nível de sensibilidade em relação à especificidade que mais vale a pena para o objetivo do programa.

### 2.3 Curvas ROC

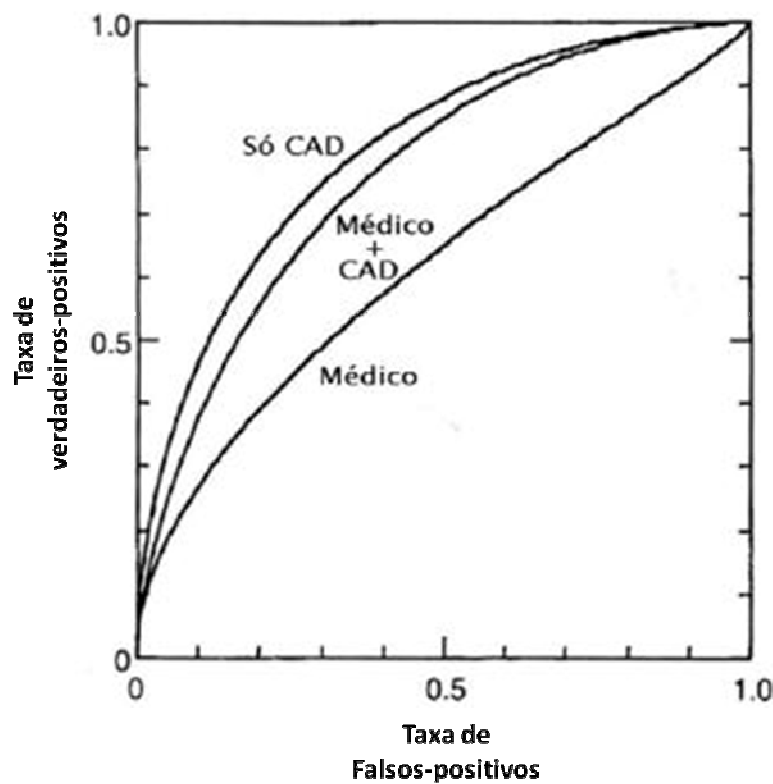
Para facilitar esta análise, normalmente são utilizadas curvas ROC (*Receiver Operating Characteristic*) (CHAN et al., 1999; HADJIISKI; CHAN; SAHINER, 2004; NUNES; SCHIABEL; BENNATI, 2002). O método dessas curvas foi desenvolvido para avaliar detecção de sinais em radar e na psicologia sensorial, pois é um método que leva em conta o aspecto subjetivo envolvido em um determinado evento. Na prática a curva ROC é uma curva relacional entre o número de verdadeiros-positivos contra o número de falsos-positivos para um determinado evento. No caso da análise de CAD mamográfico a curva ROC representa a redução da especificidade necessária para aumentar a sensibilidade do programa.

O exemplo mostrado na (Figura 1) ilustra curvas ROC comparando 3 casos: médico sem auxílio, médico com auxílio de um esquema CAD e o esquema CAD sozinho. De acordo com os autores do estudo (VYBORNÝ; GIGER; NISHIKAWA, 2000) o esquema CAD somente apresenta resultados melhores do que o conjunto médico + CAD porque os radiologistas selecionados não tinham experiência prévia com CAD e tiveram dificuldade em adaptar seu treinamento ao uso do sistema. Estudos posteriores (LUO; QIAN; ROMILLY, 2005) indicam que o treinamento no uso dos esquemas CAD amplia consideravelmente a qualidade dos resultados.

Tendo em mãos a curva ROC de um programa é recomendado que o mesmo seja calibrado para o ponto em que a sensibilidade seja a maior possível para uma ocorrência

baixa de falsos-positivos. Além disso, a área sob a curva ROC pode ser utilizada como um índice para avaliação de um esquema CAD, pois quanto maior for a área embaixo da curva melhor é a taxa de acertos do método avaliado, até o máximo ideal de 1 (onde o sistema tem sensibilidade de 100% e nenhum falso-positivo).

Com as informações da curva ROC a comparação de esquemas CAD parece simples, no entanto estas curvas, apesar de importantes, ignoram o terceiro fator que Vyborny (VYBORNY; GIGER; NISHIKAWA, 2000) requer para avaliação: Natureza da base de dados. Este último elemento será discutido nos próximos capítulos.



**Figura 1: Desempenho de um esquema CAD para detectar e classificar microcalcificações comparados ao desempenho dos radiologistas sem auxílio (VYBORNY; GIGER; NISHIKAWA, 2000)**

## Capítulo 3: Banco de dados

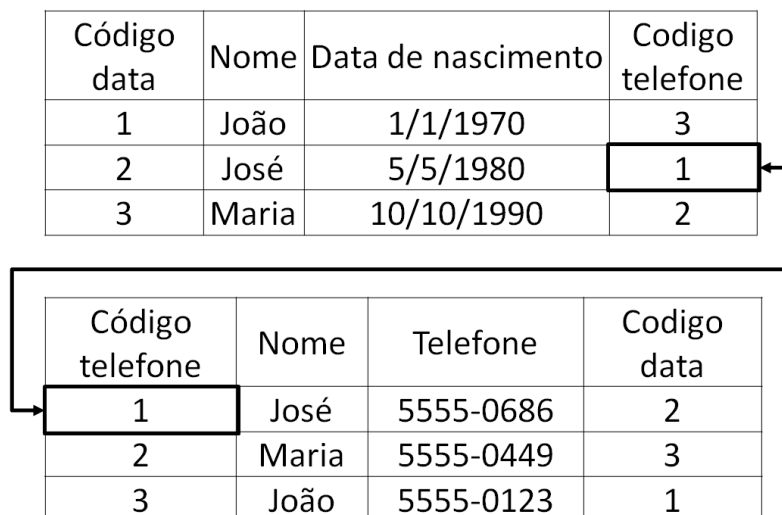
### 3.1 Definição de banco de dados

Um banco de dados (ou base de dados) pode ser definido como um conjunto de registros dispostos em uma estrutura regular que possibilita reorganização e extração de informações. Os registros podem ser qualquer tipo de dados: números, nomes, endereços ou mesmo imagens.

Estes registros podem ser armazenados de variadas maneiras, chamadas modelos de dados. Um modelo de dados define como os registros vão ser armazenados no banco e como eles vão ser associados entre si. Apesar de antigamente existirem vários modelos de dados, a melhora na capacidades dos computadores fez com que apenas um modelo ainda seja usado em larga escala: o modelo relacional.

### 3.2 Modelo relacional

O modelo relacional (CODD, 1970) estrutura os dados em tabelas onde cada linha é um registro e cada coluna um atributo do registro. Em todas as tabelas é necessário estabelecer um subconjunto das colunas como sendo chave, que deve ser única para cada registro de uma mesma tabela. Para relacionar um registro ao outro, basta que um deles tenha colunas com a mesma chave do outro (Figura 2).



**Figura 2: Exemplo de tabela de modelo relacional**

Qualquer tipo de informação pode ser armazenada em sistemas como este, bastando que o sistema de gerenciamento garanta que as chaves estão corretamente associadas.

Este modelo permite muito mais liberdade para construir o banco de dados e fazer suas relações, porém a complexidade de seu desenvolvimento consideravelmente maior que o de outros modelos antigos. Por conta disso, apesar do modelo ter sido apresentado em 1969 (CODD, 1970), apenas na década de 80 a tecnologia se desenvolveu o suficiente para que o modelo se tornasse viável.

Para manejar as informações dentro de um banco de dados se utiliza um linguagem própria chamada linguagem de consulta.

### 3.3 Linguagem de consulta

A linguagem de consulta de um banco de dados é a linguagem utilizada para realizar a comunicação entre o usuário e o sistema de gerenciamento. Esta linguagem precisa permitir pelo menos três comandos básicos para uso de uma base de dados (CODD, 1970):

- **Inserção:** Comando para inserir um registro no banco. Dependendo do modelo utilizado pode exigir informações sobre registros pais, filhos e/ ou chaves.

- **Seleção:** Comando que permite escolher um ou mais registros para leitura ou processamento. Este comando seleciona uma parte da estrutura do banco, seja um ramo da árvore hierárquica, um conjunto de grafos ou algumas linhas da tabela e reconstrói como sendo um banco separado.

- **Remoção:** Comando para remoção de um registro do banco. Este comando necessita da identificação dos registros a serem removidos utilizando previamente o comando Selecionar. O uso combinado do comando Remover e Inserir pode gerar um comando de atualização se os dados forem corretamente preservados.

A linguagem de consulta também deve garantir a segurança do sistema, exigindo acesso autorizado e permitindo que apenas usuários específicos sejam capazes de inserir, remover ou mesmo acessar certas informações privilegiadas.

Caso a base tenha como objetivo permitir inserção por múltiplos usuários é importante que o processo de inserção e remoção seja rastreável, mas isso é considerado responsabilidade do desenvolvedor dos aplicativos que estejam utilizando a base e não da linguagem de consulta, apesar dela fornecer as ferramentas para tal.

### 3.4 SQL

Do mesmo modo que o modelo relacional se tornou tão comum que pode ser dito universal, o mesmo acontece com a linguagem de consulta mais comum para seu manejo (BERNSTEIN; NEWCOMER, 2009), atualmente praticamente a única linguagem usada para o modelo relacional é SQL (CHAMBERLIN; BOYCE, 1974).

Inicialmente nomeada SEQUEL, a linguagem SQL foi desenvolvida para manipular a base relacional projetada pela IBM, chamada System R e comprovou a funcionalidade do modelo relacional para bancos de dados (SELINGER et al., 1979). Inicialmente concorrendo com a linguagem QUEL do MIT, o SQL se tornou a única linguagem para bancos relacionais amplamente utilizada (OPPEL, 2004).

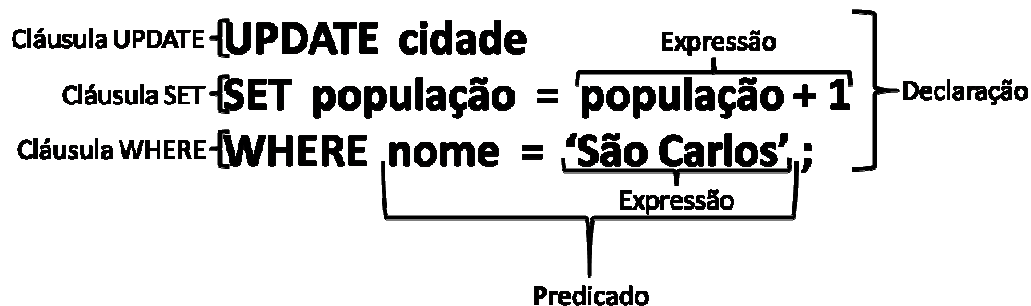


Figura 3: Anatomia de uma declaração exemplo em SQL

Como mostrado na Figura 3, a linguagem é subdividida em vários elementos (ISO/IEC 9075(1-4,9-11,13,14):2008), sendo os principais:

- **Declarações** (*Statements*): São requisições para alterar estruturas do banco de dados, controlar transações, conexões ou realizar diagnósticos.
- **Consultas** (*Queries*): São pedidos de informações ou atividades ao SGBD sem (em geral) causar alterações a base. São compostas por uma ou mais declarações.
- **Cláusulas** (*Clauses*): São os componentes básicos que formam as declarações e consultas. Uma cláusula pode conter palavras restritas com instruções ao SGBD, expressões, predicados e informações necessárias ao processo.
- **Expressões** (*Expressions*): Podem produzir valores escalares, booleanos, *strings* ou mesmo tabelas ou partes delas.
- **Predicados** (*Predicates*): São estruturas lógicas que serão avaliadas pelo SGBD. Um sistema SQL permite que a resposta seja verdadeiro, falso ou nulo (*True*,

*False* ou *Null*) utilizando um sistema de lógica de três valores (3VL). A resposta permite selecionar e limitar os efeitos das declarações e consultas.

Espaços em branco além do mínimo separador são ignorados pelo sistema.

A linguagem de consulta SQL é definida no padrão SQL assim como todo um conjunto de informações pertinentes ao uso do SQL: como usar o modelo relacional, estruturação de dados como definido em SQL/Schemata (ou Information and Definition Schemas), mecanismo de transação estruturado na própria linguagem e, obviamente, o uso da linguagem SQL. Este padrão teve sua atualização mais recente em 2008 (*ISO/IEC 9075(1-4,9-11,13,14):2008*).

Apesar do padrão SQL ter si tornado praticamente universal, vários grupos de pesquisa e empresas interpretaram o padrão de modo diferente, criando os chamados dialetos, ou seja, variações da linguagem que, apesar de seguir as regras do padrão SQL, são diferentes em estrutura e/ou uso e não são compatíveis entre si.

Abaixo são descritos alguns dialetos estudados para este trabalho:

### **3.4.1 Oracle**

O *software* e dialeto Oracle foram desenvolvidos em 1978 e lançados em 1979 pela Oracle Corporation (na época chamada de Relational Software, Inc.), sendo o primeiro sistema comercial de gerenciamento de base relacional (ORACLE CORPORATION, 2007).

Oracle apresenta a estrutura SQL de modo bastante completo, implementando quase todas as recomendações e características do padrão SQL (ORACLE CORPORATION, 2009). Além disso, apresenta uma grande quantidade de ferramentas, estruturas e comandos únicos, ampliando as capacidades do sistema e dando uma grande liberdade ao desenvolvedor.

Por ser um sistema comercial com amplo uso em empresas de grande porte, a segurança é uma prioridade, contendo, mesmo em suas estruturas mais básicas, sistemas redundantes, com o intuito de impedir que a informação seja alterada incorretamente ou perdida. Infelizmente, estas estruturas e ferramentas tendem a tornar as bases Oracle maiores e computacionalmente mais pesadas.

A grande quantidade de ferramentas e opções disponíveis exige que o usuário tenha um grande conhecimento prévio do funcionamento do sistema Oracle para poder

configurá-lo e utilizá-lo corretamente. Isto tende a dificultar o uso para desenvolvedores iniciantes ou de projetos de pequeno e médio tamanho (PETRI, 2005).

Para facilitar o uso do seu sistema, a Oracle Corporation lançou um conjunto de aplicativos para auxiliar a configuração, que recebeu o nome de *Oracle Developer Suite*. Os aplicativos incluem sistemas para lidar diretamente com a base assim como para auxiliar a criação de telas para seu uso.

Em 2005 a Oracle lançou o *Oracle Database XE*, que é uma versão gratuita daquele sistema. Esta versão é mais restrita em termos de ferramentas e não oferece suporte pela empresa, contando apenas com fóruns em que outros desenvolvedores podem corresponder. Também não permite bases de mais de 4 GB de informações e não utiliza mais do que 1GB de RAM, independentemente da quantidade de memória do servidor.

Oracle se tornou o sistema mais utilizado por grandes empresas e projetos, mas o fato de ser um sistema pago (ou ter restrições no uso) e da dificuldade em aprender e utilizar as ferramentas (mesmo com a ajuda do conjunto de aplicativos) torna-o de uso restrito a projetos maiores.

### 3.4.2 MySQL

Enquanto o sistema Oracle é um sistema proprietário de grande porte focado em grandes empresas, o MySQL é um sistema *open-source* (pela *GNU General Public License*) (STALLMAN, 2007) focado no uso por pequenas empresas e projetos, tendo sua estrutura voltada para velocidade e facilidade de utilização.

Criado em 1995 por Michael Widenius e David Axmark (WIDENIUS; AXMARK, 2010), MySQL apresenta grande parte das características do SQL implementadas ou prevista para implementação em próximos lançamentos (MYSQL AB, 2008). Certos elementos SQL foram propositalmente não implementados para aumentar a velocidade ou para facilitar o uso do sistema. MySQL às vezes é criticado por não seguir à risca o padrão SQL (PETRI, 2005).

O foco principal do MySQL são sistemas de pequeno e médio porte voltados para *web*. Como, em geral, a grande preocupação de aplicações *web* é a velocidade de resposta, este sistema tende a sacrificar a amplitude de uso e segurança – comparado ao Oracle – em troca de velocidade e facilidade de uso. Apesar de ser considerado menos



seguro e não tão robusto, MySQL frequentemente está à frente na implementação de novas técnicas.

Suas características estruturais e a fácil integração com PHP (ACHOUR et al., 2010) tornou o sistema comum em sites, mesmo os de grande porte como Google, Wikipedia e Facebook(WIDENIUS; AXMARK, 2010), tornando MySQL o mais utilizado sistema *open-source* para gerenciamento de bases relacionais no mercado.

## Capítulo 4: Bases de imagens mamográficas

### 4.1 Características de Bancos de imagens médicas

Um banco de imagens médicas pode ser utilizado e construído de várias maneiras. Atualmente a grande maioria dos hospitais tem seus arquivos físicos associados a sistemas computacionais para facilitar o processo de arquivamento e localização das informações. Este sistema inclui exames com imagens em filme e laudos provenientes de diferentes equipamentos, como raios X convencionais, mamógrafos, *CT*, *MRI*, etc. Alguns hospitais e clínicas até provêm as imagens digitalmente, seja por mídia física (em geral CD) ou *online* com uma senha exclusiva do paciente e médico que requisitou.

Quando utilizado para desenvolvimento de programas CAD, é necessário que o banco seja capaz de lidar de modo seguro com uma grande quantidade e variedade de informações. Além das imagens propriamente ditas, o sistema precisa ser capaz de armazenar dados sobre o exame realizado, o laudo resultante, paciente (no caso de bases hospitalares, estas informações podem ser confidenciais), características dos equipamentos utilizados (equipamento de obtenção de imagem, digitalizador quando necessário, etc.) e características específicas da imagem (resolução, profundidade de contraste, tamanho, etc.).

Testes apresentados na literatura já comprovaram que o desempenho de esquemas CAD é consideravelmente influenciado pela quantidade e qualidade das imagens na base (SCHIABEL et al., 2000; SCHIABEL et al., 2001). Isto significa que um banco de dados precisa conter um conjunto de imagens consistente, confiável e suficientemente amplo de modo a não exercer influência sobre os resultados dos sistemas testados.

### 4.2 Requerimentos de Bases mamográficas

O desenvolvimento de um banco de dados único padrão para desenvolvimento e testes de sistemas CAD mamográficos vem sendo buscado desde a década de 90, mas estudos (NISHIKAWA, 1994; NISHIKAWA, 1998; ELTER et al., 2007) indicam que não existe nenhuma base de imagens mamográficas com a capacidade para ser utilizada universalmente. Para este uso a base deve ter o conjunto de características discutida a seguir.

#### 4.2.1 Conjuntos de imagens

É importante para uma base de imagens mamográficas que ela contenha não só os arquivos das imagens, mas também os laudos radiológicos e, quando possível, acompanhamento em futuros exames e/ou confirmação por biópsia. Se os laudos não estiverem disponíveis ou não forem confiáveis, a base perde o sentido para testes de CAD.

As imagens devem ter alto contraste, ou seja, devem apresentar resolução de contraste de pelo menos 10 bits (1024 níveis de cinza), de acordo com Nishikawa e Schiabel (NISHIKAWA, 1998; SCHIABEL et al., 2000). Contrastes menores não contêm informação suficiente para análises detalhadas em CAD, porém profundidades maiores do que 14 bits (16.384 níveis de cinza) não parecem causar diferenças significativas no comportamento da maioria dos programas CAD (SCHIABEL et al., 2000).

Quando se lida com imagens digitais é necessário garantir que a resolução espacial seja suficiente para que as imagens sejam consideradas de boa qualidade. Normalmente este valor, em imagens mamográficas digitais ou digitalizadas, varia entre 0,035mm e 0,200mm. Nunes (NUNES et al., 2001) mostrou que o uso de resoluções melhores do que 0,100mm tendem a aumentar a quantidade de falsos-positivos sem aumento perceptível na sensibilidade, mas para percepção humana acurada a literatura (NISHIKAWA, 1998) recomenda entre 0,030mm e 0,060mm. É importante notar que a resolução espacial de uma imagem está relacionada com o tamanho da mesma em pixels e espaço de memória; por consequência, imagens com alta resolução espacial tendem a consumir grandes quantidades de memória, podendo atrasar consideravelmente seu processamento.

Nishikawa (NISHIKAWA, 1998) recomenda que uma base de imagens mamográficas deve conter pelo menos 1.000 imagens para poder ser utilizada de modo eficiente. Também é necessário que a base tenha grande variedade de imagens, contendo diferentes tipos de achado em quantidade para que testes de sistemas de detecção de achados específicos possam ser facilmente testados. Na literatura, não existe uma quantidade recomendada de imagens para cada achado em uma base mamográfica, mas um bom ponto de referência parece ser a frequência do caso na literatura médica.

Para o uso da base ser universal, o formato das imagens deve ser de fácil e livre uso. Como as imagens não podem sofrer perdas por compactação, os formatos tendem a ser diferentes do usual, como por exemplo, formato TIFF, LJPEG (JPEG *lossless*), RAW, e outros. A escolha do formato utilizado pela base vai depender das ferramentas disponíveis pelo programador, mas é importante que o formato escolhido seja capaz de lidar com imagens de pelo menos dois bytes (9 a 16 bits) de contraste e exija pouco ou nenhum pré-processamento para a imagem ser utilizada.

#### **4.2.2 Interface**

É fundamental que a base ofereça acesso livre e gratuito para qualquer desenvolvedor a nível mundial que tenha interesse em utilizar a base para sua pesquisa. Para tal, ela precisa ser acessível via internet e ser simples para o usuário utilizá-la.

É importante lembrar que muitas informações envolvendo pacientes são confidenciais e por isso a segurança da base se torna de grande importância. É necessário garantir que nenhum usuário possa ter acesso a informações sem autorização.

Para facilitar o uso da base, é importante que o usuário seja capaz de encontrar facilmente conjuntos de imagens que supram a necessidade específica do seu programa, ou seja, a base precisa ter um sistema de busca e recuperação de dados eficiente e ampla. O principal foco das buscas seriam os tipos de achados, como microcalcificação, linfonodo ou fibrose, e características específicas da imagem, como resolução espacial ou resolução de contraste.

Obter facilmente imagens específicas facilita consideravelmente o processo de teste e avaliação do programa, já que torna desnecessário visualizar as informações uma a uma procurando pelo tipo correto de imagem.

### **4.3 Bases existentes**

Atualmente, as bases mais utilizadas por desenvolvedores são as seguintes: MIAS, DDSM, LLNL/UCSF e CALMa.

#### **4.3.1 MIAS database**

MIAS (*Mammographic Image Analysis Society*) (SUCKLING, 1994) é a mais antiga base ainda em uso, criada em 1994. Esta é uma pequena base, contendo apenas 320 imagens de 8 bits de contraste (256 níveis de cinza) e 0,050mm de resolução

espacial. O projeto da base MIAS não parece ter sido continuado e a base, sem nenhuma atualização praticamente desde seu lançamento, tornou-se ultrapassada, oferecendo apenas imagens de resolução de contraste insuficiente para sistemas CAD modernos.

As imagens nesta base estão estruturadas em uma única pasta de dados com um arquivo de texto explicando o laudo e a posição dos achados em código. Não existe nenhum sistema de busca, classificação BI-RADS<sup>1</sup> ou qualquer outra ferramenta para auxílio ao desenvolvedor.

Apesar de não recomendada e estar com seu *site* oficial fora do ar há anos, esta base ainda vem sendo utilizada até recentemente (ABUBAKER et al., 2006), especialmente para testes iniciais utilizando sua versão “mini-MIAS” (SUCKLING, 2003), onde todas as imagens são padronizadas em 0,200mm, 1024 x 1024 pixels centralizada na matriz. Esta versão, de ainda mais baixa qualidade, é utilizada por ser de baixo custo computacional, já que as imagens apresentam apenas 1 byte por pixel.

#### 4.3.2 DDSM

DDSM (*Digital Database for Screening Mammography*) (HEATH et al., 2001) foi publicada em 2001. Contendo mais de 10.000 imagens, esta é a maior base pública no momento. Por volta de 20% das imagens são de 16 bits, enquanto o resto é de 12 bits, com resolução espacial entre 0,042mm e 0,050mm.

A organização da base DDSM é um pouco melhor do que a MIAS. As imagens são divididas em 43 “volumes” com entre 80 e 450 imagens em cada um. Os volumes são classificados simplesmente como: normal, benigno, câncer e benigno sem retorno; todos os casos de um dado volume são do tipo especificado. Para mais detalhes, o pesquisador precisa selecionar a descrição de caso por caso até encontrar o tipo de achado que interessa a ele. Considerando a quantidade de imagens oferecidas pela DDSM, este trabalho pode ser extremamente demorado e exaustivo e um sistema de busca facilitaria consideravelmente o uso da base. Infelizmente, o sistema de busca apresentado no *site* (ROSE et al., 2003) não funciona e foi atualizado pela última vez em 2003.

Cada imagem da base DDSM inclui um arquivo anexo que delimita os achados como uma máscara para a imagem, definindo uma borda ordenada ao redor de cada

---

<sup>1</sup> BI-RADS (*Breast Image Reporting and Data System*) é um método de classificação numérica de imagens médicas relacionadas a câncer de mama apresentado pelo *American College of Radiology*, atualmente sendo utilizado como padrão pelo FDA (AMERICAN COLLEGE OF RADIOLOGY, 2003).

achado. Estas regiões foram delimitadas utilizando um sistema automático (HEATH et al., 2001), onde os casos falsos-positivos foram manualmente removidos. Está é uma das mais importantes características da base DDSM, facilitando consideravelmente o uso das imagens por desenvolvedores interessados.

O formato de arquivo escolhido pelos criadores da DDSM foi uma versão do formato LJPEG (*Lossless JPEG*) não-padrão, dificultando consideravelmente o uso da base, pois programas usuais da área, como MATLAB®, não suportam esta versão do LJPEG. A DDSM apresenta um conjunto de programas para conversão da imagem para o formato PNM, porém os programas são em C não compilado, exigindo que o desenvolvedor CAD não só saiba utilizar C como os utilize em uma plataforma LINUX, já que, pela afirmação do próprio criador dos programas, o sistema não está pronto para uso em WINDOWS®. Isto torna difícil o uso dos programas de conversão, mas nenhuma outra opção é oferecida pela DDSM.

Também é importante ressaltar que o *download* de imagens na base DDSM precisa ser feito individualmente, navegando pelo sistema FTP da base, tendo apenas os códigos numéricos como guia. Novamente, a quantidade de imagens oferecida requer um sistema mais eficiente para localização e *download* das imagens desejadas.

#### 4.3.3 LLNL/UCSF

LLNL/UCSF (*Lawrence Livermore National Laboratories / University of California at San Francisco*) (LAWRENCE LIVERMORE NATIONAL LIBRARY) é uma base paga (US\$ 100), contendo 198 imagens de 12 bits de resolução de contraste e 0,035mm de resolução espacial. Em termos de resolução espacial, estas são as melhores imagens disponíveis *online* (<http://marathon.csee.usf.edu/Mammography/OtherResources.html>). O foco desta base são microcalcificações em clusters e espalhadas.

As imagens incluem um conjunto de casos variando de normais, normais com dificuldade (mama densa, implantes ou tecido assimétrico), microcalcificações obviamente benignas, microcalcificações benignas suspeitas e microcalcificações malignas (apesar de descritas deste modo no *link* acima é importante lembrar que microcalcificações não são benignas ou malignas, mas podem estar associadas a tumores benignos ou malignos). Todas as imagens incluem resultados de biópsia ou, pelo menos, três anos de acompanhamento posterior à obtenção da primeira imagem.

O formato dos arquivos das imagens é desconhecido, pois a base não foi adquirida para tal avaliação e esta informação não está disponível em seus artigos ou seu *site*.

O fato desta base ser paga torna seu uso mais difícil e muitas vezes inaceitável, considerando as outras bases disponíveis gratuitamente.

#### **4.3.4 CALMa**

CALMa (*Computer-Assisted Library for Mammography*) (AMENDOLIA et al., 2001) é uma base privada contendo 3.000 imagens, desenvolvida para produção do esquema CAD publicada em associação ao projeto CALMa. As imagens presentes na base são de 12 bits de resolução de contraste e 0,085mm de resolução espacial.

O projeto CALMa foi financiado pelo INFN (*Instituto Nazionale di Fisica Nucleare* – Instituto Nacional de Física Nuclear, Itália) em 1997 com o intuito de criar uma base italiana de imagens e um esquema CAD para detecção de microcalcificações e massas. Este projeto foi concluído em 2001 desenvolvendo algumas novas técnicas (FANTACCI et al., 2002; LAURIA et al., 2004). O projeto foi dado continuidade durante 2002 a 2003 sendo renomeado para projeto GPCALMA e depois de 2004 a 2006 como MAGIC-5, que manteve e deu continuidade ao desenvolvimento da base.

A estruturação da base, assim como o tipo de arquivo utilizado, não é conhecida.

Seu uso é exclusivo das Universidades associadas ao projeto CALMa, não permitindo o acesso de nenhum desenvolvedor público, apesar do projeto MAIGC-5 ter ampliado seu uso para os hospitais associados ao projeto, nunca houve previsão de tornar a base pública.

#### **4.4 Considerações finais**

As bases descritas acima não mais apresentam suporte técnico ou atualizações há alguns anos, pois seus respectivos projetos foram fechados. Isso implica que não existe auxílio para o usuário.

É fácil notar que nenhuma das bases descritas, apesar de populares, preenche os requisitos de uma base universal padrão.

## Capítulo 5: Metodologia

### 5.1 Histórico do projeto

Os trabalhos que levam ao projeto de que trata esse texto tiveram uma primeira publicação em 2001 (BENATTI; SCHIABEL; NUNES, 2001), descrevendo o desenvolvimento de uma base relacional (local) para uso dos pesquisadores do LAPIMO<sup>2</sup>, buscando facilitar o arquivamento de um conjunto de mais de 7.400 imagens adquiridas ao longo de cinco anos de pesquisa na área.

Esta base foi desenvolvida para ser um banco físico portátil, composto por 30 CDs contendo imagens (por volta de 3.300 imagens) e um programa desenvolvido em *Delphi* com uma biblioteca SQL para construção e manipulação da base de dados.

Esta estrutura permitia ao usuário realizar buscas pelo hospital de origem, idade do paciente no momento do exame e estruturas de interesse, podendo realizar a busca por qualquer combinação destes elementos.

Os resultados são apresentados em uma lista de exames que, quando selecionados, mostra o conjunto de quatro imagens do exame (em uma miniatura de 8 bits de contraste, apenas para visualização e não para processamento), o laudo médico e outras características pertinentes da paciente e dos equipamentos utilizados para obtenção da imagem e digitalização. O usuário era então oferecido a opção de copiar a imagem completa, recortar uma região de interesse ou selecionar um novo exame para visualização.

Este programa era um executável construído em DELPHI, portanto funcionava apenas em plataformas Windows. Este primeiro projeto foi concluído em 2003, com uma dissertação de Mestrado (BENATTI, 2003)

Em continuidade – dada a dinâmica do processo em si – foi avaliada a possibilidade de ampliar a base local para um sistema *online*. O projeto inicial seria a adaptação direta do sistema desenvolvido previamente, só que então para acesso *online*. Porém, uma vez que seria necessário reescrever toda a estrutura do programa para torná-lo compatível com um sistema voltado para redes, decidiu-se ampliar as capacidades da estrutura original.

---

<sup>2</sup> LAPIMO (Laboratório de Análise e Processamento de Imagens Médicas e Odontológicas) é um laboratório de pesquisas do Departamento de Engenharia Elétrica da Escola de Engenharia de São Carlos da Universidade de São Paulo. Este trabalho foi completamente desenvolvido neste laboratório.



Apesar da idéia básica da base física ter sido mantida, ou seja, as tabelas fundamentais e as relações entre elas terem sido conservadas na transição, a estrutura de *software* da base teve que ser reconstruída e novas tabelas adicionadas para poder ser utilizada de modo seguro via *web*, incluindo a possibilidade de inserção de imagens e informações remotamente. Mesmo com as diferenças de estrutura e plataforma, é importante ressaltar que a base continua mantendo ferramentas de busca rápidas e de fácil uso.

Entre o fim de 2003 e 2004 a base foi aberta ao público e testada (NUNES et al., 2003; NUNES et al., 2004), sendo batizada de “BancoWeb” (NUNES et al., 2004).

A base gerou interesse de pesquisadores na área de desenvolvimento de CAD, pesquisas médicas e ensino de radiologia. A opinião destes usuários foi coletada – sem nenhum protocolo específico – pelo envio de e-mail para os administradores. Estas avaliações foram levadas em conta e foi decidido que o projeto, apesar de ter continuado no ar, precisava de uma reformulação extensiva.

Uma vez reestruturada, mas ainda conservando características da base física original, ela foi reapresentada em 2006 (SCHIABEL; ESCAPINARTI; FREITAS, 2006), com um formato mais próximo do modelo atual. Novamente opiniões de usuários (tanto os inscritos na primeira versão quanto novos usuários) foram colhidas e foi dada continuidade ao desenvolvimento do projeto (FREITAS et al., 2007; MATHEUS; SCHIABEL, 2008).

Apesar de oficialmente aberta, a base continua sendo desenvolvida, com novas ferramentas sendo acrescentadas e erros corrigidos, conforme a necessidade se apresenta. É importante que a base tenha o seu desenvolvimento continuado para mantê-la atualizada e relevante.

## **5.2 Aquisição das imagens**

O processo de aquisição das imagens da atual base foi realizado ao longo de anos para fornecer ao grupo de pesquisa material necessário ao desenvolvimento de esquemas CAD. Essas imagens foram obtidas através de convênio de colaboração principalmente com os seguintes hospitais: Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto (HCFMRP-USP), Santa Casa de Misericórdia de São Carlos, Hospital São Paulo e, mais recentemente, Hospital das Clínicas da Faculdade de Medicina da UNESP, em Botucatu/SP. As imagens inseridas na versão atual são em

parte da base física original e em parte imagens digitalizadas posteriormente. Até a presente data o BancoWeb conta com apenas imagens do Hospital São Paulo (1222 imagens) e do Hospital das Clínicas de Botucatu (215 imagens).

A escolha de obter imagens de vários hospitais tinha como intuito aumentar a variedade de imagens e de equipamentos utilizados. Deve-se lembrar que, independentemente do número de imagens em uma base, é importante que ela contenha grande variedade de casos.

Um fator importante ao obter as imagens é conhecer o mamógrafo utilizado na época que o exame foi realizado. Informações úteis sobre os aparelhos incluem conhecer a marca, modelo e tamanho do ponto focal.

O tamanho do ponto focal é, normalmente, fornecido pelo fabricante do aparelho, mas preferencialmente deve ser medido no ponto central do campo e a partir disto calculado por toda a extensão do campo. O tamanho do ponto focal de um mamógrafo vai ajudar a avaliar a capacidade de detecção do aparelho, pois quanto menor for o ponto focal mais detalhes serão visíveis e quanto maior mais borrada a imagem se torna.

Atualmente é possível utilizar o conhecimento do ponto focal do aparelho para corrigir distorções e tornar imagens mais nítidas (VIEIRA, 2005).

Na Tabela 5.1 encontra-se a lista de mamógrafos e seus respectivos hospitais de origem conforme determinado durante o processo de aquisição das imagens. No caso do HCFMRP-USP os casos obtidos com um histórico mais longo provieram de diferentes mamógrafos, já que o hospital trocou seus aparelhos durante o período.

**Tabela 5.1: Lista de mamógrafos, local de origem e ponto focal**

<b>Mamógrafo</b>	<b>Local</b>	<b>Tamanho do ponto focal (mm)</b>
<b>Philips Mammodiagnost</b>	HCFMRP-USP	1,20 x 0,75
<b>CGR Senographe 500 T</b>	HCFMRP-USP	1,50 x 0,90
<b>GE Senographe DMR</b>	HCFMRP-USP	0,45 x 0,37
<b>CGR Senographe 600 T</b>	Hosp. S. Paulo	0,68 x 0,54
<b>CGR Senographe 600 T</b>	HC de Botucatu	0,68 x 0,54
<b>Lorad M III</b>	Sta Casa S. Carlos	0,60 x 0,40

Outras características que influenciam na qualidade da imagem resultante e, por isso, nos processamentos dizem respeito ao digitalizador (*scanner*) utilizado para digitalizar imagens coletadas em filme. O digitalizador irá definir a resolução espacial e de contraste da imagem digital final, ambos fatores extremamente importantes para avaliação da qualidade da imagem, como previamente comentado.

Para a digitalização de filmes mamográficos é necessário um equipamento especializado, pois os *scanners* convencionais funcionam por reflexão de luz no papel não sendo recomendados para a varredura de filmes radiográficos. Para digitalizá-los, recomenda-se varredura de transmissão, ou seja, um sistema que detecte a luz que passa através filme (em vez de refletida) e gere a imagem digital.

As imagens inseridas na base foram digitalizadas usando um de dois *scanners* a laser da marca Lumisys (Lumiscan 50 e Lumiscan 75), ambos capazes de uma resolução de contraste de 12 bits (4.096 níveis de cinza) segundo o fabricante. Testes indicaram que ambos são sistemas que produzem imagens realmente de 12 bits, embora limitados a 3.500 (Lumiscan 50) e 3.800 (Lumiscan 75) níveis de cinza (ESCARPINARTI, 2002). Ainda assim, foi comprovado que 3.600 níveis de cinza não geram perdas significativas nas imagens, pois os objetos importantes para o processamento mamográfico são estruturas claras, definidas nos níveis de cinza próximos ao zero.

Quanto à resolução espacial, os *scanners* apresentam alguma diferença, já que o Lumiscan 50 é capaz de digitalizar imagens com *pixels* de 0,150mm e o Lumiscan 75, com *pixels* de 0,085mm, quando operando com filmes típicos de 18 x 24cm.

Em geral os digitalizadores permitem que o tipo de arquivo final possa ser escolhido pelo usuário, a partir de formatos típicos. No caso desta base, foi dada preferência ao formato TIFF (ADOBE DEVELOPERS ASSOCIATION, 1992) – como, inclusive, recomendava trabalho prévio desenvolvido na equipe (ESCARPINARTI, 2002).

O formato TIFF apresenta grandes vantagens sobre seus concorrentes por ser um formato com suas características determinadas por *tags* (*Tag Image File Format*). Por isso, pode ser adaptado a variados usos através de um cabeçalho (*header*) detalhado e uma estrutura chamada IFD (*Image file directory*) que conta os detalhes da imagem, como tamanho, resolução de contraste, esquemas de compressão utilizados, resolução espacial (no eixo X e Y), entre outras informações.

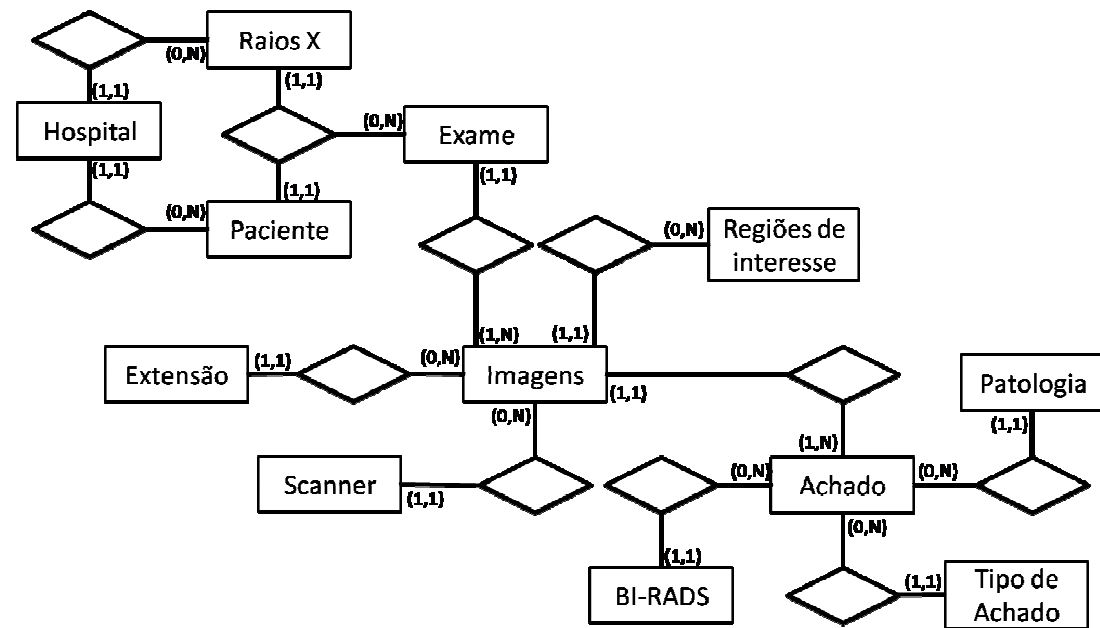
Este formato é bem documentado e apresenta uma grande quantidade de ferramentas e bibliotecas para leitura e processamento das imagens. Apesar da preferência pelo TIFF, a base está estruturada para cadastrar e armazenar imagens em qualquer outro formato, requerendo mínimas modificações.

### 5.3 Estruturação do banco

Baseando-se nas informações obtidas através dos laudos, experiência com desenvolvimento de CAD do grupo e em contato com radiologistas e ginecologistas, foram decididas quais informações seriam importantes para estruturar a base, assim como quais informações deveriam compor o sistema de busca.

O modelo relacional seguido foi baseado naquele elaborado inicialmente para a versão física desta base (BENATTI, 2003), ampliado para acomodar as novas ferramentas e capacidades, incluindo a possibilidade de acesso *online*. Na Figura 4 está apresentado o modelo, onde os retângulos escuros são tabelas com informações de configuração da base e as setas indicam a relação. Na próxima seção serão discutidas em mais detalhes as informações de cada tabela.

Apesar das vantagens de segurança e amplas ferramentas apresentadas pelo sistema Oracle, a dificuldade de configuração e uso e o fato do sistema ser pago tornou o sistema MySQL mais interessante para o uso da base. Foi importante para esta decisão o foco *web* do sistema e a preferência pelo uso apenas de ferramentas *open-source*.



```

    v ladinew.achado
    # achado_codigo : int(8)
    # exame_numero_registro : int(10)
    # imagem_nome : varchar(60)
    # achado_localizacao : varchar(30)
    # achado_birads : int(2)
    # coordenada_xse : varchar(10)
    # coordenada_yse : varchar(10)
    # coordenada_xid : varchar(10)
    # coordenada_yid : varchar(10)
    # achado_tipo : varchar(120)
    # achado_tipo_eng : varchar(120)
    # achado_patologia : varchar(60)
    # achado_patologia_eng : varchar(60)
    
```

```

    v ladinew.usuario
    # usuario_codigo : int(8)
    # usuario_login : varchar(20)
    # usuario_nome : varchar(60)
    # usuario_data_nascimento : varchar(10)
    # usuario_escolaridade : varchar(40)
    # usuario_profissao : varchar(60)
    # usuario_interesse : longtext
    # usuario_justificativa : longtext
    # usuario_email : varchar(60)
    # usuario_senha : varchar(20)
    # usuario_nivel : int(2)
    
```

```

    v ladinew.paciente
    # codigo_auxiliar : int(8)
    # paciente_codigo : varchar(12)
    # hospital_codigo : varchar(30)
    # paciente_data_nascimento : varchar(10)
    # paciente_sexo : varchar(10)
    # paciente_telarca : varchar(12)
    # paciente_lactacao : varchar(12)
    # paciente_cancer_familia : varchar(12)
    # paciente_anticoncepcional : varchar(12)
    # paciente_rep_hormonal : varchar(12)
    # paciente_nodulo : varchar(12)
    
```

```

    v ladinew.imagem
    # scanner_codigo : int(10)
    # imagem_posicao : varchar(20)
    # imagem_birads : int(2)
    # exame_numero_registro : int(10)
    # exame_codigo : int(8)
    # imagem_nome : varchar(60)
    # imagem_pixel : varchar(10)
    # imagem_tipo : varchar(30)
    # imagem_qtde_bits : int(3)
    # imagem_mama : varchar(60)
    # imagem_achado : varchar(10)
    
```

```

    v ladinew.hospital
    # hospital_codigo : int(8)
    # hospital_nome : varchar(30)
    # hospital_endereco : varchar(40)
    # hospital_complemento : varchar(30)
    # hospital_bairro : varchar(30)
    # hospital_CEP : varchar(8)
    # hospital_telefone : varchar(14)
    # hospital_fax : varchar(14)
    # hospital_contato : varchar(30)
    # hospital_radiologista : varchar(60)
    
```

```

    v ladinew.tipoachado
    # codigo : int(3)
    # tipo_de_achado : varchar(100)
    # tipo_de_achado_eng : varchar(100)
    
```

```

    v ladinew.recorte
    # Codigo : int(5)
    # Nome : varchar(100)
    # Nome_original : varchar(100)
    # Ext : varchar(10)
    # Pos_X : int(6)
    # Pos_Y : int(6)
    # Altura : int(6)
    # Largura : int(6)
    # Motivo : text
    
```

```

    v ladinew.exame
    # exame_codigo : int(8)
    # paciente_codigo : int(20)
    # raiox_codigo : int(8)
    # exame_numero_registro : int(10)
    # exame_data : varchar(10)
    # exame_motivo : varchar(60)
    # exame_laudo : longtext
    # paciente_codigo_auxiliar : int(10)
    # exame_densa : varchar(20)
    
```

```

    v ladinew.patologia
    # codigo : int(3)
    # patologia : varchar(100)
    # patologia_eng : varchar(100)
    
```

```

    v ladinew.raiox
    # raiox_codigo : int(8)
    # hospital_codigo : int(10)
    # raiox_sala : varchar(30)
    # raiox_marca : varchar(30)
    # raiox_tpf_paralelo : int(10)
    # raiox_tpf_perpendicular : int(10)
    # raiox_ftm : longtext
    # raiox_winner : longtext
    
```

```

    v ladinew.scanner
    # scanner_codigo : int(8)
    # scanner_nome : varchar(30)
    # scanner_ftm : longtext
    # scanner_winner : longtext
    
```

```

    v ladinew.birads
    # categoria : float
    # descricao : longtext
    
```

```

    v ladinew.extensao
    # codigo : int(7)
    # imagem_nome : varchar(100)
    
```

Figura 4: Modelo relacional da base

#### 5.4 Interface e uso do BancoWeb

A base considerada neste trabalho foi denominada **BASE DE IMAGENS MAMOGRÁFICAS LAPIMO EESC/USP**, ou simplesmente, **BancoWeb**. Daqui em diante ela será referida apenas como BancoWeb por praticidade. Foi construída utilizando apenas ferramentas *open-source*, incluindo linguagem PHP 5.1 em um servidor APACHE rodando em Linux Debian. HTML e Javascript para apresentação ao usuário, e MySQL 4.0 para sua estruturação interna. Todas as imagens se encontram no servidor e existem 2 *backups* da base completa, incluindo as imagens inseridas.

BancoWeb foi construída de modo a obedecer todos os parâmetros descritos na seção anterior, a começar pelo seu acesso. A base está disponível *online* no *site* [HTTP://lapimo.sel.eesc.usp.br/bancoweb](http://lapimo.sel.eesc.usp.br/bancoweb). Ao acessar o *link*, o visitante será apresentado à página inicial da base (Figura 5) onde poderá selecionar o idioma em que deseja utilizar a base (português ou inglês), contatar o administrador, inserir seu usuário e senha para entrar no sistema ou se cadastrar para uso da mesma.



Figura 5: Pagina inicial

### 5.4.1 Procedimento de cadastro

O cadastro que permite o uso da base é gratuito, pedindo apenas um pequeno conjunto de informações, mostradas na Figura 6, incluindo nome completo, nível de escolaridade, idade, motivo e justificativa para desejar acesso. Estas informações têm como intuito facilitar a compreensão do público interessado na base, com fins de futura pesquisa estatística.

Uma vez cadastrado, o usuário deve aguardar a autorização oficial de um administrador, quando receberá por *e-mail* uma senha aleatória e o nível de acesso “usuário”. Estes níveis de acesso são divididos do seguinte modo:

- “**usuário**”<sup>3</sup>: é o nível inicial para todos os usuários, permitindo acesso para visualização e cópia de todas as imagens, assim como seleção de regiões de interesse para *download*. O menu principal para este nível é mostrado na Figura 7.

Cadastro do Usuário

**Informações Pessoais**

Nome Completo:

Data de Nascimento: Dia  Mês  (ex:2004)

Escolaridade:  escolha..

Profissão:

Interesse:

Justificativa:

**Cadastro do Usuário**

E-mail:

Login:

**Verificação de dados**

Digite o texto que está ao lado:  08VDOK

Informamos que este cadastro será autorizado pelo Administrador.  
Aguarde a liberação do cadastro em seu e-mail.

**Figura 6: Cadastro do usuário**

- “**pesquisador**”<sup>4</sup>: é o segundo nível de acesso, liberando, além dos acessos acima, a possibilidade de inserir novas imagens e exames e demarcar novas regiões de

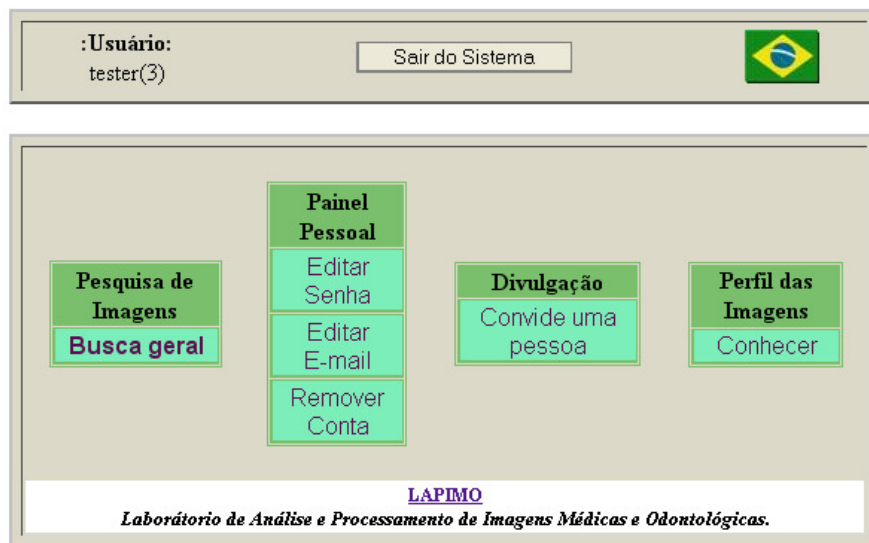
<sup>3</sup> Para evitar confusão, quando estiver se referindo ao nível de acesso o termo “usuário” estará entre aspas. Se o termo aparecer sem aspas se refere aos utilizadores da base como um todo.

interesse. Este nível é para pesquisadores convidados que tenham interesse em contribuir para o crescimento da base. O menu principal é mostrado na Figura 8.

- **“administrador”**: é o mais alto nível possível; permite, além dos acessos acima, eliminação de exames e imagens e possibilidade de editar e autorizar qualquer usuário. Este nível é restrito para os membros desenvolvedores da base. O menu principal é mostrado na Figura 9.

As imagens referentes aos menus dos usuários (Figura 97, 8 e 9) mostradas a seguir incluem todas as informações mostradas na tela para o usuário, incluindo a observação sobre a base.

É importante ressaltar que quando o "administrador" elimina algum exame ou paciente todas as informações relacionadas são eliminadas em conjunto, garantindo a integridade da base ao evitar informações incompletas ou mal referenciadas.



OBS: As imagens contidas nessa base não apresentam qualquer identificação pessoal sobre o paciente. Ao fazer o seu cadastramento, o usuário concorda que a utilização que porventura fizer das imagens dessa base não terá fins comerciais, tendo sua aplicação restrita à pesquisa científica. O usuário deverá citar em qualquer publicação ou divulgação de pesquisa decorrente do uso dessas imagens, além dos dados vinculados a elas, a Base de Imagens Mamográficas do LAPIMO, Registro INPI Nro. 09939-3.

**Figura 7: Menus para nível de acesso "usuário"**

**(OBS.: o texto abaixo da janela aparece sempre que o usuário acessa o sistema)**

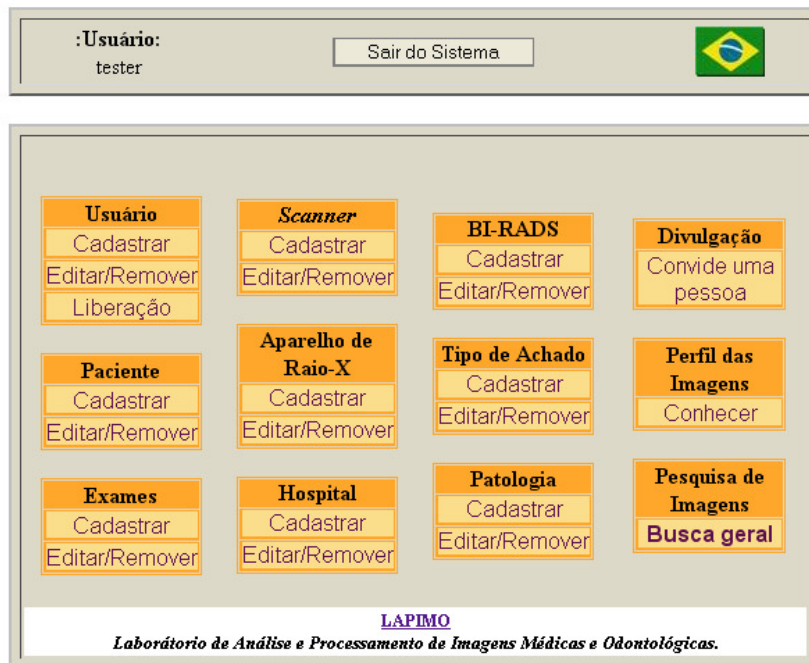
<sup>4</sup> Similar a anterior o termo “pesquisador” estará entre aspas quando se referir ao nível de acesso, sendo deixado sem aspas quando referir a desenvolvedores da área





OBS: As imagens contidas nessa base não apresentam qualquer identificação pessoal sobre o paciente. Ao fazer o seu cadastramento, o usuário concorda que a utilização que porventura fizer das imagens dessa base não terá fins comerciais, tendo sua aplicação restrita à pesquisa científica. O usuário deverá citar em qualquer publicação ou divulgação de pesquisa decorrente do uso dessas imagens, além dos dados vinculados a elas, a Base de Imagens Mamográficas do LAPIMO, Registro INPI Nro. 09939-3.

Figura 8: Menus para nível de acesso "pesquisador"



OBS: As imagens contidas nessa base não apresentam qualquer identificação pessoal sobre o paciente. Ao fazer o seu cadastramento, o usuário concorda que a utilização que porventura fizer das imagens dessa base não terá fins comerciais, tendo sua aplicação restrita à pesquisa científica. O usuário deverá citar em qualquer publicação ou divulgação de pesquisa decorrente do uso dessas imagens, além dos dados vinculados a elas, a Base de Imagens Mamográficas do LAPIMO, Registro INPI Nro. 09939-3.

Figura 9: Menus para nível de acesso "administrador"

Para um usuário ter seu nível de acesso ampliado, ele precisa fazer um requerimento a um dos administradores explicando seus motivos. Este requerimento será cuidadosamente avaliado e o nível de acesso será concedido se julgado pertinente. O objetivo deste processo é garantir que apenas poucos usuários previamente avaliados tenham direito a inserir imagens na base, garantindo sua confiabilidade.

#### 5.4.2 Inserção de imagens

O processo de inserção de imagens foi construído de modo a ser simples e de fácil acesso, enquanto garante a integridade das informações inseridas. Para isso, a inserção é realizada na seguinte seqüência pré-estabelecida:

- 1) Primeiramente as informações do hospital ou clínica são inseridas, incluindo nome completo, endereço, telefone, nome do contato no hospital ou clínica e nome do radiologista responsável (Figura 10).

A imagem mostra uma interface web para o cadastro de um hospital. O formulário é intitulado 'Cadastro do Hospital' e contém um sub-título 'Informações sobre Hospital'. Os campos de entrada são:

- Nome: [campo de texto]
- Endereço: [campo de texto]
- Complemento: [campo de texto]
- Bairro: [campo de texto]
- CEP: [campo de texto] ex: (14620000)
- Telefone: [campo de texto] - [campo de texto]
- Fax: [campo de texto] - [campo de texto]
- Contato: [campo de texto]
- Radiologista Responsável: [campo de texto]

Na base do formulário, há três botões: 'Voltar', 'Limpar' e 'Inserir'.

**Figura 10: Inserção de dados do Hospital**

- 2) O passo seguinte é a inserção das informações do equipamento de raios X e dos dados do paciente. Ambos são diretamente associados ao hospital de origem:
  - a. No caso dos aparelhos de raios X, as informações inseridas incluem o hospital de origem, FTM e espectro de Wiener do aparelho

(levantados em testes de controle de qualidade realizados pela equipe do LAPIMO no local), sala do aparelho no hospital, marca do aparelho e tamanho do ponto focal – também medido no teste de controle de qualidade (Figura 11). Caso o aparelho de raios X já tenha sido anteriormente cadastrado, este passo pode ser pulado.

A imagem mostra uma interface web para o cadastro de um aparelho de raios X. O formulário é dividido em seções. No topo, há um cabeçalho amarelo com o texto 'Cadastro do Aparelho de Raio-X'. Abaixo dele, um sub-cabeçalho azul contém o título 'Informações sobre o Aparelho Raio-X'. O formulário contém os seguintes campos e elementos:

- Um menu suspenso 'Origem:' com a opção 'Selecione...'. Abaixo dele, um aviso em um fundo laranja: 'Entre com os campos abaixo manualmente ou se preferir clique em anexar!'.
- Dois campos de texto grandes para 'FTM:' e 'Winner:', cada um com um botão 'Anexar' ao lado.
- Campos de texto menores para 'Sala:' e 'Marca:'.
- Um campo 'Tamanho do ponto focal:' com um menu suspenso que mostra 'X' selecionado e opções '[Paralelo] X' e '[Perpendicular]'.
- Na base do formulário, há três botões: 'Voltar', 'Limpar' e 'Inserir'.

**Figura 11: Inserção de dados aparelho de Raios X**

- b.** As informações de paciente são registradas utilizando o próprio hospital de origem mais o código do paciente dado pelo hospital como referência, garantindo que este número seja a única informação confidencial armazenada na base. Importante notar que este número não será referenciado em nenhuma parte da base para nenhum usuário, quando necessário será utilizado um código auxiliar interno a base para referenciar o paciente. As informações de paciente incluem hospital ou clínica de origem, data de nascimento, idade da telarca, tempo de lactação, existência de câncer na família, uso de anticoncepcional, reposição hormonal e existência de nódulo palpável (Figura 12). Caso o paciente já tenha sido cadastrado anteriormente este passo pode ser pulado.

**Cadastro do Paciente**

**Informações do Paciente**

Código do Paciente:

Origem do Paciente:

Data de Nascimento: Dia  Mês    
(ex:2004)

Sexo:

Idade Telarca:  Anos  Não Consta

Tempo Lactação:  Anos  Não Consta

Câncer na Família?  Sim  Não  Não Consta

Uso anticoncepcional?  Sim  Não  Não Consta

Reposição Hormonal?  Sim  Não  Não Consta

Nódulo Palpável?  Sim  Não  Não Consta

**Figura 12: Inserção de dados do paciente**

- 3) A inserção seguinte são os dados do *scanner*, incluindo a marca, FTM e espectro de Wiener (Figura 13). Caso o *scanner* já tenha sido cadastrado anteriormente este passo pode ser pulado.

**Cadastro do Scanner**

**Informações sobre o Scanner**

Nome:

Entre com os campos abaixo manualmente ou se preferir clique em anexar!

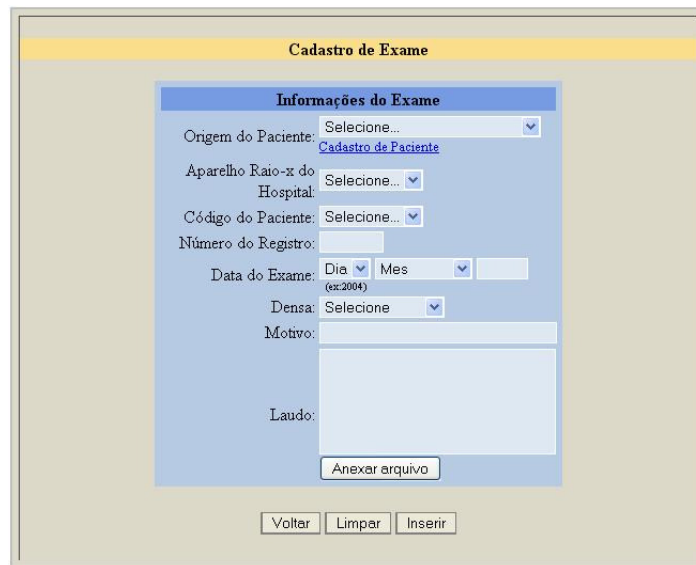
FTM:

Wiener:

**Figura 13: Inserção de dados do Scanner**

- 4) Os dados do exame incluem a seleção do hospital de origem e código do paciente e aparelho de raios X utilizado. Em seguida os dados do exame em

si são inseridos, incluindo número de registro do exame, data do exame, densidade da mama (densa ou lipossobstituída), motivo da realização do exame e o laudo radiológico (Figura 14). Assim como com o número do paciente o número de registro do exame não será referenciado dentro da base, quando necessário será substituído por um código auxiliar interno a base.



A imagem mostra uma interface web para o cadastro de exames. O formulário principal é intitulado 'Cadastro de Exame' e contém uma seção 'Informações do Exame'. Os campos incluem: 'Origem do Paciente' (menu suspenso com opção 'Selecione...' e link 'Cadastro de Paciente'), 'Aparelho Raio-x do Hospital' (menu suspenso), 'Código do Paciente' (menu suspenso), 'Número do Registro' (campo de texto), 'Data do Exame' (seletores de dia e mês com exemplo '(ex:2004)'), 'Densa' (menu suspenso), 'Motivo' (campo de texto) e 'Laudo' (área de texto grande). Há um botão 'Anexar arquivo' na base do formulário. Na parte inferior da janela, há botões 'Voltar', 'Limpar' e 'Inserir'.

**Figura 14: Inserção de dados do exame**

- 5) Uma vez inseridos os dados do exame, devem-se inserir os dados das imagens propriamente. Primeiro, é necessário selecionar o *scanner* utilizado para a digitalização da imagem em filme; em seguida, deve-se informar a posição da mama no exame (médio-lateral, crânio-caudal ou se resultante de procedimento de magnificação radiográfica), classificação BI-RADS, arquivo da imagem, tamanho do pixel (resolução espacial), tipo de arquivo, quantidade de bits (resolução de contraste), mama (direita ou esquerda), existência ou não de achados e, finalmente, a seleção do arquivo da imagem para upload (Figura 15). Caso exista algum achado na imagem, registrado no laudo, o passo seguinte (5.a) será apresentado, senão o passo de inserção da imagem (5) é reapresentado até que todas as imagens de um dado exame sejam inseridas.

- a. Quando a imagem tem pelo menos um achado, as seguintes informações são incluídas: quadrante de localização do achado

(superior externo, superior interno, inferior externo, inferior interno, aureolar ou axilar), coordenadas (em pixels, do canto superior esquerdo e inferior direito da região do achado), tipo de achado e patologia (Figura 16). Tanto os tipos de achados quanto as patologias são tabelados de modo a garantir integridade dos termos e facilitando o sistema de busca, Caso seja uma patologia ou achado ainda não catalogado no sistema, o “pesquisador” ou “administrador” pode adicioná-lo. Caso exista mais de um achado, o passo 5.a) é repetido enquanto necessário.

A imagem mostra uma interface web para o cadastro de imagens. O formulário é dividido em três seções principais:

- Scanner e Posição:** Possui dois menus suspenso para selecionar o scanner e a posição da imagem.
- BI-RADS:** Possui um menu suspenso para selecionar a categoria BI-RADS.
- Dados da Imagem:** Possui campos para: número do registro (preenchido com 999999), nome do arquivo (com um botão 'Enviar arquivo...'), tamanho de pixel, tipo, quantidade de bits e mama (todos com menus suspensos). Também possui botões de rádio para 'Imagem possui achado?' com opções 'Sim' e 'Não' (selecionada).

Na base do formulário, há três botões: 'Voltar', 'Limpar' e 'Inserir'.

**Figura 15: Inserção de dados da imagem**

As informações inseridas dependem apenas da disponibilidade das mesmas nos laudos e informações de pacientes oferecidas pelos hospitais e clínicas colaboradores com o projeto. Caso as informações não estejam disponíveis, as opções apresentadas durante a inserção incluem uma opção de “não disponível” ou equivalente.

No caso de se necessitar inserir arquivos originários do padrão DICOM (NATIONAL ELECTRICAL MANUFACTURERS ASSOCIATION, 2009), a informação deverá ser manualmente particionada e adicionada à base utilizando o

processo de inserção descrito acima. Entre tarefas de aprimoramento do sistema, previstas para um desenvolvimento futuro, está a execução de ferramenta para possibilitar inserir e recuperar conjuntos de imagens e informações diretamente no padrão DICOM.

A imagem mostra uma janela de software intitulada "Cadastro do Achado". Dentro dela, há um formulário com o título "Informações do Achado". O formulário contém os seguintes campos: "Localização:" com uma lista suspensa "Selecione..."; "Coordenadas:" com campos para "XSE:", "YSE:", "XID:" e "YID:"; "Categoria BI-RADS:" com uma lista suspensa "Selecione..."; "Tipo de Achado:" com uma lista suspensa "Selecione..."; e "Patologias detectadas:" com uma lista suspensa "Selecione...". Abaixo do formulário, há uma caixa de seleção desativada com o texto "Quero cadastrar mais um achado na imagem atual". Na base do formulário, há três botões: "Voltar", "Limpar" e "Inserir".

**Figura 16: Inserção dos achados da imagem**

### 5.4.3 Busca e recuperação de imagens

A recuperação das imagens é feita utilizando um de quatro módulos, cada um associado a um conjunto de informações inserido com o exame (Figura 17). Estes módulos são:

- 1) **Informações da imagem:** permite buscas sobre a posição da imagem (crânio-caudal, médio-lateral ou de magnificação), tamanho do pixel, tipo do arquivo, categoria BI-RADS, resolução de contraste, densidade da mama, existência, tipo e localização do achado e patologia (Figura 17a).
- 2) **Informações de paciente:** permite buscas sobre a sua origem, idade da telarca, tempo de lactação, existência de câncer na família, uso de anticoncepcional, reposição hormonal e nódulo palpável (Figura 17b).
- 3) **Raios X:** permite buscas em relação ao equipamento de raios X utilizado (Figura 17c).
- 4) **Scanner:** permite buscas pelo equipamento de digitalização utilizado (Figura 17d).

**Busca**

**(a) Busca por imagens**

Posição: Todas

Tamanho de Pixel: Todos

Tipo: Todos

Categoria BI-RADS: Todos

Resolução de contraste (em bits): Todos

Densa: Qualquer

Mama: Qualquer

Achado: Qualquer

**Informações do Achado**

Localização: Qualquer

Tipo de Achado: Qualquer

Patologias Detectadas: Qualquer

**(b) Busca por pacientes**

Origem do Paciente: Todos os Hospitais

Sexo: Qualquer

Idade Telarca: Entre e Anos

Tempo Lactação: Entre e Anos

Câncer na Família?  Sim  Não

Uso anticoncepcional?  Sim  Não

Reposição Hormonal?  Sim  Não

Nódulo Palpável?  Sim  Não

**(c) Busca por aparelhos de Raio-X**

Escolha	Marca	Origem	FTM	WIENER	Paralelo (micrometro)	Perpend. (micrometro)
<input checked="" type="checkbox"/>	Senographe 500t	Hospital São Paulo	.	.	300	300
<input checked="" type="checkbox"/>	Senographe 600t	HC de Botucatu	1	1	300	300
<input checked="" type="checkbox"/>	A especificar	Hospital das Clinicas de Ribeir	.	.	300	300

**(d) Busca por aparelhos de Scanner**

Escolha	Nome	FTM	ESPECTRO DE WIENER
<input checked="" type="checkbox"/>	Lumiscan 50 (Lumisys)	.	.
<input checked="" type="checkbox"/>	Lumiscan 75 (Lumisys)	.	.

MENU PESQUISAR

**Figura 17: Painel de busca por (a) dados da imagem; (b) dados do paciente; (c) Aparelho de raios X; (d) Scanner**

Os módulos de busca são completamente integrados, permitindo busca por qualquer conjunto de informações. A divisão em módulos existe apenas para facilitar o uso da ferramenta, tornando o sistema de busca mais simples para uso.

Os resultados da busca são apresentados em blocos de quatro imagens, cada uma com um *thumbnail* e informações sobre a imagem e o paciente (Figura 18). As informações apresentadas em cada bloco (Figura 19) incluem detalhes do exame, da imagem e dados do paciente. No canto superior esquerdo, o usuário pode passar para a página seguinte ou anterior, e no direito encontram-se as opções para baixar todas as imagens resultantes da busca (até o limite de 100 imagens) ou retornar para uma nova busca ou para o menu inicial.

#### 5.4.4 Opções e ferramentas

Para cada imagem são oferecidas ao usuário quatro opções ao lado do *thumbnail*, que são:



- 1) **Ampliar:** permite ao usuário ver a imagem no tamanho original comprimida com perdas em formato JPEG. Esta imagem não é recomendada para uso em processamento, tendo como intuito permitir a visualização do exame sem baixar a versão completa, reduzindo o tamanho do arquivo de 10MB para apenas 100KB.

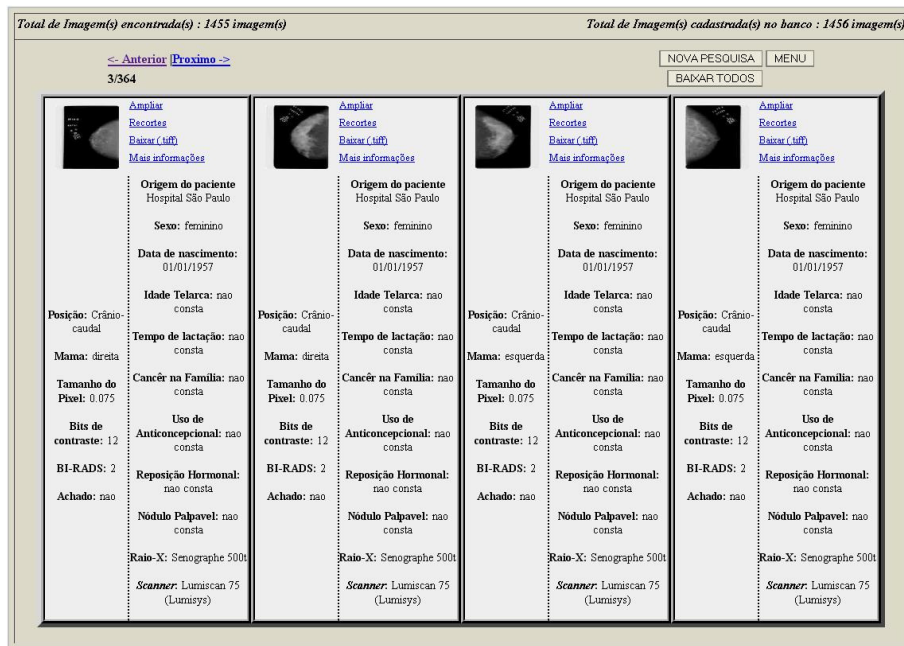


Figura 18: Exemplo de resultados da busca

- 1) **Recortes:** permite visualizar as regiões de interesse já selecionadas para a dada imagem, os motivos dados para a seleção e baixar as regiões em formato sem perdas (Figura 20). Além disso, permite acessar a ferramenta de recorte de regiões de interesse. Esta ferramenta será melhor discutida abaixo.
- 2) **Baixar:** permite baixar a imagem na versão original, completa e sem perdas. Em geral esta imagem será em TIFF, mas a base está estruturada para aceitar outros tipos de arquivo também.
- 3) **Mais informações:** permite acessar o conjunto de informações mais detalhadas sobre paciente e exame, incluindo o laudo completo e as outras imagens do mesmo exame (Figura 21).

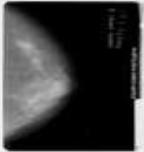
	<a href="#">Ampliar</a> <a href="#">Recortes</a> <a href="#">Baixar (.tiff)</a> <a href="#">Mais informações</a>
<b>Posição:</b> Crânio-caudal	<b>Origem do paciente</b> Hospital São Paulo
<b>Mama:</b> esquerda	<b>Sexo:</b> feminino
<b>Tamanho do Pixel:</b> 0.075	<b>Data de nascimento:</b> 01/01/1948
<b>Bits de contraste:</b> 12	<b>Idade Telarca:</b> nao consta
<b>BI-RADS:</b> 4	<b>Tempo de lactação:</b> nao consta
<b>Achado:</b> sim	<b>Cancêr na Família:</b> nao consta
<b>Localização</b>	<b>Uso de</b>
I) Não Informado II) Não Informado III) Não Informado	<b>Anticoncepcional:</b> nao consta
<b>Tipo</b>	<b>Reposição Hormonal:</b> nao consta
I) Fibrose Estromal II) Atrofia Lobular III) Focos de calcificações distrofica em lumen de ductos termina	<b>Nódulo Palpavel:</b> nao consta
<b>Patologia</b>	<b>Raio-X:</b> Senographe 500t
I) Mastite crônica inespecífica focal II) Mastite crônica inespecífica focal III) Mastite crônica inespecífica focal	<b>Scanner:</b> Lumiscan 75 (Lumisys)

Figura 19: Exemplo de bloco de resultado



Figura 20: Exemplo da opção recorte (região demarcada apenas para ilustração)

Total de Imagem(s) encontrada(s) : 22 imagem(s)

Dados do Paciente	
Hospital:	Hospital São Paulo
Data de Nascimento:	01/01/1956
Sexo:	Feminino
Telarca:	nao consta
Lactação:	nao consta
Cancêr na Família:	nao consta
Uso de Anti-Concepcional:	nao consta
Reposição Hormonal:	nao consta
Nódulo Palpavel:	nao consta

Dados do Exame I	
Data:	30/10/2000
Motivo:	
Laudo:	Conclusão: quadro radiológico compatível com substituição gordurosa. Sugere-se mamografia de rotina. Exame anterior não disponível para comparação. categoria: BI-RADS I
Equipamento de Raios-X:	Senographe 500t

Dados do Exame II	
Data:	31/09/2000
Motivo:	
Laudo:	Quadro radiológico compatível com substituição gordurosa. Exame anterior não disponível para comparação. categoria: BI-RADS I
Equipamento de Raios-X:	Senographe 500t

< Anterior | Próximo >

3/6 BAIXAR TODOS

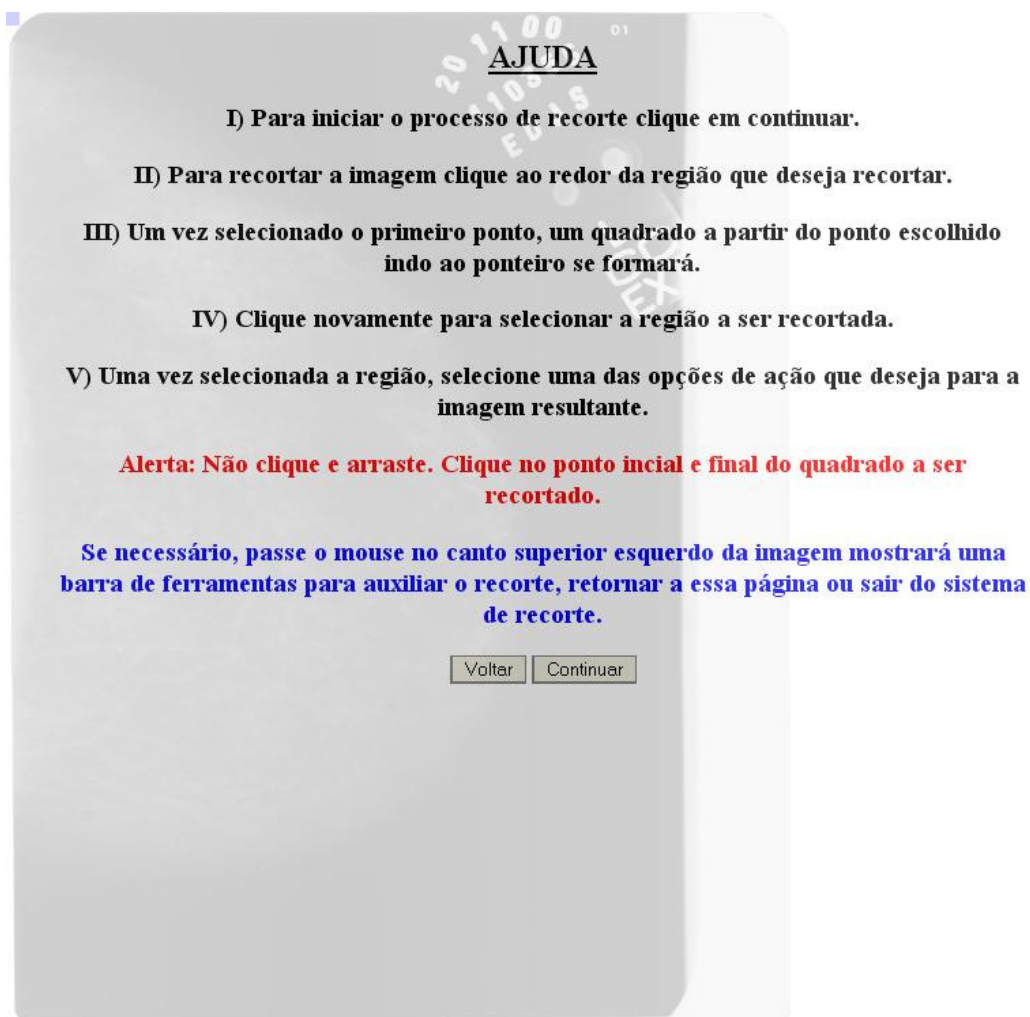
Exame I	Exame I	Exame I	Exame I
<a href="#">Ampliar</a> <a href="#">Recortes</a> <a href="#">Baixar (.tif)</a>	<a href="#">Ampliar</a> <a href="#">Recortes</a> <a href="#">Baixar (.tif)</a>	<a href="#">Ampliar</a> <a href="#">Recortes</a> <a href="#">Baixar (.tif)</a>	<a href="#">Ampliar</a> <a href="#">Recortes</a> <a href="#">Baixar (.tif)</a>
Achado: sim	Achado: sim	Achado: sim	Achado: sim
Localização: D) Não Informado II) Não Informado	Localização: D) Não Informado	Localização: D) Não Informado	Localização: D) Não Informado
Mama: direita	Mama: direita	Mama: direita	Mama: direita
Tamanho do Pixel: 0.075	Tamanho do Pixel: 0.075	Tamanho do Pixel: 0.075	Tamanho do Pixel: 0.075
Bits de contraste: 12	Bits de contraste: 12	Bits de contraste: 12	Bits de contraste: 12
BI-RADS: 4	BI-RADS: 4	BI-RADS: 1	BI-RADS: 2
Scanner: Lumiscan 75 (Lumisys)	Scanner: Lumiscan 75 (Lumisys)	Scanner: Lumiscan 75 (Lumisys)	Scanner: Lumiscan 75 (Lumisys)

Figura 21: Opção "mais informações"

A ferramenta de **Recortes**, uma vez selecionada, leva o usuário a uma nova página com instruções de como proceder para realizar o recorte de uma região de interesse da imagem (Figura 22).

Após pressionar “continuar”, o usuário, terá uma versão JPEG com perdas da imagem e poderá, nela, selecionar qualquer região quadrada (Figura 23) para recorte. Uma vez selecionada a região, o usuário poderá baixar o recorte em versão sem perdas ou refazer o recorte.

Caso o usuário tenha nível de acesso de “pesquisador” ou acima, também poderá registrar a região de interesse, explicando o motivo da seleção de tal área. Se a região for registrada, todos os usuários poderão visualizar na opção “recorte” do resultado de busca (item (2) acima) e baixar a imagem em versão sem perdas (este recorte será, então, armazenado no servidor).



**AJUDA**

I) Para iniciar o processo de recorte clique em continuar.

II) Para recortar a imagem clique ao redor da região que deseja recortar.

III) Um vez selecionado o primeiro ponto, um quadrado a partir do ponto escolhido indo ao ponteiro se formará.

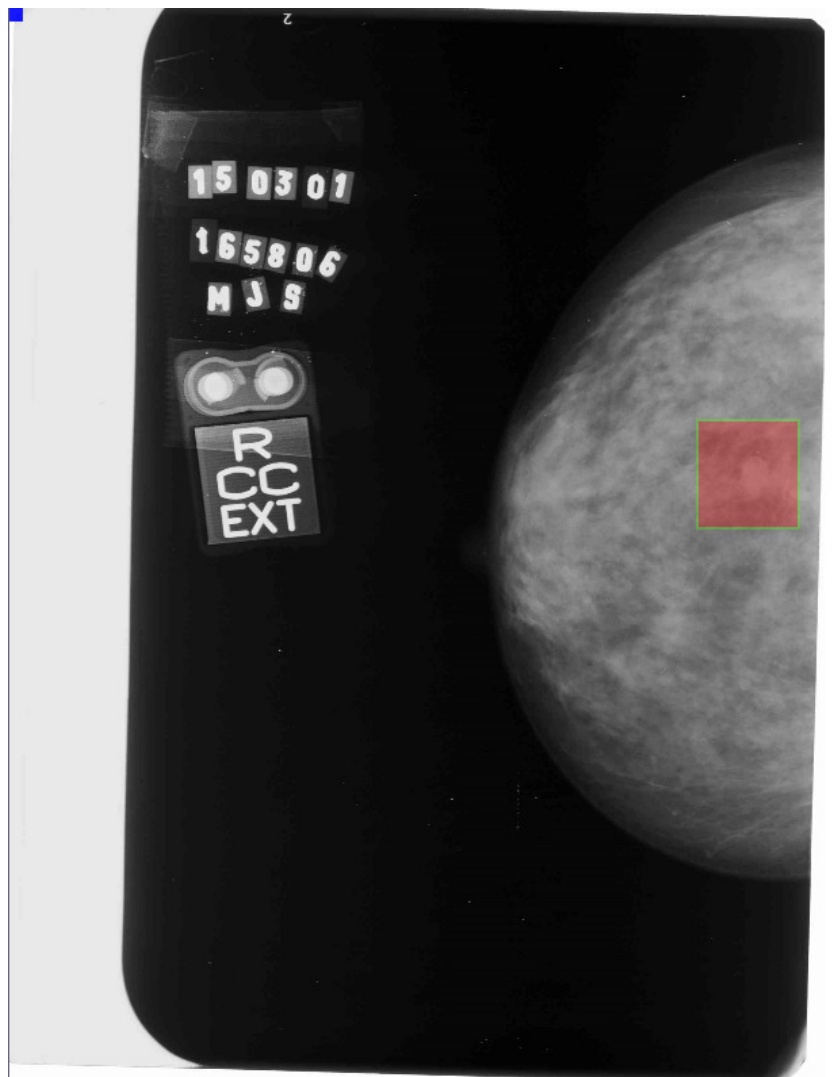
IV) Clique novamente para selecionar a região a ser recortada.

V) Uma vez selecionada a região, selecione uma das opções de ação que deseja para a imagem resultante.

**Alerta: Não clique e arraste. Clique no ponto inicial e final do quadrado a ser recortado.**

Se necessário, passe o mouse no canto superior esquerdo da imagem mostrará uma barra de ferramentas para auxiliar o recorte, retornar a essa página ou sair do sistema de recorte.

Figura 22: Instruções para uso do sistema de recorte



**Figura 23: Recorte sendo realizado**

Esta ferramenta foi desenvolvida utilizando PHP, Javascript e ImageMagik, de modo que nenhum *download* ou *plug-in* seja necessário para seu uso.

O BancoWeb também inclui uma ferramenta de perfil da base, ou seja, uma ferramenta que permite a visualização da distribuição de imagens na base de acordo com as suas características (Figura 24). Ao selecionar uma das distribuições, ao usuário é apresentado um gráfico de fatias mostrando as porcentagens das características escolhidas (Figura 25).

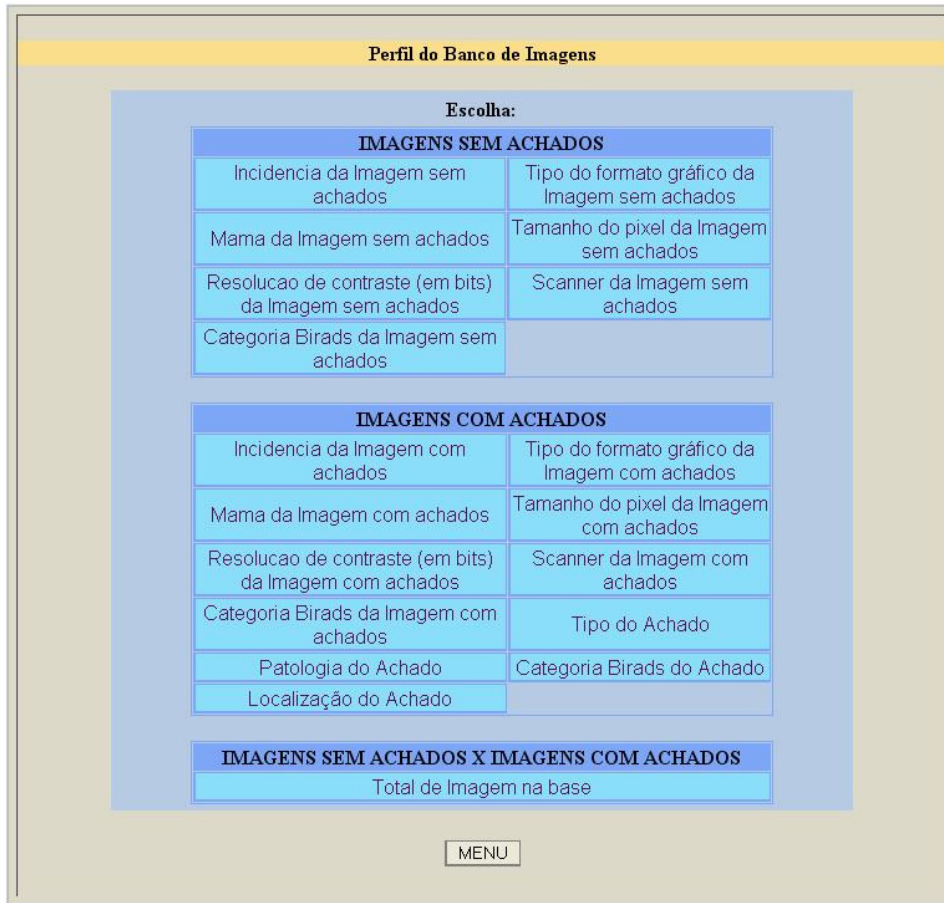


Figura 24: Ferramenta de perfil estatístico

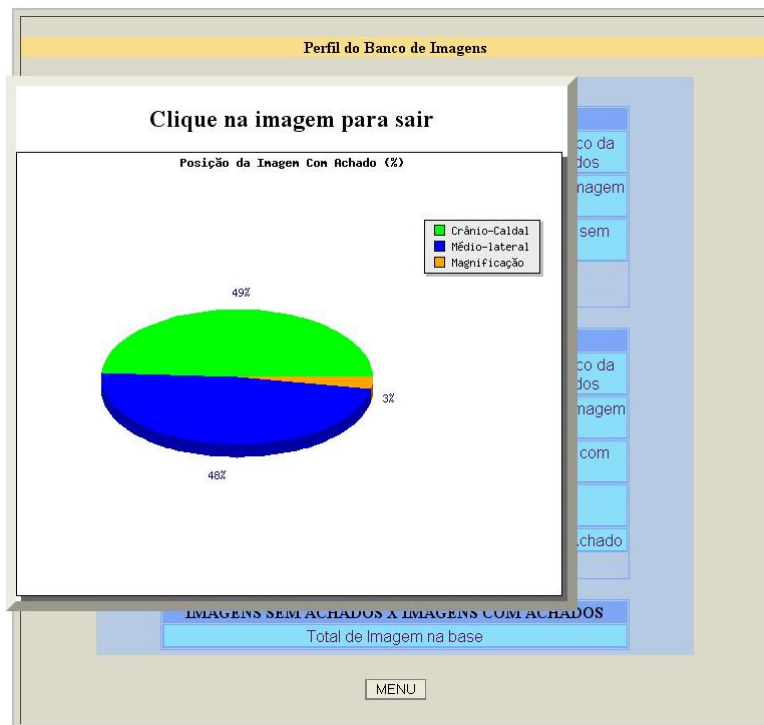


Figura 25: Exemplo do uso da ferramenta de perfil estatístico



## Capítulo 6: Resultados

O BancoWeb se encontra completamente funcional, com todas as ferramentas listadas anteriormente funcionais em todos os *browsers* testados (Internet Explorer 7 ou superior, Mozilla Firefox, Google Chrome e Safari), sem ser necessária a instalação de nenhum *software* ou *plug-in* extra, conforme planejado.

Para acessar a base basta utilizar o seguinte *link*: <http://lapimo.sel.eesc.usp.br/bancoweb>.

Para a avaliação, a base foi comparada a algumas já divulgadas e comumente descritas ou mencionadas em artigos e projetos da área, conforme discutido no capítulo 4. Informações resultantes dessa comparação são sumarizadas na Tabela 6.1.

### 6.1 Acesso, manutenção e atualização

Das bases comparadas, existem três públicas (MIAS, DDSM e BancoWeb), uma paga (UCSF) e uma restrita a pesquisadores e hospitais associados (CALMa). A base UCSF cobra U\$ 100 para "custos de reprodução", pois só é oferecida em forma física composta por 12 CDs. O BancoWeb é a única das bases públicas desse comparativo a exigir cadastro dos seus usuários.

Das bases públicas consideradas aqui apenas o BancoWeb continua sendo um projeto em aberto, que pretende manter-se em desenvolvimento pelo menos pelos próximos anos. Tanto o projeto da base MIAS como da DDSM estão fechados de acordo com seus respectivos *sites* e os desenvolvedores não mais realizarão atualizações e não se responsabilizam por problemas técnicos que a base possa apresentar.

A base UCSF não pretende atualizar seu sistema, mas por ser um produto comercial, tem oferecimento de assistência. Já o projeto CALMa, apesar de ter mudado de nome para MAGIC-5, parece continuar ativo.

### 6.2 Quantidade de imagens

Até a presente data, o BancoWeb contém 1.436 imagens (proveniente de 411 pacientes), mais do que as 1.000 recomendadas por Nishikawa (Nishikawa, 1998). Em comparação com as outras bases disponíveis a CALMa supera o BancoWeb atual, mas como a base é fechada nenhuma destas imagens está disponível para o público geral. A DDSM, no entanto, supera consideravelmente todas as outras, oferecendo 10.480 imagens livres para qualquer interessado.



Tabela 6.1: Resultados comparativos

	MIAS (UK)	DDSM (USA)	UCSF / LLNL (USA)	CALMa (IT)	BancoWeb (BR)
<b>Quantidade de imagens</b>	320	10.480	198	3.000	1.436
<b>Acesso aos arquivos</b>	Livre	Livre	Pago (US\$ 100)	Fechada	Livre após cadastro
<b>Acesso a inserção</b>	Apenas administradores	Apenas administradores	Apenas administradores	Apenas administradores	Profissionais selecionados
<b>Resolução de contraste (bits/pixel)</b>	8	12 (80%) 16 (20%)	12	12	12
<b>Resolução espacial (mm)</b>	0,050	0,042 (19%) 0,043.5 (44%) 0,050 (37%)	0,035	0,085	0,075 (32%) 0,150 (68%)
<b>Formato da imagem armazenada</b>	PGM	LJPEG <i>Não convencional</i>	*	*	TIFF
<b>Sistema de busca</b>	Não	Sim, mas não funcional	*	*	Sim
<b>Manutenção e atualização</b>	Não	Não	Sim	Sim	Sim
<b>Regiões de interesse</b>	Sim	Sim	Sim	*	Em demarcação

É importante ressaltar que das bases públicas o BancoWeb é o único que ainda está com o projeto ativo, ou seja, a quantidade de imagens apresentada neste trabalho consiste da quantidade inserida até a data de publicação do mesmo. Apenas considerando as imagens catalogadas na base física do LAPIMO contamos com mais de 7.400 imagens que estão gradativamente sendo inseridas no BancoWeb, garantindo crescimento da base.

### 6.3 Resolução espacial e de contraste

As imagens até o momento inseridas no BancoWeb foram digitalizadas em um de dois digitalizadores a LASER, Lumiscan 50 (68% – 976 imagens) com 3.500 níveis de cinza e 0,150mm de resolução espacial, e Lumiscan 75 (32% – 460 imagens) com 3.800 níveis de cinza e 0,075mm de resolução espacial. Nenhum destes valores implica uma limitação do sistema ou estrutura da base, que permite a inserção de imagens de qualquer resolução espacial ou de contraste; apenas, até o momento, nenhuma imagem foi inserida com características diferentes destas.

Das bases analisadas no comparativo, a única a apresentar imagens com menos de 12 bits de resolução de contraste é a MIAS, que é a mais antiga entre as avaliadas neste trabalho. Ainda assim, ela apresenta uma resolução espacial de 0,050mm, superior, por exemplo, à do BancoWeb.

A base DDSM é a única a apresentar imagens de 16 bits (65.536 níveis de cinza) apesar de ser uma porcentagem pequena das imagens da base (apenas 20%) significa uma quantidade de imagens maior do que a atual quantidade de imagens do BancoWeb. A resolução espacial também é a melhor das bases públicas analisadas, variando de 0,050mm até 0,042mm.

A base UCSF apresenta a melhor resolução de todas as bases analisadas, de 0,035mm, sendo esta a sua grande vantagem sobre as outras existentes no mercado.

### 6.4 Formato dos arquivos de imagem

Conforme discutido anteriormente, o tipo de arquivo preferencial do BancoWeb foi escolhido como o formato TIFF, mas a base está estruturada para receber qualquer formato, necessitando apenas da inserção nas tabelas do formato escolhido.

A base MIAS utiliza o formato PGM, que é restrito a 8 bits de resolução de contraste, por isso não sendo recomendado para imagens mamográficas atualmente – ao menos, do ponto de vista de análise computacional.

A DDSM utiliza um formato não-tradicional do LJPEG, que é um formato não suportado pela maioria dos programas e ferramentas comerciais. Para utilizar este formato, o desenvolvedor precisa empregar as bibliotecas em linguagem C oferecidas no *site* da base em uma plataforma LINUX. A dificuldade em utilizar as imagens no

formato da DDSM restringe o uso da base por desenvolvedores não treinados com as ferramentas oferecidas.

O formato das imagens nas bases UCSF e CALMa não é informado em seus artigos e/ou *sites*.

### **6.5 Sistemas de busca**

A única base que claramente apresenta um sistema de buscas amplo e funcional no momento é o BancoWeb, fornecendo opções variadas de busca e a possibilidade de escolher conjuntos diferenciados e categorizados de imagens para *download*. O sistema de busca é extremamente eficiente, para a quantidade de imagens atual leva menos de 1 segundo para realizar uma busca.

A base DDSM afirma apresentar um sistema de busca, mas que não funciona atualmente, sendo que sua última atualização remonta a 2003. Como tal projeto está fechado e os autores não mais respondem a questionamentos técnicos da base, pode-se assumir que o sistema de busca da DDSM não deve voltar a funcionar em futuro próximo. Considerando a quantidade de imagens disponibilizadas nesta base, a falta de um bom sistema de busca torna difícil a procura por características específicas.

A base MIAS, apesar de tabelar cada informação, não oferece um sistema de busca, mas, dada a pequena quantidade de imagens, é fácil procurar diretamente a característica desejada. O mesmo vale para a base UCSF cujo tamanho também permite uma fácil busca direta.

### **6.6 Regiões de interesse**

A MIAS apresenta as anormalidades porventura existentes nas imagens na forma de coordenadas em sua tabela verdade, informando o ponto central da anormalidade e o seu raio aproximado.

No caso da base UCSF, cada imagem com anormalidades é associado a um mapa binário das regiões de interesse, ou seja, uma imagem binária de mesma altura e largura da imagem original onde as regiões de interesse são “1”s e as regiões consideradas normais são “0”s.

A DDSM oferece suas regiões de interesse demarcadas como um arquivo contendo um (ou mais, dependendo da quantidade de regiões numa mesma imagem) vetor(es) de coordenadas, formando uma borda enumerada ao redor da anormalidade.

De acordo com o *site* da base, estas coordenadas foram obtidas por conversão da demarcação realizada por um radiologista experiente. A demarcação das regiões em todas as imagens disponíveis é um dos pontos mais fortes da base DDSM, facilitando muito o trabalho para o desenvolvedor sem amplo treinamento em detecção de anormalidades, além de facilitar a automatização dos testes e avaliações dos esquemas CAD.

O BancoWeb, no entanto, apesar de fornecer 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. A maioria das imagens, porém, apresenta o quadrante em que a anormalidade foi encontrada, conforme indicado no laudo médico.

## **6.7 Outras ferramentas**

O BancoWeb também apresenta outras ferramentas únicas e de grande utilidade para o desenvolvedor da área conforme descrito a seguir.

### **6.7.1 Inserção remota**

Como foi descrito no capítulo 5, o BancoWeb contém um sistema de inserção de imagens na base estruturado para que qualquer usuário com a correta autorização (“pesquisador” ou “administrador”) possa inserir imagens de modo simples e sem se preocupar em conhecer a estrutura interna da base ou o método de arquivamento.

Esta ferramenta foi implementada com o intuito de dar maior longevidade e amplitude à base, permitindo que contatos sejam feitos com diferentes hospitais, clínicas e centros de pesquisa a nível mundial e autorizar usuários confiáveis a ampliar a base. Com isto, mesmo que o projeto seja considerado fechado, outros “administradores” e “pesquisadores” podem dar continuidade ao crescimento e variedade da base.

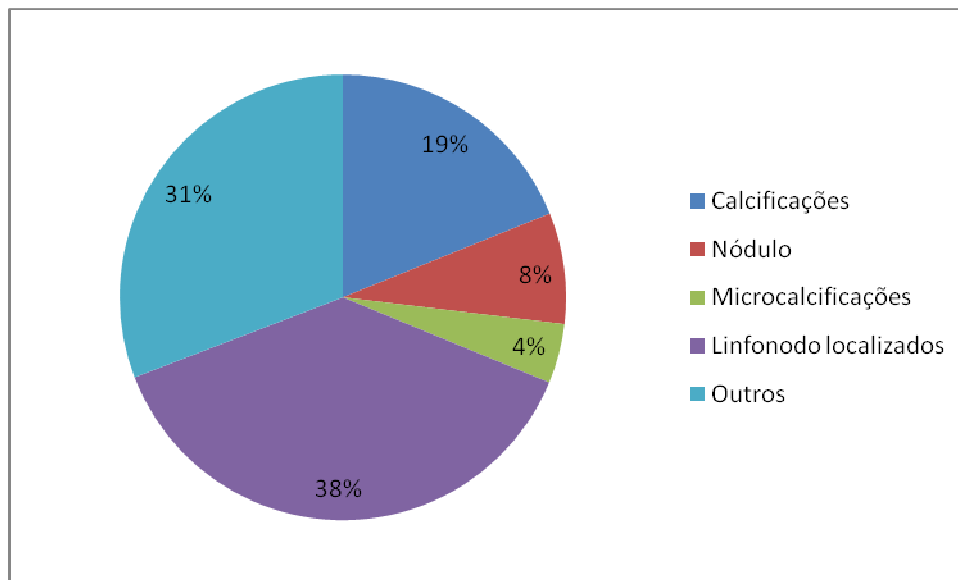
Apesar da estrutura para inserção remota já estar completamente pronta, no momento só contamos com membros integrantes locais ao LAPIMO na categoria “pesquisador”.

### **6.7.2 Perfil estatístico**

Para facilitar a avaliação da distribuição estatística de imagens na base e permitir ao usuário uma melhor compreensão das imagens como um todo, foi criada uma ferramenta que permite a qualquer usuário observar em gráficos algumas das mais importantes informações dentro da base.

Estas informações são obtidas em tempo real na base a cada pedido realizado, garantindo que, independentemente de quantas imagens novas sejam inseridas, os gráficos estejam sempre atualizados.

As imagens da base estão distribuídas em 30% com achado e 70% sem, uma distribuição próxima da realidade clínica. As imagens sem achados foram acrescentadas à base para permitir a avaliação de resultados falsos-positivos do esquema CAD. Os achados estão distribuídos como na Figura 26, lembrando que a mesma imagem pode conter mais de um achado.

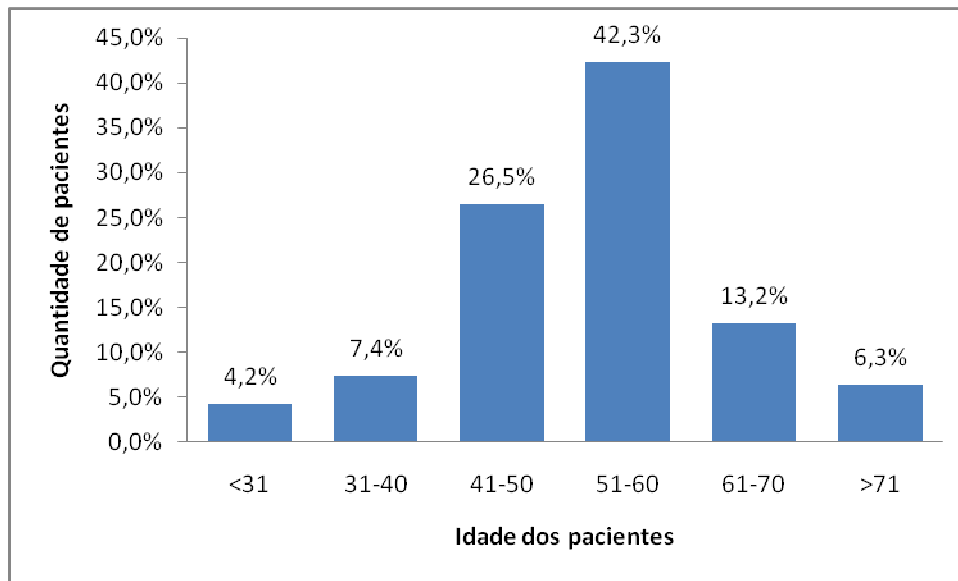


**Figura 26: Distribuição estatística dos achados**

Comparando a idade das pacientes no momento do exame, temos a distribuição mostrada na Figura 27. O gráfico mostra claramente que a maioria das pacientes (aproximadamente 68%) tinham idades entre 40 e 60 anos no momento do exame. Isto se deve à recomendação de que mulheres a partir dos 40 anos de idade realizem exames mamográficos a cada 2 anos, devido ao aumento gradual do risco de câncer de mama.

Além disso, estatísticas clínicas indicam que pacientes com mais de 60 anos, que não apresentam histórico de doenças mamárias, têm uma probabilidade muito baixa de

desenvolver qualquer patologia. Por conta disto, a quantidade de pacientes que realizam exame mamográfico após os 60 anos cai drasticamente, de acordo com o INCA.



**Figura 27: Distribuição dos pacientes de acordo com a idade no momento do exame**



## Capítulo 7: Discussão e Conclusões

Para testar de modo coerente um esquema CAD é necessário avaliar seus resultados perante uma grande quantidade e variedade de imagens, nem sempre disponíveis para o desenvolvedor. A dificuldade de acesso aos arquivos, em geral confidenciais, de hospitais e clínicas e a necessidade de equipamentos apropriados para digitalização das imagens mostra a necessidade de formar uma base pública ampla e de fácil acesso.

Em anos de pesquisa na área, o LAPIMO acumulou uma grande quantidade de imagens de vários hospitais, digitalizadas em diferentes equipamentos. Esta biblioteca local de imagens foi associada a um *software* de busca e recuperação para permitir o fácil uso da ampla quantidade de imagens disponíveis, formando a primeira versão da base de imagens do LAPIMO.

A linguagem DELPHI foi utilizada para o desenvolvimento do sistema de interação com o usuário na base local original (BENATTI, 2003), enquanto o padrão SQL (utilizando uma biblioteca para acesso direto pelo programa em DELPHI) foi empregado para a estruturação e administração da base. Ao ser completada em 2003, esta base era composta por 20 CDs contendo aproximadamente 3.300 imagens e um programa agregado para sua busca e visualização (este programa tinha aproximadamente 10Mb e podia ser facilmente gravado junto aos CDs de imagens).

O programa continha um sistema de buscas que permitia resgatar imagens pelo hospital de origem, idade da paciente (no momento do exame) e estrutura de interesse. Quando uma imagem era selecionada, ao usuário eram apresentadas as informações pertinentes ao caso, como laudo médico e equipamentos utilizados para obtenção e digitalização; também uma versão com resolução de contraste de 8bits da imagem era apresentada com vistas a uma análise visual (já que monitores convencionais não são capazes de lidar com mais do que 256 níveis de cinza). Por ser um programa executável construído em DELPHI, ele funcionava apenas em plataformas Windows.

Uma vez que esta base local estava funcionando, foi planejada a expansão dela para uma base *online* que permitisse a qualquer interessado o acesso às imagens. Pesquisando a literatura, foram encontradas algumas bases, mas não foi encontrada nenhuma que cumprisse com as exigências para o desenvolvimento de CAD e



oferecesse ferramentas úteis para auxiliar o desenvolvedor, especialmente ferramentas de busca.

A nova versão foi desenvolvida em PHP e HTML para interação remota com usuário, e MySQL como dialeto para a base de dados propriamente. O programa e as imagens são armazenados em um servidor Apache em uma plataforma LINUX, podendo ser acessado através de qualquer plataforma e *browser* de *internet*, sem requerer nenhum *plug-in*. Esta nova estrutura mantém todas as ferramentas e funcionalidade anteriores, além de outras desenvolvidas posteriormente. A esta base foi dado o nome de BancoWeb.

Infelizmente, testes com usuários interessados (na sua maioria, pesquisadores da área) mostrou que a base necessitava de reformulação em seu modelo, tanto para adicionar novas informações pertinentes quanto para permitir buscas mais detalhadas na base. Realizar tais alterações exigiriam a construção de uma base praticamente nova. Por isso a base continuou ativa, mas em paralelo foi construída uma nova versão, apresentada neste trabalho, que acabou por substituir o BancoWeb original.

O trabalho desenvolvido aqui também levou à correção de erros, expansão e aprimoramento desta base, seguindo as recomendações da literatura e as sugestões dos usuários iniciais. No momento, o BancoWeb é uma base pública que dá acesso livre a uma quantidade suficiente de imagens com seus respectivos diagnósticos e outras informações dos equipamentos utilizados. O foco é não só fornecer uma ampla quantidade de imagens, mas também oferecer ferramentas únicas que facilitem a utilização e, ao mesmo tempo, o desenvolvimento futuro da própria base.

As imagens oferecidas na base *online* foram selecionadas e inseridas de acordo com as exigências encontradas na literatura. Até a presente data, ela proporciona para *download* cerca de 1.500 imagens, provenientes de 3 hospitais e digitalizadas usando 2 *scanners* a laser diferentes. Todas as imagens têm 12 bits de resolução de contraste (3.500 a 3.800 níveis de cinza) e resolução espacial entre 0,150mm e 0,075mm.

Comparações do BancoWeb com outras bases mostraram que o projeto é compatível com o que o existe disponível a nível mundial em termos de quantidade e qualidade e oferece ferramentas e características únicas que podem facilitar muito o uso por desenvolvedores.

Em termos de resolução de contraste, as imagens estão na qualidade recomendada pela literatura (acima de 10 bits), mas a resolução espacial, apesar de

dentro dos parâmetros da literatura (próximo a 0,100mm), está no limite inferior em comparação a outras bases existentes. Conforme a base continuar a crescer e se desenvolver, mais imagens de melhor resolução espacial se tornarão disponíveis.

É importante ressaltar que a resolução espacial não é o maior limitador de qualidade das imagens mamográficas digitais, sendo a resolução de contraste um fator muito mais importante para a correta visualização e processamento de imagens. A maior dificuldade em uma imagem mamográfica é garantir suficiente diferença de contraste entre os elementos anormais e os tecidos normais da mama. Mesmo no caso de estruturas particularmente pequenas, como microcalcificações, a capacidade de detecção (seja ela visual ou por CAD) dos elementos será mais afetada pelos limites de resolução do mamógrafo e as variações locais do ponto focal – ou seja, do processo de aquisição – do que pela resolução espacial da imagem digital (desde que, evidentemente, ela apresente pixels de tamanho inferior ao menor tamanho de ponto focal de tubo utilizado para a aquisição da mamografia).

Outro fator a considerar, no momento, é a inexistência de imagens provenientes de mamógrafos digitais, ou seja, imagens obtidas de sistemas mamográficos digitais (diretos – como os do tipo FFDM\* – ou indiretos – como os CR\*\*). Apesar destes sistemas já serem comuns em outros países, ainda não o são no Brasil; por isso, obter tais imagens se torna mais difícil – inclusive também porque os equipamentos desse tipo existentes no País, em particular os de aquisição direta, encontram-se frequentemente em clínicas ou hospitais privados. É importante ressaltar, no entanto, que a base está preparada para inserção de imagens provenientes de mamógrafos digitais e que se projeta para um futuro próximo o início da inserção desse tipo de arquivo na base. Também para um futuro próximo é esperado preparar a base para a inserção e recuperação direta de arquivos seguindo o padrão DICOM.

As ferramentas implementadas atualmente incluem um sistema de busca amplo e fácil de usar, que permite ao usuário selecionar exatamente as características da imagem que deseja localizar na base. Até o momento, esta busca só pode ser feita por meio de informações inseridas na base (ver seção 5.4.2, para mais detalhes das informações inseridas), pois nenhum sistema de recuperação de imagem baseado em conteúdo foi

---

\* De *Full Field Digital Mammography*, com aquisição direta da imagem em placas eletrônicas e subsequente conversão do sinal em digital.

\*\* De *Computed Radiography*, com aquisição do sinal em placas sensoras do tipo *Imaging Plates*, descarregadas por varredura laser, com a transformação posterior do sinal luminoso em digital.

implementado. O BancoWeb é a única base pública de imagens mamográficas localizada na literatura a oferecer um sistema de busca que seja funcional atualmente.

A possibilidade de inserção remota apresentada nesta base também é única entre as bases pesquisadas. Esta ferramenta permite que qualquer usuário autorizado possa inserir imagens na base de qualquer lugar do mundo, seguindo o respectivo protocolo de inserção. O objetivo deste sistema é permitir que a base continue crescendo independentemente da continuidade ou não do projeto inicial, mantendo o BancoWeb uma base válida e útil. Além disso, vislumbra-se a possibilidade de que um usuário autorizado ao status de “pesquisador” possa utilizar o próprio sistema aqui descrito para estabelecer uma base pessoal de armazenamento remoto, cujo acesso seja definido por ele (obviamente, em conjunto com a administração do BancoWeb). Até o momento, não existem usuários externos ao LAPIMO autorizados a tal inserção – até mesmo por inexistência de pedidos. A expectativa é que futuras publicações sobre a base (incluindo este trabalho) e relações com outros laboratórios e pesquisadores tragam mais usuários qualificados interessados em adicionar imagens à base.

Outra ferramenta disponível na base que merece uma discussão é o sistema de recorte de imagens, que permite não só recortar e baixar apenas regiões de interesse do usuário como também a associação a um conjunto de regiões de interesse previamente demarcadas que também podem ser baixadas diretamente. Até a data de publicação deste trabalho, poucas regiões de interesse foram demarcadas para a base de recortes, apesar do sistema estar completamente funcional.

Apesar de algumas restrições terem sido discutidas aqui, é importante ressaltar que elas podem ser superadas sem grandes esforços, dependendo apenas principalmente da inserção de mais e imagens diferentes – e, eventualmente, adição de outras ferramentas que se mostrarem necessárias de acordo com a utilização dos arquivos –, o que deve ser feito conforme o projeto tiver continuidade.

O BancoWeb se propunha em seu projeto inicial a formar uma base de imagens padrão para avaliação comparativa de esquemas CAD em mamografia, ou seja, uma base que pudesse ser utilizada mundialmente para testes de CAD de modo que os resultados possam ser corretamente comparados.

Na literatura, a base CALMa, apesar de apresentar uma quantidade respeitável de imagens (3.000) de boa qualidade, é uma base fechada e, aparentemente, sem pretensão de ser aberta ao público desenvolvedor. A UCSF tem as melhores resoluções

dentre as bases pesquisadas, mas não apresenta uma quantidade suficiente de imagens e é paga (US\$ 100 por 198 imagens). A MIAS, sendo a mais antiga base entre as que puderam ser comparadas, apresenta uma resolução de contraste muito baixa (apenas 8 bits) e poucas imagens (320) para um trabalho completo. A DDSM apresenta a maior quantidade de imagens disponível publicamente (mais de 10.000), de boa qualidade e com grande variedade de achados demarcados, porém o uso de um formato não convencional para os arquivos e a falta de um sistema de busca (especialmente considerando a quantidade de imagens disponíveis) dificultam o uso da mesma.

Deve-se comentar ainda que, no Brasil, em 2004, foi publicada a descrição de um projeto chamado AMDI (GULIATO et al., 2004) que descreve um atlas digital de imagens mamográficas. O projeto pretendia desenvolver uma base de imagens com ferramentas de busca por informações textuais e por conteúdo, para pesquisa com CAD e também apresentar grande quantidade de informações sobre o paciente para análises estatísticas, incluindo informações detalhadas sobre hábitos (fumante, sedentário, uso prolongado de remédios, etc.), histórico familiar e outras características. No entanto, apesar do projeto ter sido dado como concluído em 2007, nenhuma nova publicação foi encontrada e o *site* do projeto ([www.lcc.ufu.br/amdi](http://www.lcc.ufu.br/amdi)) apresenta erros que impedem o acesso ao sistema. Até a presente data, não houve resposta às tentativas de contatos estabelecidas pelos mecanismos disponíveis naquele *site*.

Em conclusão, a base apresentada neste trabalho oferece uma grande quantidade de imagens da qualidade recomendada pela literatura em um formato de fácil uso e amplamente utilizado na área médica, além de contar com um sistema único de busca que permite a recuperação apenas das imagens que interessem à pesquisa do usuário. O acesso é público e livre para qualquer interessado e o projeto deve-se manter ativo no futuro próximo, garantindo inserção de novas e diferentes imagens e adição de novas ferramentas. Comparando com outras bases amplamente utilizadas na área, estas características fazem do BancoWeb uma ferramenta única e poderosa para o pesquisador da área, oferecendo mais opções do que qualquer outra base disponível.

### **Futuros desenvolvimentos**

Além dos possíveis novos desenvolvimentos discutidos acima, outros passos futuros deste projeto incluem integração com o padrão DICOM, permitindo que os

arquivos sejam inseridos diretamente na base e que exames inteiros sejam baixados já compatíveis com o padrão.

Com a base estabelecida, este projeto pretende combiná-la com um esquema CAD (em desenvolvimento pela equipe do LAPIMO), criando um sistema *online* não só para recuperação de imagens, mas de auxílio direto ao diagnóstico. Assim, para ampliar as capacidades de pesquisa da base, futuramente, ela deve conter um sistema automatizado de busca por conteúdo das imagens, analisando a imagem em si e oferecendo opções similares para o usuário.

Também é proposto que a base possa ser utilizada como um repositório remoto de imagens para qualquer usuário de nível “pesquisador”, permitindo inserção de imagens e seleção de quantas e quais delas deseja tornar pública, mantendo as outras para seu uso pessoal.

Como foi discutido anteriormente, é necessário demarcar as regiões de interesse das imagens disponíveis, formando uma base de recortes mais ampla. Esta demarcação deve ser feita por usuários com autorização de “pesquisador”. Foi sugerido usar paralelamente um sistema automático para detectar as regiões. Caso isto seja feito, as regiões demarcadas por “pesquisadores” e por sistemas automáticos devem ser claramente diferenciadas para o usuário final.

É interessante ressaltar que a estrutura de inserção e recuperação apresentada neste trabalho não precisa se restringir a imagens mamográficas. Seria necessária apenas a adição de novas tabelas no modelo para acrescentar à base imagens de mama proveniente de ultrassom ou ressonância magnética, por exemplo, ou até mesmo imagens de outros órgãos que possam ser úteis a pesquisadores de CAD em diferentes áreas.

## Referências

- ABUBAKER, A. A. et al. Efficient Pre-processing of USF and MIAS Mammogram Images. **Journal of Computer Science**, v. 2, n. 3, p. 67-75, 2006.
- ACHOUR, M. et al. **PHP Manual**, 2010. Disponível em: <<http://www.php.net/manual/en/>>. Acesso em: 23 fevereiro 2010.
- ADOBE DEVELOPERS ASSOCIATION. **TIFF: REvision 6.0 - Technical specification**. [S.l.]: [s.n.], 1992.
- AMENDOLIA, S. R. et al. The CALMA project. **Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment**, v. 461, n. 1-3, p. 428-429, 2001.
- AMERICAN COLLEGE OF RADIOLOGY. **American College of Radiology (ACR) Breast Imaging Reporting and Data System Atlas (BI-RADS® Atlas)**. [S.l.]. 2003.
- ANDERSON, B. E. A. **Guideline implementation for breast healthcare in low-income and middle-income countries: overview of the Breast Health Global Initiative Global Summit 2007**. [S.l.]: [s.n.]. 2008.
- BACHMAN, C. W. **Network (CODASYL) Data Model**. CODASYL Consortium. [S.l.]. 1969.
- BENATTI, R. H. **Elaboração de banco de imagens mamográficas digitalizadas**. São Carlos: [s.n.], 2003.
- BENATTI, R. H.; SCHIABEL, H.; NUNES, F. L. S. Construção de uma base de mamogramas digitalizados para utilização em sistemas de diagnóstico auxiliado por computador. **Anais do XXIV Encontro Nacional de Física da Matéria Condensada**, São Lourenço (MG), 2001.
- BENATTI, R. H.; SCHIABEL, H.; NUNES, F. L. S. A mammographic images database for computer-aided diagnosis schemes. **World Congress on Medical Physics and Biomedical Engineering**, Sydney, 2003.
- BERNSTEIN, P. A.; NEWCOMER, E. **Principles of Transaction Processing**. 2nd edition. ed. [S.l.]: Elsevier, 2009.
- BRZAKOVIC, D.; LUO, X. M.; BRZAKOVIC, P. An approach to automated detection of tumors in mammograms. **IEEE Transactions on Medical Imaging**, v. 9, n. 3, p. 233-241, 1990.
- CHAMBERLIN, D. D.; BOYCE, R. F. **SEQUEL: A Structured English Query Language**. ACM SIGMOD Workshop on Data Description, Access and Control. [S.l.]: Ann Arbor, Michigan. 1974. p. 249-264.
- CHAN, H.-P. et al. Computer-aided detection of microcalcifications in mammograms: methodology and preliminary clinical study. **Investigative Radiology**, v. 23, n. 9, p. 664-671, 1998.
- CHAN, H.-P. et al. Improvement of radiologists' characterization of ammmographic masses by using computer-aided diagnosis: an ROC study. **Radiology**, v. 23, n. 9, p. 817-827, 1999.
- CHENG, H. D. et al. Computer-aided detection and classification of microcalcification in mammograms: a survey. **Pattern Recognition**, v. 36, n. 12, p. 2967, 2003.
- CODD, E. F. **A Relational Model of Data for Large Shared Data Banks**. Communications of the ACM. [S.l.]. 1970.
- DAVIES, D. H.; DANCE, D. R. Automatic computer detection of clustered calcification in digital mammograms. **Physics in Medicine and Biology**, v. 35, n. 8, p. 1111-1118, 1990.

ELTER, M. et al. A modern benchmark case database for the computer-aided diagnosis of breast cancer. **Int J Comput Assist Radiol Surg**, n. 2 (S1), p. 514, 2007.

ESCARPINARTI, M. C. **Investigação de formatos e compressão de imagens digitais para processamento de imagens mamográficas de mamas densas**. São Carlos: Dissertação (Mestrado em Engenharia Elétrica S. Carlos) - Universidade de São Paulo, 2002.

FANTACCI, M. E. et al. **Search of microcalcification clusters with the CALMA CAD station**. Proceedings of SPIE. [S.l.]: [s.n.]. 2002. p. 1301-1310.

FREITAS, L. P. et al. Banco de imagens mamográficas on-line para consulta e testes em técnicas de processamento de imagens e esquemas CAD. **Anais do III SIIM**, São Carlos (SP), 2007.

GEORGIAN-SMITH, D.; MOORE, R. H.; HALPERN, E. Blinded Comparison of Computer-Aided Detection with Human Second Reading in Screening Mammography. **AJR Women's Imaging**, n. 189, p. 1135-1141, 2007.

GIGER, M. I. Computer analyzes of images in the detection and diagnosis of breast cancer. **Seminars in ultrasonod, CT and MRI**, v. 25, p. 411-418, 2004.

GULIATO, D. et al. **AMDI – Um Atlas Indexado para Mamografias Digitais via WEB**. III CLAEB. João Pessoa: IFMBE Proc. 2004. p. 529-532.

HADJIISKI, L.; CHAN, H.-P.; SAHINER, B. E. A. Improvement in radiologists' characterization of malignant and benign breast masses on serial mammograms with computer-aided diagnosis: an ROC study. **Radiology**, v. 233, p. 255-265, 2004.

HEATH, M. et al. The Digital Database for Screening Mammography. **Proceedings of the Fifth International Workshop on Digital Mammography**, p. 212-218, 2001. ISSN ISBN 1-930524-00-5.

JIAN, Y.; NISHIKAWA, R. M.; PAPAIOANNOU, J. **Requirement of microcalcification detection for computerized classification of malignant and benign clustered microcalcification**. SPIE International Symposium Medical Imaging - Image Processing. [S.l.]: [s.n.]. 1998. p. 313-317.

KALLERGI, M. Computer-aided diagnosis of mammographic microcalcification clusters. **Medical Physics**, v. 31, n. 2, p. 314-326, 2004.

LAI, S. M.; LI, X.; BISCHOF, W. F. On techniques for detecting circumscribed masses in mammograms. **IEEE Transactions on Medical Imaging**, v. 8, p. 377-386, 1989.

LAPIMO, 2009. Disponível em: <<http://lapimo.sel.eesc.usp.br>>. Acesso em: 23 fevereiro 2010.

LAURIA, A. et al. The CALMA system: an artificial neural network method for detecting masses and microcalcifications in digitized mammograms. **Nuclear Instruments and Methods in Physics Research A**, n. 518, p. 391-393, 2004.

LAWRENCE LIVERMORE NATIONAL LIBRARY. **UCSF Digital Mammogram Database**. Center for Health Care Technologies Livermore. Livermore, CA, USA.

LUO, P.; QIAN, W.; ROMILLY, P. CAD-Aided Mammogram Training. **Journal of the Association of University Radiologists**, v. 12, n. 8, p. 1039-1048, 2005.

MATHEUS, B. R. N.; SCHIABEL, H. Base de imagens mamográficas na Internet para auxílio no desenvolvimento e avaliação de esquemas de diagnóstico auxiliado por computador. **Anais do XXI Congresso Brasileiro de Engenharia Biomédica**, Salvador (BA), 2008.

METHA, D.; SAHNI, S. **Handbook of Data Structures and Applications**. [S.l.]: Chapman and Hall/CRC Press, 2007.

- MYSQL AB. MySQL 5.5 Reference Manual, 2008. Disponível em: <<http://dev.mysql.com/doc/refman/5.5/en/index.html>>. Acesso em: 23 fevereiro 2010.
- NATIONAL ELECTRICAL MANUFACTURERS ASSOCIATION. **Digital Imaging and Communications in Medicine (DICOM)**. Rosslyn, Virginia: National Electrical Manufacturers Association, 2009.
- NISHIKAWA, R. M. . E. A. Effect of case selection of the performance of computer-aided detection schemes. **Medical Physics**, n. 21, p. 265-270, 1994.
- NISHIKAWA, R. M. . Y. L. M. Variations in measured performance of CAD schemes due to database composition and scoring protocol. **SPIE 99: Medical Imaging**, n. 3338, p. 840-844, 1999.
- NISHIKAWA, R. M. Mammographic databases. **Breast Disease**, n. 10(3,4), p. 137-150, 1998.
- NISHIKAWA, R. M. E. A. Performance of automated CAD schemes for the detection and classification of clustered microcalcifications. **Digital mammography, Elsevier Science Publishers B.V.**, p. 13-20, 1994.
- NUNES, F. et al. Comparisons of different contrast resolutions effects on a computer-aided detection system intended to cluster microcalcification detection in dense breast images. **Journal of Digital Imaging**, v. 14, n. 2, p. 217-219, 2001.
- NUNES, F. L. S. et al. Um sistema computacional para registrar e recuperar imagens mamográficas via-Internet. **Anais do II Simpósio Brasileiro de Qualidade de Software/III Workshop de Informática Médica**, Fortaleza (CE), 2003.
- NUNES, F. L. S. et al. BancoWeb: registrando, recuperando e avaliando o uso de uma base de imagens mamográficas para Internet. **Anais do XXXI SEMISH (Seminário Integrado de Software e Hardware)**, Salvador (BA), 2004.
- NUNES, F. L. S. et al. Base de imagens na Internet para avaliação de diagnóstico auxiliado por computador em mamografia. **Anais do III Congresso Latino-Americano de Engenharia Biomédica (XIX Congresso Brasileiro de Engenharia Biomédica)**, João Pessoa (PB), 2004.
- NUNES, F. L. S.; SCHIABEL, H.; BENNATI, R. H. Contrast enhancement in dense breast images using the modulation transfer function. **Medical Physics**, v. 29, n. 12, p. 2925-2936, 2002.
- NUNES, F. L. S.; SCHIABEL, H.; GOES, C. E. Contrast enhancement in dense breasts images to aid clustered microcalcifications detection. **Journal of Digital Imaging**, v. 20, p. 53-66, 2007.
- OPPEL, A. **Databases Demystified**. San Francisco: McGraw-Hill Osborne Media, 2004.
- ORACLE CORPORATION. Profit: the executive's guide to oracle applications. [S.l.]: [s.n.], 2007. p. 26-33.
- ORACLE CORPORATION. **Oracle Database 11g Release 2 (11.2) Documentation**. [S.l.]. 2009.
- PEIXOTO, J. E.; CANELLA, E.; AZEVEDO, A. C. **Mamografia: da prática ao controle**. Rio de Janeiro: Gráfica Esdeva, 2007.
- PETRI, G. A comparison of Oracle and MySQL. **SELECT Journal**, n. 1st Qtr, p. 41-48, 2005.
- QIAN, W. E. A. Computer assisted diagnosis for digital mammography. **IEEE Engineering in Medicine and Biology**, v. 14, n. 5, p. 561-569, 1995.
- ROSE, C. et al. Web services for the DDSM and digital mammography research, 2003. Disponível em: <<http://marathon.csee.usf.edu/~ddsm/search.html>>. Acesso em: 2010.
- SAMPAT, M. P.; MARKEY, M. K.; BOVIK, A. C. Computer-Aided detection and diagnosis in mammography. In: BOVIK, A. C. (. **Handbook of image and video processing**. [S.l.]: Elsevier academic press, 2005. p. 1195-1217.



- SANDER, L. G. **Data Modeling**. [S.l.]: [s.n.], 1995.
- SCHIABEL, H. et al. Performance of a processing scheme for clustered microcalcifications detection with different images database. **Proceedings of the 22nd Annual International Conference of the IEEE**, 2, n. 2000, 2000.
- SCHIABEL, H. et al. Investigations on the effect of different characteristics of images sets on the performance of a processing scheme for microcalcifications detection in digital mammograms. **Journal of Digital Imaging**, v. 14, n. 2 (1), p. 224-225, 2001.
- SCHIABEL, H.; ESCAPINARTI, M. C.; FREITAS, L. P. Base de imagens mamográficas digitalizadas para consulta e uso em testes de esquemas CAD. **XXXV Congresso Brasileiro de Radiologia**, Curitiba (PR), v. 39, p. 4-5, 2006.
- SELINGER, P. G. et al. **Access Path Selection in a Relational Database Management System**. Proceedings of the 1979 ACM SIGMOD International Conference on Management of Data. [S.l.]: [s.n.]. 1979. p. 23-24.
- STALLMAN, R. **GNU GENERAL PUBLIC LICENSE, version 3**, 2007. Disponível em: <<http://www.gnu.org/licenses/gpl-3.0.txt>>. Acesso em: 23 fevereiro 2010.
- SUCKLING J, E. A. The Mammographic image analysis society digital mammogram database. **Exerpta Medica International Congress**, n. 1069, p. 375-378, 1994.
- SUCKLING, J. PEIPA, the Pilot European Image Processing Archive. **The mini-MIAS database of mammograms**, 2003. Disponível em: <<http://peipa.essex.ac.uk/info/mias.html>>. Acesso em: 2009.
- TAYLOR, W. R.; FRANK, R. L. CODASYL Data-Base Management Systems. **ACM Computing Surveys (CSUR)**, v. 8, n. 1, p. 67-103, March 1976.
- TSICHRITZIS, D. C.; LOCHOVSKY, F. H. Hierarchical Data-Base Management: A Survey. **ACM Computing Surveys (CSUR)**, v. 8, n. 1, p. 105-123, March 1976.
- U. S. FOOD AND DRUG ADMINISTRATION. **Summary and safety of Effectiveness Data: R2 Technologies**. [S.l.]. 1998. (P970058).
- U. S. FOOD AND DRUGS ADMINISTRATION. **Summary and safety of Effectiveness Data: CADx Medical Systems**. [S.l.]. 2002. (P010034).
- U. S. FOOD AND DRUGS ADMINISTRATION. **Summary and safety of Effectiveness Data: ISSI**. [S.l.]. 2002. (P010038).
- VIEIRA, M. A. C. **Metodologia baseada nas funções de transferência para pré-processamento de imagens mamográficas digitais e sua aplicação em esquema computacional de auxílio ao diagnóstico**. [S.l.]: Tese (Doutorado em Engenharia Elétrica) - USP - Escola de Engenharia de São Carlos, 2005.
- VYBORNÝ, C. J.; GIGER, M. L.; NISHIKAWA, R. M. Computer-aided detection and diagnosis of breast cancer. **Radiologic Clinics of North America**, v. 38, n. 4, p. 725-740, 2000.
- WIDENIUS, M.; AXMARK, D. **MySQL**, 2010. Disponível em: <<http://www.mysql.com/>>. Acesso em: 23 fevereiro 2010.
- WIEDERHOLD, G.; EL-MASRI, R. The structural model for database design. In: \_\_\_\_\_ **Entity-Relationships Approach to Systems Analysis and Design**. [S.l.]: [s.n.], 1980. p. 237-257.
- WINSBERG, G. F. et al. Detection of radiographic abnormality in mammograms by means of optical scanning and computer analysis. **Radiology**, v. 89, p. 211-215, 1967.

# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)