UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE SÃO CARLOS
Departamento de Engenharia Elétrica

PÂMELA SERRA RODRIGUES

**Ferramentas de interação e gerenciamento do *BancoWeb* -
uma base de imagens para auxílio a pesquisas na área de
mamografia digital.**

São Carlos
2018

PÂMELA SERRA RODRIGUES

Ferramentas de interação e gerenciamento do *BancoWeb* - uma base de imagens para auxílio a pesquisas na área de mamografia digital.

Dissertação de mestrado apresentada à Escola de Engenharia de São Carlos da Universidade de São Paulo, como parte dos requisitos para a obtenção do título de mestre em Ciências no Programa de Engenharia Elétrica.

Área de concentração: Processamento de Sinais e Instrumentação

Orientador: Prof. Dr. Homero Schiabel

São Carlos

2018

Trata-se da versão corrigida da dissertação. A versão original se encontra disponível na EESC/USP que aloja o programa de Pós-Graduação de Engenharia Elétrica.

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

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da
EESC/USP com os dados inseridos pelo(a) autor(a).

R696f   Rodrigues, Pâmela Serra
            Ferramentas de interação e gerenciamento do
BancoWeb - uma base de imagens para auxílio a pesquisas
na área de mamografia digital. / Pâmela Serra
Rodrigues; orientador Homero Schiabel. São Carlos,
2018.

            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,
2018.

            1. Mamografia. 2. base de imagens. 3. Diagnóstico
auxiliado por computador (CAD). 4. imagens de Phantom.
I. Título.

Eduardo Graziosi Silva - CRB - 8/8907

FOLHA DE JULGAMENTO

Candidata: Tecnóloga **PÂMELA SERRA RODRIGUES.**

Título da dissertação: "Ferramentas de interação e gerenciamento BancoWeb- uma base de imagens para auxílio a pesquisas na área de mamografia digital".

Data da defesa: 23/11/2018.

Comissão Julgadora:

Resultado:

Prof. Associado **Homero Schiabel**
(Orientador)
(Escola de Engenharia de São Carlos/EESC)

APROVADA

Profa. Dra. **Ana Cláudia Patrocínio**
(Universidade Federal de Uberlândia/UFU)

APROVADA

Profa. Dra. **Renata de Freitas Góis**
(Faculdade de Tecnologia de Araçatuba/FATEC)

Aprovada

Decano da CCP:
Prof. Associado **Luís Fernando Costa Alberto**

Presidente da Comissão de Pós-Graduação:
Prof. Associado **Luís Fernando Costa Alberto**

Dedico...

Ao meu namorado Rafael Marchi e aos meus pais, Adão e Susana.

AGRADECIMENTOS

À minha família, que sempre me apoiou em todas as minhas decisões, sempre esteve comigo em todos os momentos.

Ao meu namorado Rafael Marchi, por todo apoio, auxílio e confiança que depositou em mim durante este mestrado.

Ao Prof. Dr. Homero Schiabel, pelo auxílio e direção durante a orientação deste mestrado, acarretando conhecimento para a minha trajetória profissional.

Ao Prof. Dr. Bruno Roberto Nepomuceno Matheus, por me encaminhar neste projeto.

À Dra. Maria Angélica Zucareli Sousa, pelo apoio durante o projeto.

Aos membros do LAPIMO, pelas sugestões e auxílio sempre que precisei.

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

RESUMO

Rodrigues, P. S., **Ferramentas de interação e gerenciamento do *BancoWeb* - uma base de imagens para auxílio a pesquisas na área de mamografia digital**. 2018. 79 f. Dissertação (Mestrado) – Escola de Engenharia de São Carlos – Universidade de São Paulo, São Carlos, 2018.

O *BancoWeb* é uma base de imagens para auxílio a pesquisa na área de mamografia digital. Oferece o gerenciamento da base de dados com um conjunto amplo de imagens mamográficas de qualidade apropriada ao processamento digital e, principalmente, comparação entre diferentes esquemas. O sistema de busca da base de dados/imagens mamográficas está disponível desde 2010, sendo *on-line* e gratuito. Com os avanços em pesquisas durante todos estes anos, a necessidade e demanda de melhorias, mudanças e incrementos no sistema de gerenciamento e buscas foram surgindo. Por isso, os objetivos deste trabalho estão em desenvolver e aperfeiçoar recursos na busca de imagens, tipos de achado, permitir a classificação pelo sistema CAD, além de implementar um novo sistema de gerenciamento e armazenamento de imagens de um novo *phantom*, desenvolvido como resultado de pesquisa ainda mais recente no grupo. Testes comparativos com outras bases análogas que disponibilizam imagens mamográficas para download faz do *BancoWeb* uma ferramenta única e poderosa para o pesquisador e interessado da área, oferecendo mais opções que as demais bases disponíveis atualmente.

Palavras Chaves: Mamografia, base de imagens, Diagnóstico auxiliado por computador (CAD), imagens de *Phantom*.

ABSTRACT

Rodrigues, P. S. **Interaction and management tools of BancoWeb - a database of images to aid research in the area of digital mammography**. 2018. 79 f. Dissertação (Mestrado) – Escola de Engenharia de São Carlos – Universidade de São Paulo, São Carlos, 2018.

BancoWeb is a database of images to aid research in the field of digital mammography. It offers the management of the database with a wide set of mammographic images of appropriate quality to the digital processing and, mainly, comparison between different schemes. The database search system / mammographic images are available since 2010, being online and free. With advances in research over the years, the need for and demand for improvements, changes, and increases in the management and search system have been emerging. Therefore, the objectives of this work are in development and improvement of resources in the search of images, types of find, allow the selection of CAD, in addition to implementing a new system of management and storage of images of a new phantom, developed as a result of even more recent research in the group. Comparative tests with other analogous databases that provide mammographic images for download makes BancoWeb a unique and powerful tool for the researcher and the interested party, offering more options than the other databases currently available.

Key words: mammography, images database, computer-aided diagnosis (CAD), images *Phantom*.

LISTA DE ILUSTRAÇÕES

Figura 1: Exemplo do modelo relacional	21
Figura 2: Exemplo de modelo orientado a objeto.....	23
Figura 3: Exemplo de anatomia de um SQL	26
Figura 4: Modelo relacional da base - versão original	38
Figura 5: Painel de busca, informação do hospital oculta (informação era exibida entre o campo de sexo e idade; também na tabela de dados do aparelho de Raios-X).	43
Figura 6: Resultado da busca, informação do hospital oculta.	44
Figura 7: Mais informação, informação do hospital oculta.	44
Figura 8: Cadastro do Tipo de Achado, inclusão da característica adicional.....	45
Figura 9: Informações do Tipo de Achado no cadastro de exame.	45
Figura 10: Resultado da busca, com a informação da característica adicional.	46
Figura 11: Tabela de Tipo de Achado, inclusão da característica adicional	46
Figura 12: Cadastro de exame, inclusão da fonte da imagem	47
Figura 13: Busca de imagens, informação da fonte da imagem.....	47
Figura 14: Tabela de Exame, inclusão da fonte da imagem	48
Figura 15: Recorte da imagem, (a) botão para inverter imagem; (b) imagem negativa.	48
Figura 16: Recorte, recurso de ajuda	49
Figura 17: Ferramenta para classificar imagem com CADx	50
Figura 18: Resultado da classificação	50
Figura 19: Reconstrução do resultado da busca	52
Figura 20: Resultado da busca, detalhes expandidos	53
Figura 21: Recurso (em destaque) para baixar exame e modelo do arquivo .TXT com os dados respectivos	54
Figura 22: Diagrama novas tabelas para arquivo e gerenciamento das imagens de <i>phantom</i> ..	56
Figura 23: Menu com opções adicionais para cadastro e acesso a imagens de <i>phantom</i>	56
Figura 24: Cadastro de Achado de <i>Phantom</i>	57
Figura 25: Cadastro de equipamento.	58
Figura 26: Cadastro de tipo de achado, inclusão do campo "pertence a".....	58
Figura 27: Cadastro de hospital	59
Figura 28: Cadastro de imagens de <i>phantom</i>	59
Figura 29: Cadastro de achado, vínculo com a imagem.....	60

Figura 30: Busca por imagens de <i>phantom</i>	61
Figura 31: Resultado da busca de imagens de <i>phantom</i>	61
Figura 32: Resultado da busca, informações expandidas	62
Figura 33: Gabarito da imagem	62
Figura 34: Mais informações da imagem de <i>phantom</i>	63
Figura 35: Arquivo .TXT disponibilizado ao usuário	64
Figura 36: Perfil de tipo de achado para imagens de <i>phantom</i>	64
Figura 37: Sobre as imagens de <i>phantom</i>	65
Figura 38: Estatística tipo de achado das imagens reais	71
Figura 39: Estatística tipo de achado das imagens de <i>phantom</i>	72

LISTA DE TABELAS

Tabela 1: Lista de mamógrafos e local de origem.....	37
Tabela 2: Comparativo de bases mamográficas disponíveis no mundo.....	67
Tabela 3: Comparativos (2018).....	68

LISTA DE ABREVIATURAS E SIGLAS

BI-RADS	<i>Breast Imaging-Reporting and Data System</i>
BCDR	<i>Breast Cancer Digital Repository</i> (Repositório Digital de Câncer de Mama)
CAD	<i>Computer-aided Diagnosis</i> (Diagnóstico auxiliado por computador)
CADX	<i>Computer-Aided Diagnosis</i> (Diagnóstico auxiliado por computador)
CALMa	<i>Computer-Assisted Library for Mammography</i>
CC	Crânio-caudal
CR	<i>Computed Radiography</i>
DDSM	<i>Digital Database for Screening Mammography</i>
DICOM	<i>Digital Imaging and Communications in Medicine</i>
DR	<i>Digital Radiography</i>
ECM	Exame clínico de mama
FFDM	Full-Field Digital Mammography
HCFMRP-USP	Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto
HD	<i>Hard drive</i> (disco rígido)
HTML	Hyper Text Markup Language (Linguagem para Marcação de Hiper-texto)
INCA	Instituto Nacional de Câncer
INFN	Instituto Nazionale di Fisica Nucleare (Instituto Nacional de Física Nuclear)
IIS	Internet Information Service
LAPIMO	Laboratório de Análise e Processamento de Imagens Médicas e Odontológicas
LLNL/UCSF	<i>Lawrence Livermore National Laboratories/University of California at San Fransisco</i>
LP/MM	Pares de linhas por milímetro
MIAS	<i>Mammographic Image Analysis Society</i>
MLO	Médio-Lateral Oblíqua
MYSQL	Dialeto do SQL
OO	Orientado a objeto
PVC	Cloreto de Polivinil

SQL

Structured Query Language

SGBD

Sistema Gerenciador de Banco de Dados

SUMÁRIO

Capítulo 1: Introdução	17
1.1 Câncer de Mama	17
1.2 Detecção e diagnóstico de câncer de mama	17
1.3 Esquemas CAD	18
1.4 Objetivos	19
1.5 Disposição do trabalho	19
Capítulo 2: Base de Dados	20
2.1 Definição de base de dados	20
2.2 Modelo de dados	20
2.2.1 Modelo relacional	21
2.2.2 Modelo orientado a objetos	22
2.3 Estrutura de dados	23
2.4 Mecanismo de transação	24
2.5 Linguagem de consulta	24
2.6 SQL	25
Capítulo 3: Bases de imagens mamográficas	27
3.1 Características de uma Base de imagens médicas	27
3.2 Requisitos de Bases de imagens mamográficas	27
3.2.1 Conjuntos de imagens	28
3.2.2 Interface	29
3.3 Bases existentes	29
3.3.1 Base de dados MIAS	29
3.3.2 DDSM	30
3.3.3 LLNL/UCSF	31

3.3.4 MammoGrid (antiga CALMa).....	31
3.3.5 INBreast.....	32
3.3.6 BCDR	33
3.3.7 <i>BancoWeb</i>	33
3.4 Conclusão.....	34
Capítulo 4: Metodologia	35
4.1 Histórico do projeto	35
4.1.1 Aquisição das imagens	36
4.1.2. Estruturação da base	38
4.1.3. Ferramentas utilizadas no desenvolvimento.....	40
4.1.4. Implementações do BancoWeb	40
4.2. Novas Implementações do <i>BancoWeb</i>	42
4.2.1 Alterações quanto à origem da imagem cadastrada.....	42
4.2.2 Cadastro de Tipo de Achado – Inclusão da característica adicional	44
4.2.3 Inclusão da fonte do exame DR, CR ou DG	46
4.2.4 Recorte – Recurso para obter o negativo na imagem	48
4.2.5 Ferramenta para classificar imagem, integração com esquema CADx	49
4.2.6 Reconstrução da Busca por imagens	51
4.2.7 <i>Download</i> completo do exame	53
4.2.8 Implementação do sistema de gerenciamento de arquivos de imagens de <i>Phantom</i>	55
Capítulo 5: Resultados	66
5.1 Acesso, manutenção e atualização	66
5.2 Quantidade de imagens	66
5.3 Sistemas de busca.....	69
5.4 Classificação com CADx	70
5.5 Imagens de <i>Phantom</i>	70
5.6 Perfil estatístico.....	71

5.7 Outras ferramentas	72
Capítulo 6: Discussão e Conclusões	73
Referências Bibliográficas	78

Capítulo 1: Introdução

1.1 Câncer de Mama

A taxa de mortalidade por câncer de mama em mulheres no mundo a cada ano, vem se agravando. De acordo com Instituto Nacional do Câncer do Brasil (INCA, 2018), este câncer é o segundo mais frequente no mundo, depois do de pele não melanoma, respondendo por cerca de 28% dos casos a cada ano. O câncer de mama também acomete homens, porém é raro, representando apenas 1% do total de casos da doença.

Segundo as estimativas do GLOBOCAN (BRAY, et al., 2018), haverá 18,1 milhões de novos casos de câncer, sendo 11,6% de casos de câncer de mama feminino (11,6%). Na maioria dos casos, o diagnóstico ocorre em estágios avançados, ocasionando a morte da paciente. O Governo (2017) relata que no Brasil, as mulheres enfrentaram, em 2017, 57.960 novos casos e 59.700 em 2018, de acordo com a estimativa do INCA (INCA, 2018). Na maioria dos casos, o diagnóstico ocorre em estágios avançados, ocasionando a morte da paciente.

A taxa de sobrevivência é relativa ao cuidado que as pessoas possuem, pois quanto antes forem diagnosticadas, maiores serão as chances de um tratamento efetivo, além de manter uma alimentação saudável e praticar exercícios. A American Cancer Society relata que, para mulheres diagnosticadas com câncer de mama após 5 anos, é de 91% as chances de sobrevivência, após 10 anos, 86%, e após 15 anos, 80% (AMERICAN CANCER SOCIETY, 2017).

1.2 Detecção e diagnóstico de câncer de mama

Existem dois métodos que são utilizados há muito tempo para o diagnóstico de câncer de mama são através do exame clínico de mama (ECM) e mamografia (OSHIRO, M. L., et al., 2014). Além destes métodos, existem autoexame das mamas, ultrassonografia, ressonância nuclear magnética, tomossíntese mamária e termografia.

O exame clínico de mama (ECM) consiste na inspeção visual e palpação das mamas feitas por um profissional de saúde capacitado, podendo detectar tumores de até 1,0cm, se superficiais. Segundo Silva e Hortale (2012), poucos países com programas de rastreamento ainda utilizam o ECM como exame diagnóstico. A melhora da acurácia do ECM está

relacionada à capacidade dos profissionais de saúde e à padronização dos achados e registros do mesmo.

A mamografia é realizada utilizando-se um equipamento de raios X chamado mamógrafo. Este exame é o principal indicado, já que possibilita o diagnóstico do câncer de mama em estágio inicial. Assim, favorece o tratamento precoce, mais efetivo e menos agressivo. Entretanto, com foco na detecção e avaliações de lesões extremamente pequenas, o processo de aquisição da imagem deve ser cauteloso, ou seja, deve ser feito com muita atenção. Portanto, a calibração e manutenção do equipamento é fundamental e depende da qualificação de técnicos especializados para realizar o exame.

De acordo com MIGOWSKI et al (2018), o aumento da sensibilidade dos exames de mamografia ao longo das últimas décadas pode ter resultado na eficácia do tratamento de casos de câncer localmente avançado identificados em decorrência de sinais ou sintomas, porém, ainda existem os casos de exames diagnosticado erroneamente.

1.3 Esquemas CAD

Existem inúmeros casos registrados de diagnóstico equivocado de câncer de mama (YUAN et al., 2018), alguns por falta de experiência dos médicos e profissionais da área e outros por falha dos exames. Para auxiliar os profissionais de saúde como uma segunda opinião, têm sido desenvolvidos os esquemas computadorizados de auxílio ao diagnóstico (*computer-aided diagnosis*), ou CAD, que constituem a aplicação de um algoritmo na leitura da imagem mamográfica digital, para detecção de possíveis lesões. O sistema computacional analisa as informações disponíveis e tenta apresentar um diagnóstico com base nos dados fornecidos pela imagem processada.

O objetivo do CAD, segundo Calas, Gutfilen e Albuquerque (2012), não é de diagnóstico, mas sim, alertar o radiologista para áreas suspeitas, reduzindo a taxa de resultados falsos-negativos em mamografias de rastreamento.

Os sistemas CAD podem ser utilizados para detecção ou para classificação. Normalmente, para a detecção, o computador encontra uma lesão não perceptível ao radiologista. Por outro lado, para a classificação, o radiologista indica uma área suspeita e o programa irá avaliar a probabilidade de malignidade.

O desenvolvedor de um sistema CAD necessita de um conhecimento em estruturas da mama e acesso a significativa quantidade de variedades de imagens com laudos confiáveis.

Durante os testes do desenvolvedor, o laudo do exame torna-se essencial, pois ele poderá comparar os resultados obtidos pelo CAD com os resultados dos laudos médicos.

A qualidade da imagem também é de suma importância, pois uma imagem com baixa qualidade altera os resultados esperados através do CAD. Ao obter uma imagem, as principais características, como quantidade de níveis de cinza, resolução espacial, qualidade do mamógrafo, níveis de ruído, e até mesmo a forma de armazenamento são de extrema importância (MATHEUS, 2010).

1.4 Objetivos

Os objetivos deste trabalho estão em desenvolver e aperfeiçoar recursos em uma base de dados/imagens mamográficas, denominada *BancoWeb*. As implementações propostas são na busca de imagens, tipos de achado, ocultar a origem da paciente (hospital), informar a fonte do exame, recurso para obter o negativo da imagem, permitir a classificação pelo sistema CAD, download completo do exame, além de implementar um novo sistema de gerenciamento e armazenamento de imagens de um novo *phantom*, desenvolvido como resultado de pesquisa ainda mais recente no grupo (SOUZA, MATHEUS, SCHIABEL, 2018).

Assim, embora o sistema já tivesse uma variedade de imagens com qualidade, oferecendo um conjunto básico de ferramentas computacionais para facilitar o estudo das imagens e localização das informações pertinentes a cada caso, este trabalho apresenta avanços que irão complementar e/ou aprimorar processos pré-existentes, (e) adicionar uma nova base de imagens (as de *phantom*) no sistema, além de oferecer novos recursos de gerenciamento e também processamento digital das imagens.

1.5 Disposição do trabalho

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

Capítulos 2 e 3: apresentam revisões bibliográficas sobre Base de Dados e Base de imagens mamográficas.

Capítulo 4: apresenta a metodologia do trabalho, incluindo as implementações realizadas e os detalhes.

Capítulo 5: apresenta os resultados do projeto e a comparação com as principais bases existentes.

Capítulo 6: apresenta as conclusões do projeto em relação aos objetivos estabelecidos.

Capítulo 2: Base de Dados

2.1 Definição de base de dados

Uma base de dados é um repositório de informações relacionadas com determinado assunto ou finalidade, ou seja, uma coleção de dados (CALDEIRA, C. P, 2017), sendo eles físicos ou digitais, como uma agenda de contatos, um livro, HD (*Hard drive* – disco rígido) externo ou um diretório armazenado na “nuvem”. Esse tipo de repositório na área médica é fundamental para possibilitar testes e pesquisas variadas; particularmente, no atual estado da arte dos processamentos computacionais, são fonte para validação e verificação da eficácia de diversos tipos de técnicas e ferramentas para auxílio, por exemplo, ao diagnóstico.

Na área de pesquisas relacionadas a CAD, bases de imagens/dados são fundamentais para testes tanto de desenvolvimento como de desempenho de esquemas computacionais que visam ao auxílio na identificação de diversos tipos de patologias em variadas áreas da Medicina (SCHIABEL, 2014).

Para computação, base de dados é um conjunto de informações que são manipuladas e acessadas por um *software* chamado de Sistema Gerenciador de Banco de Dados (SGBD). O SGBD é composto por quatro partes que são: o modelo de dados, estrutura de dados, linguagem de consulta e mecanismos de transação. Para Valente e Fujino (2016), as características desejáveis numa base de dados são:

- a) Controle de redundância;
- b) Compartilhamento de dados;
- c) Controle de acesso aos dados;
- d) Múltiplas interfaces;
- e) Representação de associações complexas;
- f) Garantia de restrições de integridade;
- g) Recuperação de falhas.

Uma base de dados torna-se segura e confiável quando todas as características são atendidas, principalmente quando se trata de uma base muito grande e que contém milhões de registros.

2.2 Modelo de dados

O modelo de dados define como os registros vão ser armazenados na base de dados e como será a sua associação. Para cada uso deve-se ter o cuidado da escolha certa do modelo de

dados. Assim, o objetivo e os tipos de registros que serão inseridos podem influenciar na escolha. Uma escolha inadequada pode causar lentidão e inconsistência de dados.

Segundo Takai, Italiano e Ferreira (2005), os modelos mais utilizados são:

- a) Relacional
- b) Orientado a objetos

O modelo hierárquico e o de rede não são mais utilizados. O hierárquico tem associação entre dois registros apenas. Já o de rede, é uma extensão do modelo hierárquico permitindo associar mais de dois registros.

2.2.1 Modelo relacional

O modelo relacional faz relações diretas entre cada tabela. Este modelo, que é o mais comum, classifica dados em tabelas, também conhecidas como relações, cada uma das quais consiste em colunas e linhas. Cada tabela estabelece uma das colunas como sendo um atributo-chave e este atributo deve ser único para cada registro de uma mesma tabela. O relacionamento de um registro com outro é através da chave, basta que um deles tenha uma coluna com esta chave (Figura 1).

Figura 1: Exemplo do modelo relacional

Fonte do autor.

No modelo relacional (ELMASRI e NAVATHE, 2005) cada linha na tabela representa um fato que corresponde a uma entidade ou relacionamento. O nome da tabela e os nomes das

colunas são usados para ajudar na interpretação do significado dos valores em cada linha. Por exemplo, a tabela de ALUNO, na Figura 1, possui as colunas *nome*, *número*, *turma* e *curso* e cada linha representa os fatos sobre uma entidade aluno em particular. Os nomes das colunas especificam como interpretar os valores de dados em cada linha, com base na coluna em que cada valor está. Todos os valores em uma coluna são do mesmo tipo de dado.

Uma linha é chamada de tupla, na terminologia do modelo relacional formal. Um cabeçalho de coluna é conhecido como atributo, e a tabela é chamada relação. O tipo de dado que descreve os tipos de valores que podem aparecer em cada coluna é representado pelo domínio de valores possíveis. Desta forma, qualquer tipo de informação pode ser armazenado facilmente, bastando que as chaves estejam corretamente associadas.

Este modelo permite maior liberdade na construção de uma base de dados (MATHEUS, 2010); em contrapartida, o custo computacional é consideravelmente elevado. Apenas na década de 1980 este sistema começou a valer a pena, pois os computadores se desenvolveram para tal modelo de dados.

2.2.2 Modelo orientado a objetos

De acordo com Takai, Italiano e Ferreira (2005) o modelo orientado a objetos (OO) começou a se tornar comercialmente viável em meados de 1980. O surgimento deste modelo ocorreu em função dos limites de armazenamento e representação semântica impostas no modelo relacional. Este modelo define a base de dados como uma coleção de objetos ou elementos de *software* reutilizáveis, com recursos de métodos associados.

Para um modelo ser considerado orientado a objeto ele deve conter alguns conceitos básicos (GALANTE, MOREIRA, BRANDÃO, 2007), como abstração, encapsulamento, herança e polimorfismo. A capacidade de modelar coisas do mundo real é chamada de abstração. O encapsulamento tem como benefício evitar que interferências externas afetem na manipulação dos dados. A herança é o mecanismo que permite reutilização de códigos e o polimorfismo significa, etimologicamente, várias formas.

Figura 2: Exemplo de modelo orientado a objeto.

Fonte do autor.

Uma empresa deseja controlar dados dos seus clientes. O programador colhe os dados genéricos dos clientes e os molda no que se chama de classe, conforme Figura 2. Uma classe descreve um grupo e objetos com propriedades (atributos), similares comportamentos (operações), relacionamentos comuns com outros objetos e uma semântica comum (ROCHA, 2001).

Em 2004, as bases de dados orientados a objeto tiveram um crescimento devido ao surgimento de bases de dados OO livres. Estes modelos ganharam espaço nas áreas como base de dados espaciais, telecomunicações, e nas áreas científicas como física de alta energia e biologia molecular (GALANTE, MOREIRA, BRANDÃO, 2007).

2.3 Estrutura de dados

Após selecionar o modelo de dados, é necessário estabelecer como estes dados serão armazenados, lidos e manipulados na memória do computador. Não devemos pensar apenas nos registros, mas também nas estruturas, principalmente no caso de tabelas do modelo relacional e orientado a objetos.

Elmasri e Navathe, 2005, relatam que a estrutura de dados deve ser não só apropriada para armazenamento, mas também, garantir um bom desempenho ao consultar e manipular os dados. Devido ao tamanho que a base de dados pode atingir, estas informações estarão armazenadas em um equipamento de memória permanente (o mais comum seria um HD).

Para uma aplicação, é fundamental que a estrutura de dados seja altamente otimizada para não se tornar um problema à velocidade do sistema. Atualmente, o aumento da

complexidade destes algoritmos e o desenvolvimento de bibliotecas otimizadas exige muito conhecimento de um desenvolvedor, além de compreender as funções e ferramentas que permitem trabalhar com tal estrutura.

2.4 Mecanismo de transação

Uma transação é uma sequência de operações executadas como uma única unidade lógica de trabalho, ou seja, deve garantir que funções operacionais realizadas em uma base de dados tenha coerência, confiabilidade e independência em relação a outros processos simultâneos ou não. Este mecanismo é responsável por garantir que, caso ocorram falhas do sistema e interrupção das execuções, não haja corrupção dos dados nem problemas para o uso da base.

O mecanismo de transação deve garantir ACID (Atomicidade, Consistência, Isolamento e Durabilidade) (STAIR,2011).

- **Atomicidade:** quando um conjunto de operações associadas for executado, todas devem funcionar ou nenhuma delas, deve ser indivisível. Todas as informações devem ser trazidas de volta ao estado anterior, caso ocorra algum problema.

- **Consistência:** a base de dados deve ser íntegra, as regras do modelo, estrutura e linguagem não podem ser rompidas em nenhuma operação.

- **Isolamento:** a regra de isolamento garante que transações que manipulem ou leiam os mesmos registros da base não sejam realizadas ao mesmo tempo. A prioridade é estabelecida pelo mecanismo de transação e da linguagem utilizada.

- **Durabilidade:** uma vez realizada uma transação na base, não há como desfazer as alterações, ou seja, o efeito de uma transação será permanente.

As regras do ACID só serão implementadas em todas as operações realizadas caso o mecanismo de transação garanta isto, sejam elas de alteração dos registros ou das estruturas da base de dados.

2.5 Linguagem de consulta

A comunicação entre o usuário e o sistema de gerenciamento é através de uma linguagem de consulta de uma base de dados. Esta linguagem permite pelo menos três comandos básicos para uso de uma base de dados (WADE, CHAMBERLIN 2012):

- **Inserção:** comando para inserir um registro na base. De acordo com o modelo utilizado, exige informações sobre registros pais, filhos e/ou chaves.

- **Seleção:** comando para seleção de registro, ou seja, consultar um ou mais registros para leitura ou processamento. Este comando seleciona uma parte da estrutura da base, de acordo com as cláusulas impostas, e reconstrói uma consulta como sendo uma base separada.

- **Remoção:** comando para remoção de registro na base. Este comando necessita que seja informado a chave do registro a ser removido, para que apenas um registro em específico seja removido.

Para realizar tais comandos mencionados, é necessário estruturar a segurança do sistema, exigindo acesso autorizado e permitindo que apenas usuários específicos e/ou autorizados sejam capazes de executar estes comandos.

Caso a base possa ser acessada por múltiplos usuários, o controle de acesso e permissão deve ser bem controlado pelo desenvolvedor, pois algumas ações podem ser realizadas indevidamente. Não é responsabilidade da linguagem de consulta controlar tais acessos ou ações.

2.6 SQL

A manipulação dos registros na base de dados ocorre pela linguagem de consulta, denominada SQL (*Structured Query Language*) (TAKAI, ITALIANO e FERREIRA, 2005). A linguagem SQL é um padrão de linguagem de consulta comercial que usa uma combinação de construtores em Álgebra e Cálculo Relacional e possui as seguintes partes:

- a) Linguagem de definição de dados (DDL);
- b) Linguagem interativa de manipulação de dados (DML);
- c) Incorporação DML (*Embedded SQL*);
- d) Definição de visões;
- e) Autorização;
- f) Integridade;
- g) Controle de transações.

A linguagem foi desenvolvida para manipular a base relacional projetada pela IBM, chamada *System R*, e comprovou a funcionalidade de um modelo relacional para base de dados (MATHEUS, 2010).

Figura 3: Exemplo de anatomia de um SQL

Fonte do autor.

Na Figura 3, exemplifica-se como a linguagem é subdividida em vários elementos (ISO/IEC 9075:2011). Como principais:

- **Cláusulas** (*Clauses*): são os componentes básicos que formam as declarações e consultas. Podem conter palavras restritas com instrução, expressões, predicados e informações necessárias ao processo.
- **Expressões** (*Expressions*): podem representar valores, escalares, booleanos, *strings* ou mesmo tabelas ou partes delas.
- **Predicados** (*Predicates*): são estruturas lógicas que serão avaliadas pelo SGBD. O SQL permite que resposta booleana seja verdadeiro, falso ou nulo (*True*, *False* ou *Null*).
- **Consultas** (*Queries*): são solicitação de informação ao SGBD.
- **Declarações** (*Statements*): são requisições para alterar estruturas da base de dados, controlar transações, conexões ou realizar diagnósticos.

Este padrão teve sua atualização em 2008 e outra em 2011. Em 2016, houve alteração em outros aspectos do SQL, como JSON e funções de tabelas polimórficas (ISO/IEC 9075-1:2016).

O padrão SQL define todo um conjunto estrutural de dados, mecanismo de transação e permite o uso da linguagem SQL para manipulação dos dados. Com este recurso o desenvolvedor pode facilmente consultar os dados inseridos, alterados e validar uma remoção. A maioria dos SGBD possuem as mesmas estruturas de anatomia do SQL, porém alguns possuem algumas particularidades que devem ser analisadas, como por exemplo, ser *case-sensitive*, sensível a caixa de letras, maiúsculas e minúsculas.

Capítulo 3: Bases de imagens mamográficas

3.1 Características de uma Base de imagens médicas

Os hospitais possuem duas formas de armazenar exames dos pacientes: a primeira é armazenada em papel por até 20 anos, a segunda forma é em sistemas *on-line* para arquivar permanentemente (BRASIL STORAGE, 2014). Atualmente, a grande maioria dos hospitais tem seus arquivos físicos de exames associados a sistemas computacionais para facilitar o processo de arquivamento e localização das informações. Este sistema inclui exames com imagens em filmes e laudos de diversos equipamentos, como raios X convencionais, mamógrafos, *CT*, *MRI*, entre outros. Alguns hospitais e clínicas até proveem as imagens digitalmente, seja por mídia física (em geral *CD*) ou *on-line*, com uma senha exclusiva do paciente e do médico.

Para o desenvolvimento de programas que armazenam bases de imagens e permitem consultá-las, é necessário que a base seja capaz de lidar com uma grande quantidade de informações e imagens, de modo seguro. Este tipo de sistema deve ser capaz de armazenar as imagens propriamente ditas, o exame realizado, o laudo resultante, dados da paciente (informações não confidenciais e com utilização autorizada), características dos equipamentos utilizados (equipamento de aquisição da imagem, digitalizador quando necessário, entre outros) e características específicas da imagem (resolução, contraste, tamanho, etc.).

Testes relatados na literatura por Benatti, Schiabel e Nunes, 2001; Schiabel, *et al.*, 2000; Schiabel, Escarpinati, Benatti, 2001, já comprovavam que o desempenho de esquemas CAD é grandemente influenciado pela quantidade e qualidade das imagens na base. Isto indica que uma base de dados precisa conter um conjunto de imagens consistente, confiável e amplo, de modo a não exercer influência sobre os resultados dos sistemas testados.

3.2 Requisitos de Bases de imagens mamográficas

Desde a década de 1990, tem-se buscado a elaboração de uma base única, padrão para desenvolvimento e testes de esquemas CAD mamográficos (NISHIKAWA, 1994; NISHIKAWA, 1998), pelo fato de não existir, ainda, nenhuma base de imagens mamográficas com capacidade para ser utilizada universalmente. Para servir a este fim, a base deveria apresentar um conjunto de características discutidas a seguir:

3.2.1 Conjuntos de imagens

A base de imagens mamográficas não é apenas composta pelas imagens propriamente, mas também são importantes os laudos radiológicos e, quando possível, acompanhamento em futuros exames e/ou confirmação por biópsia. Caso os laudos e a imagem não sejam confiáveis, os testes de CAD não serão adequados, perdendo a finalidade.

Nishikawa (NISHIKAWA, 1998) propõe que uma base de imagens mamográficas deve conter pelo menos 1000 imagens para poder ser utilizada de modo eficiente. Também é importante não apenas quantidade, mas grande variedade de imagens com diferentes características para que os sistemas possam ser testados em todos os aspectos.

De acordo com Nishikawa e Schiabel (Nishikawa, 1998; Schiabel, *et al.*, 2000) as imagens devem ser de alto contraste, devem apresentar resolução de contraste de pelo menos 10 bits (1024 níveis de cinza). Resoluções maiores do que 12 bits (4.096 níveis de cinza) não parecem causar divergências significativas no comportamento do CAD, porém, contrastes menores não apresentam informações suficientes para análises detalhadas (MATHEUS, 2010).

Para imagens digitais, é necessário garantir que a resolução espacial seja suficiente para que as imagens sejam consideradas de boa qualidade. Nunes (NUNES, SCHIABEL, *et al.*, 2001) mostrou que o uso de resoluções menores do que 0,100mm tendem a aumentar a quantidade de falsos-positivos sem aumento notável na sensibilidade. A resolução de imagens mamográficas digitais ou digitalizadas varia entre 0,035mm e 0,200mm (MATHEUS, 2010); a literatura recomenda entre 0,030mm e 0,060mm (NISHIKAWA, 1998).

Segundo Magalhães (2010), o tamanho e quantidade de pixels são fatores determinantes para a análise de detalhes, uma boa resolução requer pixels pequenos, que são obtidos ao se selecionar matrizes de grandes dimensões. Outro fator é o número de bits, quanto maior o número de bits por pixel, maior será a escala de tons de cinza, ou níveis de brilho da imagem.

O radiologista deve ser treinado, experiente e capaz de realizar todos os aspectos da imagem da mama, amostragem de tecidos e procedimentos de localização sob controle de imagem. Os requisitos de volume devem ter no mínimo de 1000 imagens por ano ou 5000 para aqueles que participam de programas de rastreamento (PERRY *et al.*, 2008). A literatura cita que por orientações europeias a resolução espacial seria > 12 lp/mm em um nível aceitável, para um nível desejável seria >15 lp/mm. Em relação a visibilidade do contraste de limite o nível aceitável e desejável é de $<1,5\%$ (PERRY *et al.*, 2008).

O *BancoWeb* lida com aproximadamente 6000 imagens no formato TIFF e DICOM, obtidas por digitalização com resoluções espaciais de 150 μ m ou 75 μ m e 12 bits de resolução de contraste a partir de filmes mamográficos. São convertidas apenas para exibição no *browser*

de internet, que não consegue interpretar adequadamente o formato TIFF. Contudo, cabe ao desenvolvedor avaliar as ferramentas disponíveis para tal formato; como as imagens não podem sofrer perdas por compactação, os formatos tendem a ser de extrema importância. É importante ser capaz de lidar com imagem de pelo menos dois bytes (9 a 16 bits) de contraste e exija pouco ou nenhum pré-processamento para a imagem ser utilizada.

3.2.2 Interface

Ao desenvolver uma base livre e gratuita em nível mundial, difere das demais bases existentes porque proporciona, aos desenvolvedores, recursos de esquemas de processamento de imagens, permitindo estabelecer comparações de seus desempenhos com um conjunto de dados comum. Uma interface deve ter segurança e usabilidade aos usuários que estão utilizando.

Muitas informações contidas nos exames e laudos, envolvem pacientes, são confidenciais e, por isso, a segurança da base se torna de grande importância. É necessário garantir que nenhum usuário externo possa ter acesso a tais informações.

O conjunto de imagens disponíveis na base deve ser de fácil consulta. Para isto, o programa precisa contemplar uma busca e recuperação de dados ampla e eficiente. O principal foco das buscas seriam os tipos de achados, como microcalcificação, linfonodos, nódulos ou lesões, e características específicas da imagem, como resolução espacial ou de contraste. A busca por imagem precisa ser clara e consistente, para que facilite o processo de teste e avaliação do programa a ser utilizado no seu processamento. Portanto, é importante que as informações sejam claras e de fácil entendimento.

3.3 Bases existentes

As bases mais conhecidas por desenvolvedores atualmente nessa área são as seguintes: MIAS, DDSM, LLNL/UCSF, CALMa, INBreast, BCDR e *BancoWeb*.

3.3.1 Base de dados MIAS

A MIAS (Mammographic Image Analysis Society) (SUCKLING et al., 1994) foi criada em 1994 e é a base mais antiga em uso e de acesso público. Esta base contém apenas 322 imagens de 8 bits de contraste (256 níveis de cinza) e 0,050 mm de resolução espacial. A MIAS consiste em 161 casos que geraram as 322 imagens digitalizadas a partir de filme.

As imagens são armazenadas e estruturadas em uma única pasta de dados com arquivo de texto explicando o laudo e posição dos achados em código. As informações são textuais, sem a existência de uma busca ou ferramenta para auxílio ao usuário.

Apesar de ser a mais antiga base de dados disponível, ainda é utilizada na literatura (MOREIRA, I. C. et al; 2012) por ser de uso livre (<http://peipa.essex.ac.uk/info/mias.html>). As informações existentes na base não são suficientes para alguns estudos, além da resolução das imagens digitalizadas ser reduzida, tornando-a inadequada para experimentos como a detecção de microcalcificações.

De acordo com a literatura recente, ainda há autores que utilizam a base MIAS para a sua pesquisa, como Souza (2018), que apresenta o diagnóstico de câncer de mama a partir de imagens de mamografia 2d utilizando descritores de forma 3d (obteve aproximadamente 97,4% de acerto).

3.3.2 DDSM

A DDSM (*Digital Database for Screening Mammography*) (Heath, Bowyer, *et al.*, 2001) é de acesso público para a comunidade de pesquisa e análise de imagens mamográficas. O projeto teve uma doação do Programa de Pesquisa de Câncer de Mama (*Breast Cancer Research Program*) do US Army Medical Research e Material Command.

O projeto DDSM é um esforço colaborativo envolvendo o Hospital Geral de Massachusetts (D. Kopans, R. Moore), Universidade do Sul da Flórida (K. Bowyer) e Sandia National Laboratories (P. Kegelmeyer), publicado em 2001. Outros casos da Escola de Medicina da Universidade de Washington foram fornecidos por Peter E. Shile, MD, Professor Assistente de Radiologia e Medicina Interna. Outras instituições colaboradoras incluem a Escola de Medicina da Universidade Wake Forest (Departamentos de Engenharia Médica e Radiologia), Sacred Heart Hospital e ISMD, Inc.

O principal objetivo da base de dados é facilitar a pesquisa sólida no desenvolvimento de algoritmos computacionais para auxiliar no rastreo. Os objetivos secundários da base de dados podem incluir o desenvolvimento de algoritmos para auxiliar no diagnóstico e no desenvolvimento de auxiliares de ensino ou treinamento. A base de dados contém mais de 10.000 imagens com laudos detalhados, contendo as coordenadas das bordas dos achados em um arquivo de texto associado a cada imagem. Cada estudo existente na base inclui duas imagens de cada mama, juntamente com algumas informações associadas à paciente (idade no

momento do estudo, classificação ACR de densidade de mama, classificação de sutileza para anormalidades, descrição de anomalia de ACR) e informações de imagem.

A DDSM é uma das mais utilizadas atualmente, estando acessível através de <http://marathon.csee.usf.edu/Mammography/Database.html>. Algumas pesquisas que utilizaram os dados dessa base e que podem ser mencionadas: Carvalho (2012), que teve como objetivo desenvolver uma metodologia de discriminação e classificação de regiões extraídas de mamografias em massa e não massa; Duarte, *et al.* (2013), que apresentou um sistema de auxílio ao diagnóstico (CAD) para classificação de lesões de mamografias, sendo as ROI's obtidas na base de dados DDSM; Neto (2016), que propôs uma metodologia computacional para auxiliar na descoberta de massas em mamas densas e não densas; Diniz (2017), que desenvolveu uma metodologia automática de detecção de regiões de massa em pares de mamografias digitalizadas adaptada à densidade da mama.

3.3.3 LLNL/UCSF

A LLNL/UCSF do Laboratório Nacionais de Lawrence Livermore (LLNL), juntamente com o Departamento de Radiologia da Universidade da Califórnia em San Fransisco (UCSF) (MAMMOGRAPHY IMAGE DATABASES, 2018) desenvolveram uma biblioteca de CD de 12 volumes de mamografias digitalizadas com microcalcificações. Esta base possui 198 filmes de 50 pacientes, sendo cada pixel de 12 bits de escala de cinza, com cerca de 50 megabytes de tamanho, para um total de 6 gigabytes para toda a base.

As imagens apresentam 5 casos normais, médios e saudáveis (mamografias normais anteriores e sem história de ultrassom, vistas de ampliação, biópsia, etc.), 5 casos normais, mas difíceis (com mamas densas ou fibrosas, implantes ou tecido assimétrico), 20 casos de microcalcificações obviamente benignas, 12 casos de microcalcificações suspeitas e benignas e 8 casos com um grupo maligno de microcalcificações, com biópsia comprovada (MAMMOGRAPHY IMAGE DATABASES, 2018).

Esta base não é de acesso público, e o conjunto é disponível pelo custo de US \$100 para cobrir os custos de reprodução.

3.3.4 MammoGrid (antiga CALMa)

A MammoGrid, anteriormente denominada como CALMa, foi elaborada em 2001 com um total de 3000 imagens (Amendolia, Bisogni, *et al.*, 2001). Os filmes foram digitalizados com uma resolução de 12 bits e salvos em arquivo no padrão DICOM. O MammoGrid é uma

colaboração entre o Reino Unido, a Itália, Suíça, Holanda e Noruega, com imagens padronizadas usando a mamografia padrão. Possui imagens digitalizadas de filmes e imagens FFDM, as anotações estão disponíveis nas estações de trabalho dos hospitais participantes (WARREN et al., 2007).

O início do projeto foi intitulado como CALMa (Computer-Assisted Library for Mammography) e financiado pelo INFN (Istituto Nazionale di Fisica Nucleare – Instituto Nacional de Física Nuclear, Itália) em 1997. Foi concluído em 2001 (FANTACCI et al., 2002; LAURIA et al., 2004); em 2002 a 2003 foi renomeado para GPCALMa e, depois, de 2004 a 2006 como MAGIC-5, sendo atualmente conhecido como MammoGrid e com um total de 3369 imagens (MOREIRA, I. C. et al., 2012).

A MammoGrid inicialmente não era uma base de imagens, mas sim um grid de processamento que associa vários bancos de dados de mamografia, permitindo que os médicos desenvolvam novas abordagens comuns e colaborativas para a análise de mamografias. Afim de compartilhar recursos entre vários centros médicos colaboradores. Tudo isso é fornecido por meio de uma nova infra-estrutura de informações de software e hardware que garante a integridade e a segurança dos dados médicos (ODEH, et al., 2004; SOLOMONIDES, 2018).

A principal limitação da MammoGrid é que está disponível apenas para instituições associadas, ou seja, é exclusiva do projeto e dos hospitais associados a ele e não há previsão de se tornar base pública.

3.3.5 INBreast

A INBreast foi desenvolvida no Breast Centre do CHSJ (Centro Hospitalar São João), no Porto, Portugal, com permissão do hospital e do Comitê de Ética e Comitê Nacional de Proteção de Dados (MOREIRA, I. C. et al., 2012). As imagens foram adquiridas entre abril de 2008 e julho de 2010, o equipamento utilizado foi MammoNovation Siemens (um mamógrafo do tipo FFDM), com tamanho de pixel de 70 μm e resolução de contraste de 12 bits. A matriz da imagem é 3328 x 4084 ou 2560 x 3328 pixels, dependendo da placa de compressão utilizada na aquisição (de acordo com o tamanho da mama da paciente).

As imagens utilizam o padrão DICOM (Digital Imaging and Communications in Medicine) que permite visualizar imagens radiológicas com nitidez, incorporação desses resultados nos computadores e armazenamento em formatos apropriados para figuras (MOREIRA, I. C. et al., 2012). Todas as informações médicas confidenciais foram removidas do arquivo, para manter a integridade das informações da paciente. Existem 410 imagens,

algumas são incidências médio-lateral-oblíqua (MLO), outras crânio-caudal (CC) e outras ainda de mulheres que fizeram mastectomia. A base de dados inclui exemplos de mamografias normais, com massas, com calcificações, com distorções arquitetônicas, assimetrias e imagens com múltiplos achados. O objetivo é fornecer imagens para pesquisadores desenvolverem, testarem e compararem seus métodos.

A INBreast é uma base gratuita e não tem um *site* estruturado onde as imagens ficam disponíveis. Para acessá-la é necessário entrar em contato com o Breast Research Group do Porto, Portugal (<http://medicalresearch.inescporto.pt/breastresearch/>) e pedir acesso. O acesso é através de um link para baixar conjuntos das imagens compactadas em RAR e um usuário e senha para poder fazer o login ao *site*.

3.3.6 BCDR

O BCDR (*Breast Cancer Digital Repository* – Repositório Digital de Câncer de Mama) foi resultado de uma colaboração entre CETA-CIEMAT, Instituto INEGI do Porto, Faculdade de Medicina da Universidade do Porto e a Universidade de Aveiro. Publicada em 2013, é a base disponível mais recente (AUGUSTO, 2014). Contém dois repositórios de domínio público. O primeiro contém mamografias digitalizadas de filmes, num total de 1010 pacientes (998 mulheres e 12 homens), com idades entre 20 e 90 anos de idade. O segundo repositório é composto por mamografias digitais, com 237 casos.

Todos os conjuntos de dados contêm dados clínicos relativos a paciente, descritivos e imagens extraídas em segmentação de cada lesão, além de classificação binária entre benigna e maligna. O objetivo desta base é estabelecer uma nova referência para desenvolver métodos de detecção/diagnóstico auxiliados por computador para câncer de mama e para treinamento de médicos.

3.3.7 BancoWeb

O *BancoWeb* é uma base de imagens desenvolvida pelo grupo LAPIMO, com acesso gratuito e *on-line*. Atualmente, conta com mais de 1900 imagens, sendo salvas no formato .tiff e DICOM. A base disponibiliza recursos para inclusão de usuário, paciente, exame, scanner, aparelho de raio x, hospital, BI-RADS, tipo de achado, patologia, equipamento e uma ferramenta de busca de imagens. Além de permitir efetuar o download de um exame completo, com imagens e laudo. Também, possui um recurso para recorte de imagem.

As imagens da base estão distribuídas em 30% com achado e 70% sem. O conjunto de imagens contém dados clínicos relativos aos pacientes, não confidenciais, para auxiliar desenvolvedores de processamento de imagens em seus testes. A base conta com 780 usuários cadastrados até o momento.

3.4 Conclusão

De todas as bases de imagens mamográficas descritas acima, apenas duas delas são mais utilizadas por desenvolvedores, que são MIAS e DDSM. Porém, não oferecem suporte ao usuário e seus respectivos projetos foram encerrados.

O *BancoWeb* é a única base descrita que preenche os requisitos de uma base de imagens médicas, em relação ao conjunto de imagens e interface, conforme citado no item 3.2.

Capítulo 4: Metodologia

4.1 Histórico do projeto

O grupo de pesquisas do Laboratório de Análise e Processamento de Imagens Médicas e Odontológicas (LAPIMO) tem trabalhado quase que de forma contínua, desde 1997, na elaboração e manutenção de uma base de dados e imagens mamográficas para funcionar como apoio a testes de técnicas de processamento de imagens na área, com foco principal no desenvolvimento de um esquema CADx (SCHIABEL, 2014)

A composição de tal base de dados constituiu-se de imagens provenientes de diferentes hospitais: Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto (HCFMRP-USP), Santa Casa de Misericórdia de São Carlos, Hospital São Paulo (EPM/UNIFESP), Hospital Pérola Byington (São Paulo) e Hospital das Clínicas da Faculdade de Medicina da UNESP (Botucatu-SP) (NUNES, 2001). Todas as imagens foram digitalizadas de filmes e salvas em pasta em um diretório no computador do LAPIMO.

A primeira publicação referente foi em 2001 (BENATTI, SCHIABEL e NUNES, 2001), descrevendo o desenvolvimento de uma base relacional (local) para uso dos pesquisadores do LAPIMO, buscando facilitar o arquivamento de um conjunto de mais de 6.000 imagens adquiridas.

Posteriormente, em 2003, foi desenvolvido um sistema de gerenciamento da base de imagens utilizando-se a linguagem visual *Delphi* com três tabelas (paciente, exame e local). Havia a possibilidade de escolher o hospital de origem e faixa etária da paciente; além disso, exibia as informações do hospital, laudo e a imagem. A base era armazenada em CDs, para que pudesse se tornar o mais “portátil” possível e para que não prejudicasse o desempenho da máquina que estava sendo utilizada para visualização das imagens, já que os exames ocupavam um grande espaço computacional (BENATTI, 2003).

Em continuidade, foi avaliada a possibilidade de ampliar a base local para um sistema *on-line*. As tabelas fundamentais e as relações entre elas foram 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 inserção de imagens e informações remotamente.

Entre o fim de 2003 e 2004, a base foi aberta ao público e testada (NUNES, SCHIABEL, *et al.*, 2003; NUNES, SCHIABEL, *et al.*, 2004) com usuários para se obter um *feedback* da

estrutura e funcionamento. Com isso, houve reformulação extensiva para melhorar alguns pontos (SCHIABEL, ESCAPINARTI e FREITAS, 2006). Mais uma vez, opiniões de usuários foram colhidas para a continuidade do projeto (FREITAS, SCHIABEL, *et al.*, 2007; MATHEUS e SCHIABEL, 2008).

Enfim, em 2010, (MATHEUS, 2010) apresentou-se o sistema de gerenciamento da base de imagens mamográficas no formato do modelo atual, com uma estrutura mais robusta e com as funcionalidades desejadas e configurada de tal modo que ficou disponível para acesso livre na Internet, recebendo a denominação de *BancoWeb*, e baseada na página *Web* do LAPIMO (em <http://lapimo.sel.eesc.usp.br/bancoweb/>)

A principal finalidade foi disponibilizar os casos para desenvolvimento, testes e avaliação comparativa de esquemas computadorizados de auxílio ao diagnóstico. A base *on-line* contém uma parcela (1900 imagens) de todas as imagens adquiridas e digitalizadas ao longo de 20 anos pelos pesquisadores do LAPIMO, com grande variedade de laudos, também disponíveis na base, assim como informações de alguns dados clínicos (não confidenciais) das pacientes. Uma detalhada interface 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 (MATHEUS, 2010).

4.1.1 Aquisição das imagens

Ao longo de anos, foi realizado o processo de aquisição das imagens da atual base. Alguns dados relativos ao controle de qualidade dos equipamentos mamográficos que originaram essas imagens também estão armazenados quando disponíveis.

Na Tabela 1 (MATHEUS, 2010) apresenta-se a lista de mamógrafos e seus respectivos hospitais de origem, conforme determinado durante o processo de aquisição. O HCFMRP-USP, ao longo do período, trocou seus aparelhos. Portanto, apresenta um histórico de diferentes mamógrafos.

Tabela 1: Lista de mamógrafos e local de origem

Mamógrafo	Local
Philips Mammodiagnost ¹	HCFMRP-USP
CGR Senographe 500 T ²	HCFMRP-USP
GE Senographe DMR ³	HCFMRP-USP
CGR Senographe 600 T	Hosp. S. Paulo
Lorad M III	Sta Casa S. Carlos

Fonte: Matheus (2010).

O digitalizador (*scanner*) utilizado também influencia na qualidade da imagem resultante. O digitalizador irá definir a resolução espacial e de contraste da imagem digital final. As imagens originalmente inseridas na base foram digitalizadas usando um de dois *scanners* a laser da marca Lumisys (Lumiscan 50 e Lumiscan 75) (MATHEUS, 2010), com resolução de contraste de 12 bits (4.096 níveis de cinza). Em relação à resolução espacial, o Lumiscan 50 (GÓES, SCHIABEL e ESCARPINATI, 2006) é capaz de digitalizar imagens com *pixels* de 0,150mm e o Lumiscan 75 (PIRES, MEDEIROS, ELIAS, 2008), com *pixels* de 0,085mm, quando operando com filmes típicos de 18 x 24 cm.

No sistema de armazenamento foi dada preferência ao formato TIFF (ADOBE Developers Association, 1992), como recomendava trabalho prévio desenvolvido na equipe (ESCARPINARTI, 2002). O formato TIFF apresenta diversas vantagens, pode ser adaptado a variados usos através de um cabeçalho (*header*) e uma estrutura chamada IFD (*Image file directory*) que contém detalhes da imagem, como tamanho, resolução de contraste, resolução espacial, entre outras informações. Estas características são determinadas como etiquetas (*Tag Image File Format*). Grande quantidade de ferramentas e bibliotecas para leitura e processamento de imagens estão disponíveis para esse formato.

Ainda assim, a base possui o cadastro de imagens que está estruturado para armazenar arquivos em qualquer outro tipo de formato (lembrando que, para exibição num *browser*, é necessária uma conversão do formato TIFF para JPG).

¹ Equipamento utilizado rotineiramente até o fim dos anos 80

² Equipamento utilizado rotineiramente até 1996

³ Equipamento utilizado rotineiramente durante o processo de aquisição das imagens naquele local.

4.1.2. Estruturação da base

A base foi estruturada com as informações pertinentes aos laudos, experiências com desenvolvimento de CAD do grupo e em contato com radiologistas e ginecologistas. Com base nisso, foram decididas quais informações são importantes manter e compor o sistema.

Baseando-se naquele modelo elaborado inicialmente para a versão física desta base (BENATTI, 2003), foi escolhido o modelo relacional para estruturá-la, ampliando para acomodar novas ferramentas e capacidades.

A Figura 4 apresenta o modelo da base de imagens antes das implementações realizadas neste trabalho; os retângulos escuros são as tabelas com informações da base, ou seja, os atributos da tabela que são os campos pertinentes a ela, e as setas indicam a ordem de relação, ou seja, quais tabelas estão relacionadas com quais.

Figura 4: Modelo relacional da base - versão original

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)	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)	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)
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)	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)	ladinew.patologia codigo : int(3) patologia : varchar(100) patologia_eng : varchar(100)
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)	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	ladinew.scanner scanner_codigo : int(8) scanner_nome : varchar(30) scanner_ftm : longtext scanner_winner : longtext
ladinew.tipoachado codigo : int(3) tipo_de_achado : varchar(100) tipo_de_achado_eng : varchar(100)	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	ladinew.birads categoria : float descricao : longtext
		ladinew.extensao codigo : int(7) imagem_nome : varchar(100)

Fonte: Matheus (2010).

Todas as tabelas da base possuem *index* que, no contexto da estrutura de dados, é uma referência associada a uma chave, que é utilizada para fins de otimização, permitindo uma localização mais rápida de um registro quando efetuada uma consulta (PEREIRA, 2006). Algumas tabelas que se relacionam mais de uma vez possuem mais de um *index*, facilitando a busca e resultando em uma pesquisa ainda mais rápida.

A base fica hospedada em um servidor do LAPIMO e todas as imagens se encontram neste servidor; existem *backups* da base completa, incluindo as imagens inseridas. Além disso, a cada alteração realizada é feito um *backup* da versão anterior a da implementação.

4.1.3. Ferramentas utilizadas no desenvolvimento

O *BancoWeb* foi desenvolvido em PHP 5.1 e Java Script, HTML como linguagem de marcação para interação com o usuário e o MySQL 4.0 como base de dados. A aplicação e as imagens são armazenadas em um servidor Apache sob um sistema operacional Debian LINUX, podendo ser acessadas através de qualquer plataforma e navegador de internet (MATHEUS, 2010).

O PHP é uma linguagem de *script open source* de uso geral, especialmente adequada para desenvolvimento *web* embutida com HTML, e é uma linguagem fácil de ser implementada, pois existem diversos conteúdo na internet e *sites* criados por ela. Pode ser utilizada na maioria dos sistemas operacionais, incluindo Linux. Possui suporte à maioria dos servidores *web* atualmente. Isso inclui o Apache, o IIS (internet information service) e muitos outros (COWBURN, P. et al., 2018).

O HTML *Hyper Text Markup Language* (Linguagem para Marcação de Hiper-texto, em português) é uma linguagem de marcação para descrever documentos *web* (páginas *web*), ou seja, um conjunto de *tags* de marcação. O propósito de um navegador (Chrome, Microsoft Edge, Internet Explorer, Firefox, Safari) é ler um documento HTML e exibi-lo (W3SCHOOL HOME, 2016). Portanto, o PHP é uma linguagem de programação usada para servidor, e o HTML é a linguagem que exibe o conteúdo buscado do servidor.

O MySQL é a base de dados de código aberto mais conhecida no mundo. Segundo DB Engines (SOLID IT, 2018), foi criado em 1995, atualmente pertence à ORACLE. Com alto desempenho, confiabilidade e facilidade de uso, o MySQL tornou-se a principal opção de base de dados para aplicativos baseados na *web*, usado por propriedades da *web* de alto perfil, incluindo *Facebook*, *Twitter*, *YouTube* e todos os cinco *sites* principais (ORACLE, 2018).

A preferência pelas ferramentas apresentadas acima decorre de sua facilidade de implementação, documentação e suporte existente atualmente, além de serem ferramentas *open source*, que permitem a sua utilização sem custo com licenças, sendo também de grande utilização por diversos desenvolvedores.

4.1.4. Implementações do *BancoWeb*

A base é denominada **BASE DE IMAGENS MAMOGRÁFICAS LAPIMO EESC/USP**, ou simplesmente, ***BancoWeb***, e sua versão *on-line* está disponível no endereço (<http://lapimo.sel.eesc.usp.br/bancoweb>). Ao acessar o *link*, o visitante será apresentado a uma

página inicial, onde poderá selecionar o idioma que deseja (português ou inglês), inserir seu usuário e senha ou se cadastrar para uso.

Após o estabelecimento da versão original das funcionalidades da base e seu consequente registro⁴ e publicação em 2010 (MATHEUS e SHIABEL, 2011), verificou-se que era necessário e/ou útil, incorporar certos aspectos e recursos relativos ao seu uso no sistema de gerenciamento, assim como a necessidade de uma ampliação e diversificação dos conjuntos de imagens/casos armazenados, particularmente por causa da mudança drástica que passou a ocorrer no que se refere ao processo de aquisição das imagens mamográficas na prática do radiodiagnóstico. Afinal, em termos de Brasil, desde o início desta década, praticamente todos os procedimentos relativos aos exames radiográficos, incluída a mamografia, migraram do registro da imagem em filme para o formato de arquivo digital.

Em quase todos os centros de radiodiagnóstico, o antigo e clássico procedimento de se revelar o filme após a exposição ao feixe de raios X no equipamento foi substituído ou por metodologia utilizando sistemas CR⁵ (mais práticos para centros de menor poder aquisitivo, porque o mamógrafo convencional pode ser mantido e apenas a aquisição da imagem se dá pelo digitalizador das IPs) ou por mamógrafos digitais de campo inteiro⁶ (FFDM ou sistemas DR) (KUZMIAK, et al., 2005). Consequentemente, foi necessário estabelecer um novo módulo de armazenamento de imagens e dados para contemplar esse novo conjunto de imagens mamográficas digitais, a fim de permitir que a principal finalidade dessa base pudesse continuar a mesma – auxílio a testes de técnicas de processamento de imagens – como também manter a sua atualidade perante o novo modelo de registro da mamografia. Além disso, uma proposta adicional, uma vez estabelecido um novo protótipo de esquema CAD pela equipe do LAPIMO, passou a ser a disponibilização desse importante recurso associado ao conjunto de imagens armazenadas, a fim de ampliar as possibilidades de testes comparativos nessa linha de pesquisas.

Por essa razão, o presente trabalho foi gestado e desenvolvido, tendo como objetivos principais implementar novas ferramentas e melhorias na base para torná-la, ao mesmo tempo, mais ampla e mais completa, desenvolver novo módulo de gerenciamento de imagens

---

⁴ Reg. INPI nro. 09939-3

⁵ CR: *Computed Radiography*. Equipamentos que proporcionam a obtenção da imagem a partir da varredura com laser de uma placa de material fosfórico (IP - *imaging plate*) previamente sensibilizada durante a exposição aos raios X. A imagem gerada pela varredura é armazenada em formato digital.

⁶ Nesses mamógrafos, uma placa eletrônica formada por uma matriz de células correspondentes a sensores de radiação (ou de luz), é a responsável pelo registro da imagem, armazenada diretamente em formato digital.

mamográficas digitais provenientes de equipamentos CR e DR – paralelamente ao armazenamento desse novo conjunto de imagens provenientes de hospitais colaboradores – e associar a todo esse sistema um módulo de um esquema CADx – voltado ao processamento de classificação de achados – e seus recursos. Todas as implementações realizadas foram feitas na versão da aplicação em português e inglês.

As implementações realizadas têm como foco criar interação com o usuário e torná-la mais ampla. Essa é a razão pela qual não houve a migração da estruturada da base para o modelo orientado a objetos; optou-se por manter o modelo relacional, que já apresenta todos os recursos necessários para o funcionamento adequado do sistema de gerenciamento da base.

Abaixo serão apresentadas as modificações e atualizações realizadas, os novos recursos disponíveis, quais processos do programa são afetados e qual o motivo do desenvolvimento.

4.2. Novas Implementações do *BancoWeb*

As implementações realizadas na base, que vão desde inclusão de informações, como também incorporação de novas ferramentas, serão apresentadas nos itens abaixo.

4.2.1 Alterações quanto à origem da imagem cadastrada

Na primeira versão disponibilizada do *BancoWeb*, existiam os dados do hospital em que foi obtido o exame. Na nova versão, esta informação ficará oculta para segurança e confidencialidade da paciente. Esta implementação foi retirada das telas de pesquisa de imagem, busca geral, resultado da pesquisa e informações da paciente.

Figura 5: Painel de busca, informação do hospital oculta (informação era exibida entre o campo de sexo e idade; também na tabela de dados do aparelho de Raios-X).

Busca

Busca textual

Digite sua busca:

Busca por recortes

Digite o motivo do recorte:

Dados das 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 ▾
 Fonte da Imagem: Qualquer ▾

Informações do Achado

Localização: Qualquer ▾
 Tipo de Achado: Qualquer ▾
 Patologias Detectadas: Qualquer ▾

Dados dos pacientes

Sexo: Qualquer ▾
 Idade Telarca: Entre e Anos
 Tempo Lactação: Entre e Anos
 Câncer na Família? Indiferente ☐ Sim ☐ Não ☐
 Uso anticoncepcional? Indiferente ☐ Sim ☐ Não ☐
 Reposição Hormonal? Indiferente ☐ Sim ☐ Não ☐
 Nódulo Palpável? Indiferente ☐ Sim ☐ Não ☐

Dados dos aparelhos de Raios-X

Escolha	Marca	FTM WIENER		Paralelo (micrometro)	Perpend. (micrometro)
☒	Senographe 500r	-	-	300	300
☒	Senographe 600r	1	1	300	300
☒	GE Sengoraphe DMR	-	-	300	300
☒	deletar	---	---	0	0
☒	Senographe DS	-	-	300	300

Dados dos aparelhos de Scanner

Escolha	Nome	FTM	ESPECTRO DE WIENER
☒	Lumiscan 50 (Lumisys)	-	-
☒	Lumiscan 75 (Lumisys)	-	-

MENU
PESQUISAR

Fonte do autor.

Ao efetuar o *login*, informando o usuário e senha, existe uma opção no menu “Pesquisa de Imagens – Busca geral”. Ao abrir a tela, conforme Figura 5, exibe-se um conjunto de quadros para que possam ser preenchidas informações para realizar uma busca mais específica no sistema. Nesta tela, existia uma opção de Origem, que trata do Hospital, indicadas pela seta, e que ficará oculta do sistema.

Além da tela de busca, ao pesquisar, exibe-se o resultado da busca, permitindo visualizar “Mais informações” que são relacionadas ao exame em questão, conforme Figuras 6 e 7.

Figura 6: Resultado da busca, informação do hospital oculta.

Total de Imagem(s) encontrada(s) : 1900 imagem(s)		Total de Imagem(s) cadastrada(s) no banco : 1900 imagem(s)	
Proximo -> 1/475		NOVA PESQUISA MENU	
	Posição: Crânio-caudal	Data do Exame: 15/03/2001	Sexo: feminino
	Mama: direita	Bits de Contraste: 12	Ano de Nascimento: 1958
	Tamanho do Pixel: 0.075	BI-RADS: 3	Idade Telarca: nao consta
	Tempo de lactação: nao consta	Achado: sim	Uso de Anticoncepcional: nao consta
	Reposição Hormonal: nao consta	Câncer na Família: nao consta	Raios-X: Senographe 500t
Scanner: Lumiscan 75 (Lumisys)		Nódulo Palpável: nao consta	

Fonte do autor.

Figura 7: Mais informação, informação do hospital oculta.

Dados do Paciente	
Data de Nascimento:	01/01/1958
Sexo:	feminino
Telarca:	nao consta
Lactação:	nao consta
Câncer na Família:	nao consta
Uso de Anti-Concepcional:	nao consta
Reposição Hormonal:	nao consta
Nódulo Palpável:	nao consta

Fonte do autor.

4.2.2 Cadastro de Tipo de Achado – Inclusão da característica adicional

Quando inserimos um novo exame com os dados da paciente e a imagem, podemos informar as características pertinentes a ela, como posição, bits de contraste, tipo de digitalizador, tipo de achado, entre outras informações. Também pode-se indicar se há alguma suspeita de anormalidade na mama e, em caso positivo, é possível descrever o tipo de achado, como: nódulo, calcificações, densidades, linfonodos, entre outros.

A base apresentava vários Tipos de Achado repetidos, pois trata-se do mesmo tipo, mas com alguma característica adicional, entretanto, incluídos no cadastro de Tipo de Achado. Assim, foram unificados os cadastros e acrescentou-se a tabela de “característica adicional” atrelada aos cadastros, para informar as eventuais particularidades. Na tela de pesquisa por imagens, pode-se buscar pelo Tipo de Achado e o resultado da busca exibe o Tipo de Achado e a característica adicional para o usuário visualizar.

- 1) As Figuras 8 e 9 ilustram a nova estruturação do cadastro do Tipo de Achado, que possibilita acrescentar separadamente alguma característica adicional e, se disponível, informações quanto ao quadrante de localização do achado (superior externo, superior interno, inferior externo, inferior interno, aureolar ou axilar),

quadrante (a partir do canto superior esquerdo ou inferior direito da imagem), tipo de achado e patologia .

Figura 8: Cadastro do Tipo de Achado, inclusão da característica adicional

Fonte do autor.

Figura 9: Informações do Tipo de Achado no cadastro de exame.

Fonte do autor.

No acesso à tela de busca de imagens, pode-se expandir informações de um determinado exame, de modo a visualizar as características adicionais, conforme Figura 10, enquanto a Figura 11 ilustra o diagrama de tabelas internas do sistema.

Figura 10: Resultado da busca, com a informação da característica adicional.

Total de Imagem(s) encontrada(s) : 1900 imagem(s)		Total de Imagem(s) cadastrada(s) no banco : 1900 imagem(s)	
Proximo >> 1475		NOVA PESQUISA MENU	
		Sexo: feminino Ano de Nascimento: 1958 Idade Telarca: nao consta Raio-X: Senographie 500t	
Posição: Crânio-caudal Mama: direita Tamanho do Pixel: 0.075 Tempo de lactação: nao consta Reposição Hormonal: nao consta Scanner: Lumiscam 75 (Lumisy)		Data do Exame: 15/03/2001 Bits de Contraste: 12 BI-RADS: 3 Achado: sim Câncer na Família: nao consta Nódulo Palpável: nao consta	
Localização:	Tipo:	Característica Adicional:	Patologia:
	Nódulo	Arredondado;Calcido;Bem definido; Densificacoes discreta com varios pequenos nodulos;Suspeito espiculado;Altamente suspeito;Aspecto Benigno; Esparsos bem definidos;Suspeito; Tipo densidade de partes moles no quadrante inferior; Tipo densidade de partes moles no quadrante superior; com 0,7 cm de diametro, arredondado, com contornos bem definidos;Presenca de 2 nodulos com 0,5 cm de diametro;Dispersos;Imagens calcificas;Imagens nodulares;Imagens tubulares;Imagem arredondada radiopaca;nodular	Indefinida
	Nódulo Calcificações	Arredondado;Calcido;Bem definido; Densificacoes discreta com varios pequenos nodulos;Suspeito espiculado;Altamente suspeito;Aspecto Benigno; Esparsos bem definidos;Suspeito; Tipo densidade de partes moles no quadrante inferior; Tipo densidade de partes moles no quadrante superior; com 0,7 cm de diametro, arredondado, com contornos bem definidos;Presenca de 2 nodulos com 0,5 cm de diametro;Dispersos;Imagens calcificas;Imagens nodulares;Imagens tubulares;Imagem arredondada radiopaca;nodular Benignas;Isoladas;arredondadas;Lineares;Baixa densidade;Irregulares;Agrupadas;Distroficas;Ductos terminais;Esparsos;Irregulares agrupadas;Vasculares;Benignas;Grossas;Cutaneas;Puntiformes Benignas;Arredondadas;Puntiformes isoladas;Ductais;Lineares;Veruga;Casca de ovo;Focos de calcificacoes distrofica em lumen de ductos terminais;Grossas dispersas; Grossas agrupadas	Indefinida Indefinida

Fonte do autor.

Figura 11: Tabela de Tipo de Achado, inclusão da característica adicional

ladinew.tipoachado	
codigo	int(3)
tipo_de_achado	varchar(100)
tipo_de_achado_eng	varchar(100)
tipo_achado_caract	varchar(500)
tipo_achado_caract_eng	varchar(500)

Fonte do autor.

4.2.3 Inclusão da fonte do exame DR, CR ou DG

Entre as quase duas mil imagens cadastradas no *BancoWeb*, já há várias que foram provenientes de sistemas mamográficos digitais de campo inteiro (FFDM), a que passamos a denominar aqui de imagens do tipo DR (de *Digital Radiography*), além das originalmente digitalizadas de filmes. Para fins de pesquisa e estudos, alguns pesquisadores necessitam desta informação, que anteriormente não era disponibilizada na base.

No cadastro de exame, então, criou-se a possibilidade da inserção de imagens na base, considerando como dado-chave a sua fonte de aquisição, ou seja, a indicação se trata de uma imagem digital gerada por sistema do tipo CR, DR ou se trata de filme digitalizado. Obviamente, para imagens DR não é pertinente indicar o *Scanner*, e o campo ficará desabilitado. Na tela de Pesquisa de Imagens, também é possível filtrar a busca pelo campo de Fonte. Atualmente, existem na base cerca de 200 imagens digitais provenientes de sistemas DR e 1700 digitalizadas.

No cadastro de exame, existem várias informações referentes ao exame, laudo e paciente, e agora, a informação da fonte da imagem, conforme Figura 12. Na busca por imagens, houve a inclusão da informação da fonte da imagem para realizar uma busca mais específica, facilitando ao usuário no momento de encontrar as imagens desejadas, conforme Figura 13. A tabela de exame, sofreu uma alteração para inclusão de um novo atributo denominado “exame_fonte_imagem” para atender esta implementação, conforme Figura 14.

Figura 12: Cadastro de exame, inclusão da fonte da imagem

Fonte do autor.

Figura 13: Busca de imagens, informação da fonte da imagem

Escolha	Marca	FTM WIENER	Paralelo (micrometro)	Perpend. (micrometro)
☒	Senographe 500t	-	-	300
☒	Senographe 600t	1	1	300
☒	GE Senographe DMR	-	-	300
☒	deletar	---	---	0
☒	Senographe DS	-	-	300

Escolha	Nome	FTM	ESPECTRO DE WIENER
☒	Lumiscan 50 (Lumisys)	-	-
☒	Lumiscan 75 (Lumisys)	-	-

Fonte do autor.

Figura 14: Tabela de Exame, inclusão da fonte da imagem

Field Name	Data Type
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)
usuario	tinyint(1)
exame_fonte_imagem	varchar(5)

Fonte do autor.

4.2.4 Recorte – Recurso para obter o negativo na imagem

O *BancoWeb* possui um recurso para recorte de imagem que permite selecionar uma área da mesma e, após recortar, pode-se salvar a região selecionada na própria base de dados ou efetuar o *download*. Em alguns casos, pesquisadores e médicos que fazem a análise da imagem a fim de identificar áreas suspeitas, preferem analisar a imagem em negativo. Com isso, houve a necessidade de criar um botão na tela de recorte que possibilite aplicar o recurso de gerar o negativo da imagem, conforme ilustrado na Figura 15. Na tela de ajuda/instruções do recorte, informamos ao usuário a possibilidade do novo recurso, conforme Figura 16.

Figura 15: Recorte da imagem, (a) botão para inverter imagem; (b) imagem negativa.

Fonte do autor.

Figura 16: Recorte, recurso de ajuda

Fonte do autor.

4.2.5 Ferramenta para classificar imagem, integração com esquema CADx

Segundo Schiabel et al. (2011), a partir de pesquisas na literatura, foram encontradas algumas bases de imagens com características que não atendem adequadamente às exigências para testes de desenvolvimento de esquemas de processamento de imagens. Por isso, o *BancoWeb*, nesta nova versão, traz um recurso inédito que é a integração com um programa CADx.

Assim, ao se buscar uma determinada imagem, existe a opção de “Recorte”, conforme já mencionado, que permite selecionar parte dela e baixar apenas essa região, em vez de baixar o arquivo da imagem inteira. Além disso, é possível “Classificar” uma determinada imagem a partir desse recorte: ao selecionar uma ou duas determinadas áreas suspeitas, o sistema automaticamente envia o recorte ao esquema CADx associado, que analisa a região e retorna quando se trata de achado “Suspeito” ou “Normal”. Esta ferramenta permite enviar até duas áreas da imagem por vez para essa análise, conforme Figura 17. Se a imagem escolhida é, por exemplo, a projeção crânio-caudal esquerda, será exibida ao lado imagem da mesma mama, porém, na projeção médio-lateral esquerda. Desta forma, o usuário terá a possibilidade de selecionar uma área em cada imagem dos dois lados da mesma mama – o que permitirá aumentar o grau de eficácia da avaliação do sistema de processamento.

Figura 17: Ferramenta para classificar imagem com CADx

Fonte do autor.

Se for selecionada uma área por engano, existe a possibilidade de desmarcar a região através do ícone 'X', apresentado junto ao quadro vermelho da Figura 17; ainda assim, o usuário também poderá classificar a região utilizando apenas um recorte de uma única imagem se preferir.

Figura 18: Resultado da classificação

Fonte do autor.

Estes sistemas têm o intuito de analisar as informações fornecidas para apresentar, com base nos dados da imagem processada, uma informação diagnóstica, conforme Figura 18. Na mamografia, as marcações das regiões de interesse podem indicar ao pesquisador informações de um possível achado, além de servir como auxílio na prática médica, por exemplo, em um ambiente de ensino.

4.2.6 Reconstrução da Busca por imagens

Quando se realiza uma busca, são listadas todas as imagens encontradas com suas respectivas informações, como por exemplo: posição, se é da mama direita ou esquerda, tamanho do pixel, tipo de sistema digitalizador, número de bits de contraste, entre outras. Também há um totalizador da quantidade de imagens encontradas e quantidade de imagens na base, que atualmente está em 1900.

O resultado da busca exibia anteriormente as informações em forma de colunas na tela, o que em dados momentos dificultava ao usuário a visualização, pois para cada imagem era necessário efetuar *scroll* (rolagem) da página *web* para visualizar todo o seu conteúdo. Portanto, está implementada uma nova estruturação da apresentação do resultado da busca por imagens, mantendo as mesmas informações, por outro lado, exibidas em formato horizontal (de linha). Os dados que possuem uma quantidade maior de informações como patologia, característica adicional do tipo de achado, entre outros, são listados quando o usuário clica no botão “expandir”, facilitando assim a visualização imediata da imagem ao abrir a tela, conforme Figura 19.

Figura 19: Reconstrução do resultado da busca

Total de Imagem(s) encontrada(s) : 1900 imagem(s)		Total de Imagem(s) cadastrada(s) no banco : 1900 imagem(s)	
Proximo -> 1/475		NOVA PESQUISA MENU	
	Posição: Crânio-caudal Mama: direita Tamanho do Pixel: 0.075 Tempo de lactação: nao consta Reposição Hormonal: nao consta Scanner: Lumiscan 75 (Lumisys)	Bits de Contraste: 12 BI-RADS: 3 Achado: sim Cancêr na Família: nao consta Nódulo Palpável: nao consta	Sexo: feminino Ano de Nascimento: 01/01/1958 Idade Telarca: nao consta Uso de Anticoncepcional: nao consta Raio-X: Senographie 500t
	Posição: Crânio-caudal Mama: esquerda Tamanho do Pixel: 0.075 Tempo de lactação: nao consta Reposição Hormonal: nao consta Scanner: Lumiscan 75 (Lumisys)	Bits de Contraste: 12 BI-RADS: 3 Achado: sim Cancêr na Família: nao consta Nódulo Palpável: nao consta	Sexo: feminino Ano de Nascimento: 01/01/1958 Idade Telarca: nao consta Uso de Anticoncepcional: nao consta Raio-X: Senographie 500t
	Posição: Crânio-caudal Mama: direita Tamanho do Pixel: 0.075 Tempo de lactação: nao consta Reposição Hormonal: nao consta Scanner: Lumiscan 75 (Lumisys)	Bits de Contraste: 12 BI-RADS: 2 Achado: sim Cancêr na Família: nao consta Nódulo Palpável: nao consta	Sexo: feminino Ano de Nascimento: 01/01/1958 Idade Telarca: nao consta Uso de Anticoncepcional: nao consta Raio-X: Senographie 500t
	Posição: Médio-lateral Mama: esquerda Tamanho do Pixel: 0.075 Tempo de lactação: nao consta Reposição Hormonal: nao consta Scanner: Lumiscan 75 (Lumisys)	Bits de Contraste: 12 BI-RADS: 3 Achado: sim Cancêr na Família: nao consta Nódulo Palpável: nao consta	Sexo: feminino Ano de Nascimento: 01/01/1958 Idade Telarca: nao consta Uso de Anticoncepcional: nao consta Raio-X: Senographie 500t

Fonte do autor.

Os detalhes da imagem podem ser visualizados ao clicar na seta, que irá expandir e abrir mais informações pertinentes àquela imagem. Este recurso permite expandir e retrair as informações, facilitando a visualização da miniatura da imagem e seus dados correspondentes, como ilustrado na Figura 20.

Figura 20: Resultado da busca, detalhes expandidos

Total de Imagem(s) encontrada(s) : 1900 imagem(s)		Total de Imagem(s) cadastrada(s) no banco : 1900 imagem(s)	
Proximo -> 1/475		NOVA PESQUISA MENU	
	Posição: Crânio-caudal Mama: direita Tamanho do Pixel: 0.075 Tempo de lactação: nao consta Reposição Hormonal: nao consta Scanner: Lumiscan 75 (Lumisys)	Bits de Contraste: 12 BI-RADS: 3 Achado: sim Cancêr na Família: nao consta Nódulo Palpável: nao consta	Sexo: feminino Ano de Nascimento: 01/01/1958 Idade Telarca: nao consta Uso de Anticoncepcional: nao consta Raio-X: Senographie 500t
	Posição: Crânio-caudal Mama: esquerda Tamanho do Pixel: 0.075 Tempo de lactação: nao consta Reposição Hormonal: nao consta Scanner: Lumiscan 75 (Lumisys)	Bits de Contraste: 12 BI-RADS: 3 Achado: sim Cancêr na Família: nao consta Nódulo Palpável: nao consta	Sexo: feminino Ano de Nascimento: 01/01/1958 Idade Telarca: nao consta Uso de Anticoncepcional: nao consta Raio-X: Senographie 500t
Localização:	Tipo:	Característica Adicional:	Patologia:
	Calcificacoes	Benignas;Isoladas;arredondadas;Lineares;Baixa densidade;Irregulares;Agrupadas;Distroicas;Ductos terminais;Esparsas;Irregulares agrupadas;Vasculares;Benignas;Grosseiras;Cutaneas;Puntiformes Benignas;Arredondada;Puntiformes isoladas;Ductais;Lineares;Veruga;Casca de ovo;Focos de calcificacoes distrofica em lumen de ductos terminais;Grosseiras dispersas; Grosseiras agrupadas	Indefinida

Fonte do autor.

4.2.7 Download completo do exame

A base de imagens disponibiliza o recurso de *download* completo de um determinado exame. O *download* contém as imagens do exame e todos os seus dados, inclusive dados da paciente (não confidenciais). Deve-se ressaltar que, na primeira versão da base, este recurso apesar de estar implementado, permitia, no entanto, o *download* apenas das imagens.

Na tela de resultado da busca, há um botão para visualizar mais informações, o qual exhibe dados da paciente, todos os exames vinculados e as demais informações pertinentes ao exame. A fim de possibilitar aos pesquisadores o uso das informações para estudos, ao se clicar no botão de “Baixar Todos”, será feito o *download* de um arquivo .RAR contendo todas as imagens do exame e um arquivo .TXT com as informações da paciente e dos exames, como ilustra a Figura 21.

Figura 21: Recurso (em destaque) para baixar exame e modelo do arquivo .TXT com os dados respectivos

Total de Imagem(s) encontrada(s) : 5 imagem(s)

Dados do Paciente	
Data de Nascimento:	01/01/1958
Sexo:	feminino
Telarca:	nao consta
Lactação:	nao consta
Cancêr na Família:	nao consta
Uso de Anti-Concepcional:	nao consta
Reposição Hormonal:	nao consta
Nódulo Palpável:	nao consta

[Proximo ->](#) 1/2 BAIXAR TODOS

Exame I	Exame II	Exame III	Exame IV
Ampliar Recortes Baixar (.tiff)	Ampliar Recortes Baixar (.tiff)	Ampliar Recortes Baixar (.tiff)	Ampliar Recortes Baixar (.tiff)
Achado: sim Scanner: Lumiscan 75 (Lumisys) Posição: Crânio-caudal Mama: direita Tamanho do Pixel: 0.075 Bits de contraste: 12 BI-RADS: 2	Achado: sim Scanner: Lumiscan 75 (Lumisys) Localização: I) Não Informado II) Não Informado Mama: direita Tamanho do Pixel: 0.075 Bits de contraste: 12 BI-RADS: 3	Achado: sim Scanner: Lumiscan 75 (Lumisys) Posição: Crânio-caudal Mama: esquerda Tamanho do Pixel: 0.075 Bits de contraste: 12 BI-RADS: 3	Achado: sim Scanner: Lumiscan 75 (Lumisys) Localização: I) Não Informado II) Não Informado Mama: direita Tamanho do Pixel: 0.075 Bits de contraste: 12 BI-RADS: 3
Localização: I) Não Informado Tipo: I) Não Informado Característica Adicional: I) Não Informado Patologia: I)Indefinida	Tipo: I) Não Informado II) Não Informado Característica Adicional: I) Não Informado II) Não Informado Patologia: I)Indefinida II)Indefinida	Localização: I) Não Informado Tipo: I) Não Informado Característica Adicional: I) Não Informado Patologia: I)Indefinida	Tipo: I) Não Informado II) Não Informado Característica Adicional: I) Não Informado II) Não Informado Patologia: I)Indefinida II)Indefinida

Fonte do autor.

```

Dados do Paciente
Data de Nascimento: 01/01/1958
Sexo: feminino
Telarca: nao consta
Lactação: nao consta
Cancêr na Família: nao consta
Uso de Anti-Concepcional: nao consta
Reposição Hormonal: nao consta
Nódulo Palpável: nao consta

Exame I
Nome da Imagem: 1150301CCD1.tiff
Posição: Crânio-caudal
Mama: direita
Tamanho do Pixel: 0.075
Bits de contraste: 12
BI-RADS: 2
Achado: sim
Scanner: Lumiscan 75 (Lumisys)
Localização: I) Não Informado
Tipo: I) Não Informado
Característica Adicional: I) Não Informado
Patologia: I)Indefinida

Exame II
Nome da Imagem: 1150301CCD2.tiff
Posição: Crânio-caudal
Mama: direita
Tamanho do Pixel: 0.075
Bits de contraste: 12
BI-RADS: 3
Achado: sim
Scanner: Lumiscan 75 (Lumisys)
Localização: I) Não Informado II) Não Informado

```

Fonte do autor.

4.2.8 Implementação do sistema de gerenciamento de arquivos de imagens de *Phantom*

Um dos recursos inéditos implementados na base foi o de gerenciamento para arquivamento e manipulação (e eventual *download* se for de interesse) de imagens de um simulador mamográfico (*phantom*) desenvolvido recentemente no grupo (SOUSA, 2017) – especificamente para avaliação ou validação de sistemas de auxílio ao diagnóstico (CAdE/CAdx). Este *phantom* constitui-se de oito camadas compostas por filme de PVC submerso em parafina gel em uma distribuição não uniforme, e mais três camadas compostas somente de parafina gel. Tal distribuição permite simular regiões mais ou menos densas, de acordo com a concentração do material, resultando nas quatro categorias de classificação de densidades BI-RADS®.

Os nódulos foram simulados usando dois modelos tridimensionais impressos, um para lesões circunscritas e outro para espiculadas. Para as microcalcificações, foi utilizada a hidroxiapatita granulada distribuída em quatro *clusters* que representavam casos comumente encontrados em mamas reais (SOUSA, 2015).

Cada lesão é inserida sobre a imagem do *phantom* utilizando uma ferramenta computacional que realiza a soma da imagem radiográfica da lesão à imagem da mama simulada. A intensidade e localização de cada lesão pode variar de acordo com a escolha do usuário. A principal característica desse *phantom* é sua versatilidade em representar quantidade muito grande de diferentes projeções radiográficas, como a simular um número também muito grande de diferentes imagens mamográficas. Essa característica possibilita, no limite, gerar quantidade estatisticamente significativa de diferentes imagens com ou sem sinais representativos de estruturas suspeitas a serem convenientemente armazenadas para utilização em testes de desenvolvimento e, principalmente, desempenho de esquemas CAdE/CAdx. (SOUSA, 2017)

Com vistas, então, à criação de um módulo da base exclusivo para o conjunto dessas imagens de *phantom*, foram criados menus para cadastro de equipamento, achado de *phantom*, cadastro de imagens de *phantom*, pesquisa por imagens de *phantom*, além das devidas adaptações nos cadastros de tipo de achado, e perfil de imagens. As novas tabelas criadas para armazenar estas informações podem ser vistas na Figura 22. Além destas tabelas, houve a inclusão de um campo no cadastro de tipo de achado para distinguir os tipos de achado de imagens reais da base das imagens provenientes de *phantom*.

Figura 22: Diagrama novas tabelas para arquivo e gerenciamento das imagens de *phantom*

Fonte do autor.

A sequência de procedimentos que agora constituem esse novo módulo de gerenciamento para as imagens de *phantom* é descrita abaixo:

- No menu da base, incluíram-se opções de menu para cadastro de achado, equipamento, imagens de *phantom* e pesquisa de Imagens de *phantom*, conforme Figura 23. Ao efetuar o *login* no sistema, o usuário será redirecionado a esta tela e, dependendo do seu perfil, serão habilitadas as opções de menu a que o mesmo pode ter acesso.

Figura 23: Menu com opções adicionais para cadastro e acesso a imagens de *phantom*

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.

Fonte do autor.

- b) O novo cadastro de achado de *phantom*, conforme Figura 24, permite informar a descrição, tipo de achado e distribuição. O objetivo é indicar se para a imagem em questão, o achado encontrado, é nódulo ou *cluster* de microcalcificações, bem como suas características. O cadastro de tipo de achado será vinculado ao achado através do campo de tipo de achado, indicando se trata de simulação de massa circunscrita, espiculada, entre outros. Nesse caso, deverá exibir apenas os tipos de achado que sejam de *phantom*. A distribuição são os quadrantes da imagem para cada exposição ‘X’ representada de uma lesão.

Figura 24: Cadastro de Achado de *Phantom*

Fonte do autor.

- c) O novo cadastro de equipamento, conforme Figura 25, permite informar a descrição (como por exemplo Lorad MIV, Lorad Dimensions 2D ou GE Senographe DS), a fonte da imagem (CR ou DR) e o sistema (exemplo: Agfa 85). Estas informações caracterizam o tipo de equipamento utilizado para a aquisição das imagens de *phantom*.

Figura 25: Cadastro de equipamento.

Fonte do autor.

- d) No cadastro de tipo de achado, existem os registros que são referentes às imagens já cadastradas no sistema. Para distinguir os registros existentes de imagens reais para os registros de imagens de *phantom*, foi criado um campo “Pertence a” que permite selecionar se o tipo de achado em questão está presente em uma imagem real ou numa de *phantom*, conforme Figura 26. No cadastro de achado de *phantom*, deve-se informar qual o tipo de achado, já que só serão visíveis os que pertencem a imagens de *phantom*.

Figura 26: Cadastro de tipo de achado, inclusão do campo "pertence a".

Fonte do autor.

- e) O cadastro de hospital permaneceu com as mesmas informações que já existiam no sistema, conforme Figura 27. Apenas incluiu-se novos hospitais que foram vinculados às imagens de *phantom* cadastradas na base. O hospital é informado no momento de cadastrar uma nova imagem.

Figura 27: Cadastro de hospital

Cadastro do Hospital

Informações sobre Hospital

Nome:

Endereço:

Complemento:

Bairro:

CEP: ex: (14620000)

Telefone: -

Fax: -

Contato:

Radiologista Responsável:

Fonte do autor.

- f) O cadastro de imagem de *phantom* possibilita informar todas as suas características, como, equipamento utilizado, hospital, alvo/filtro, pré-processamento, escolher a imagem, gabarito da imagem, configuração do *phantom*, kVp, mAs, compressão (cm), densidade e intensidade da lesão (SOUSA, 2017), conforme Figura 28.

Figura 28: Cadastro de imagens de *phantom*

Cadastro de Imagens de Phantom

Informações de Equipamento

Equipamento:

Informações de Hospital

Hospital:

Informações da Imagens de Phantom

Alvo/Filtro:

Pré-processamento:

Arquivo:

Gabarito:

Configuração do Phantom:

KVp:

mAs:

Compressão (cm):

Densidade %:

Intensidade da lesão:

Fonte do autor.

Segundo Sousa (2017), os testes realizados para geração das imagens de *phantom* associam as seguintes informações:

- i) Equipamento mamográfico utilizado;
- ii) Hospital em que o equipamento estava disponível para realização dos testes;
- iii) Alvo/Filtro de tubo (Molibdênio, Ródio ou Tungstênio);
- iv) Pré-processamento, (que informa se houve aplicação de realce, suavização, aumento de altas intensidades e remoção do fundo da imagem, por exemplo);
- v) Gabarito – imagem que contém a posição exata das lesões inseridas para servir de referência;
- vi) Configuração do *phantom* – combinação das camadas utilizadas na sua confecção (SOUSA, 2017);
- vii) kVp e mAs utilizados na exposição;
- viii) Compressão (cm) – indicando o tamanho efetivo do conjunto de camadas associadas para a configuração em questão daquela imagem após aplicação do compressor para a exposição;
- ix) Densidade – correspondente ao valor percentual da simulação de tecido mais e menos fibroso, calculado a partir da imagem digital gerada (SOUSA, 2017);
- x) Intensidade da lesão inserida – durante os testes foram inseridas, computacionalmente, lesões simuladas, cujo contraste em relação ao fundo da mama simulada pode variar conforme a imagem.

Após realizar o cadastro da imagem, o usuário será redirecionado ao cadastro de “achados” daquela imagem, onde serão vinculadas as estruturas de interesse da imagem que está sendo cadastrada, conforme Figura 29. É permitido vincular mais de um achado a uma mesma imagem.

Figura 29: Cadastro de achado, vínculo com a imagem

Fonte do autor.

- g) A busca por imagens permite pesquisar por qualquer conjunto de informações. O recurso de busca por imagens de *phantom* apresenta 4 blocos que são divididos em: dados do equipamento, dados do hospital, informações de achado de *phantom* e

dados da imagem (alvo/filtro, pré-processamento, nome da imagem, configuração do *phantom*, KVp, mAs, compressão, densidade e intensidade da lesão), conforme Figura 30.

Figura 30: Busca por imagens de *phantom*

Fonte do autor.

- h) Após efetuar a busca por imagens, será possível visualizar uma lista de resultados com as imagens encontradas na pesquisa realizada dentro da base. Todos os dados cadastrados (conf. Itens f e g) pertinentes às imagens são exibidos (Figura 31).

Figura 31: Resultado da busca de imagens de *phantom*

Total de Imagem(s) encontrada(s) : 611 imagem(s)		Total de Imagem(s) cadastrada(s) no banco : 611 imagem(s)	
Proximo -> 1/153		NOVA PESQUISA MENU	
 GABARITO	Equipamento: Lorad MIV Alvo/Filtro: Mo Pré-processamento: Mo Hospital: CIDI Santa Casa de São Carlos	Configuração Phantom: 3 parafinas + camada A (BRA) KVp: 28 Compressão (cm): 4,5	Densidade (%): 17,45 Intensidade da Lesão: Não informado mAs: 80
 GABARITO	Equipamento: Lorad MIV Alvo/Filtro: Mo Pré-processamento: Mo Hospital: CIDI Santa Casa de São Carlos	Configuração Phantom: 2 parafinas + camada B + camada G (BRB) KVp: 28 Compressão (cm): 5,0	Densidade (%): 32,37 Intensidade da Lesão: Não informado mAs: 80

Fonte do autor.

Assim como na busca por imagens reais, as imagens de *phantom* possuem informações como o achado, que podem ser vinculadas a vários registros de uma única imagem. Para isto, no resultado da busca, ao expandir uma determinada imagem, visualizam-se estas informações, conforme Figura 32.

Figura 32: Resultado da busca, informações expandidas

Fonte do autor.

- i) Ao selecionar-se o botão de gabarito, deverá abrir uma tela contendo a imagem de gabarito da imagem de *phantom* relacionada, caso exista, conforme Figura 33 (se não, será exibida uma mensagem ao usuário).

Figura 33: Gabarito da imagem

Fonte do autor.

- j) Selecionando-se “Mais informação”, será exibida a imagem de *phantom* selecionada, gabarito (caso exista), equipamento, fonte da imagem, sistema, hospital, achado, distribuição, tipo de achado, característica adicional, alvo/filtro,

pré-processamento, configuração, kVp, mAs, compressão (cm), densidade (%) e intensidade da lesão, como ilustra a Figura 34.

Figura 34: Mais informações da imagem de *phantom*

Mais informações sobre a imagem de *Phantom*

Imagem de *Phantom*

Gabarito de *Phantom*

[BAIXAR TODOS](#)

Dados da Imagem	
Equipamento:	Lorad MIV
Fonte Imagem:	CR
Sistema:	Agfa 85
Hospital:	CIDI Santa Casa de São Carlos
Achado <i>Phantom</i>:	Cluster - Distribuição 1
Distribuição:	Distribuição 1
Tipo de Achado:	Circunscrita
Característica Adicional:	Circunscrita
Alvo/Filtro:	Mo
Pré-processamento:	Mo
Configuração do <i>Phantom</i>:	BRA + distribuição 1 dos clusters entre a parafina e a camada A
KVp:	28
mAs:	80
Compressão (cm):	4,5
Densidade (%):	18,81
Intensidade da Lesão:	Não informado

Fonte do autor.

- k) Em “Mais informações”, existe também o recurso para baixar todos os casos, que possibilita efetuar o *download* das imagens, gabarito e todas as informações em .TXT, compactadas em um arquivo .RAR, como mostra Figura 35.

Figura 35: Arquivo .TXT disponibilizado ao usuário

```

Imagem de Phantom

Dados da Imagem

Nome da Imagem: 42410119
Gabarito: 42410119.png
Arquivo da Imagem: 42410119_1522112394.png
Equipamento: Lorad MIV
Fonte Imagem: CR
Sistema: Agfa 85
Hospital: CIDI Santa Casa de Sao Carlos
Achado Phantom: Nódulo
Distribuição: 1
Tipo de Achado: Circunscrita
Característica Adicional:
Alvo: Mo
Filtro: Mo
Configuração do Phantom: 2 parafinas + camada B + camada G (BR2)
KVp: 28
mAs: 80
Compressão (cm): 5,0
Densidade (%): 32,37
Intensidade da Lesão: Não informado

```

Fonte do autor.

- 1) O menu de Perfil das imagens exibe opções para visualizar gráficos estatísticos de imagens sem achado, imagens com achado e imagens sem achados x imagens com achados. Foi criada uma opção para visualizar os tipos de achados que pertencem às imagens de *phantom*, conforme Figura 36.

Figura 36: Perfil de tipo de achado para imagens de *phantom*

Fonte do autor.

- m) No menu, também existe uma opção “Sobre as imagens de *Phantom*”, que exibe uma tela com informações que podem auxiliar os pesquisadores a entender melhor as características do *phantom*, conforme Figura 37.

Figura 37: Sobre as imagens de *phantom*.

Fonte do autor.

Dado o exposto, todas as implementações realizadas tornaram a base mais ampla. Desde os recursos relacionadas com as imagens reais da base como o novo módulo de imagens de *phantom* poderão proporcionar uma variedade de casos de teste durante o desenvolvimento e pesquisa de esquemas relacionados à área.

Capítulo 5: Resultados

A base se encontra reestruturada com todos os itens descritos no capítulo anterior. Foram realizados testes e está funcional em todos os navegadores testados (Internet Explorer 11 ou superior, Microsoft Edge 41 ou superior, Mozilla Firefox 59 ou superior, Google Chrome 65 ou superior e Safari), sem a necessidade de instalação de nenhum *software* ou *plug-in* extra.

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

5.1 Acesso, manutenção e atualização

O *BancoWeb* é uma base pública que exige cadastro dos seus usuários, autorizado pelo administrador do sistema. As demais bases não necessitam de cadastro para acesso.

As bases MIAS (SUCKLING et al., 1994), DDSM (Heath, Bowyer, et al., 2001), INBreast (MOREIRA, I. C. et al., 2012) e BCDR (AUGUSTO, 2014) são públicas, uma paga (UCSF) (MAMMOGRAPHY IMAGE DATABASES, 2018) e uma restrita a pesquisadores e hospitais associados (CALMa) (Amendolia, Bisogni, et al., 2001). A base UCSF cobra US\$ 100, oferecendo em forma física 12 CDs (este valor é para “custo de reprodução”).

A base UCSF informa não pretender atualizar seu sistema, já o projeto CALMa (Amendolia, Bisogni, et al., 2001), apesar de ter mudado de nome para MAGIC-5, parece continuar ativo. Já os projetos INBreast (MOREIRA, I. C. et al., 2012) e BCDR (AUGUSTO, 2014) são novos, e estão ativos e podem ter atualizações.

Das bases públicas, apenas o *BancoWeb*, até o momento, continua sendo um projeto aberto, que pretende manter-se em desenvolvimento pelo menos pelos próximos anos. Os projetos MIAS (SUCKLING et al., 1994) e DDSM (Heath, Bowyer, et al., 2001) disponibilizam as suas imagens, mas não mais realizarão atualizações e não se responsabilizam por problemas técnicos que essas bases possam apresentar.

5.2 Quantidade de imagens

No momento, o *BancoWeb* contém 1.703 imagens provenientes de filmes digitalizados a laser (Lumiscan 50 e Lumiscan 75, Lumisys, Inc.) com 12 bits de resolução de contraste (entre 3.500 e 3.800 níveis de cinza) e entre 0,085mm e 0,150mm de resolução espacial. Contém ainda 196 imagens provenientes de mamógrafos do tipo DR. Quando foi feito o registro do sistema

de gerenciamento da base em 2010, a versão *on-line* disponibilizava cerca de 300 imagens a menos, o que mostra o constante cadastramento de novos casos, num processo ainda dinâmico.

Em 2014, foi publicado um comparativo de bases desse tipo existentes (AUGUSTO, 2014), que conduziu a uma tabela resumo reproduzida aqui na Tabela 2.

Tabela 2: Comparativo de bases mamográficas disponíveis no mundo

Databases	MIAS[38]	Bancoweb[45]	DDSM(IRMA)[41]	LLNL/UCSF[49]	inBREAST[43]	BCDR-D[44]	BCDR-F[44]
Year	1994	2010	1999 Unknown	2010	2013	2013	
Origin	UK	Brazil	USA	USA	Portugal	Portugal	Portugal
Number of Cases	161	320	2620	50	115	243	288
Number of Images	322	1400	10480	198	410	392	388
Views	MLO	CC MLO	CC MLO	CC MLO	CC MLO	CC MLO	CC MLO
Contain Mastectomy	No	Unknown	No	Yes	Yes	No	No
Mammogram Type	Screen Film	Screen Film	Screen Film	Screen Film	FFDM	FFDM	Screen Film
Contrast Resolution	8	12	12 and 16	12	8	8	Unknown
Image Format	PGM	TIFF	PNG	DICOM	DICOM	TIF	TIF
Patient data	No	Yes	Age	Yes	Yes	Yes	Yes
Abnormally Categorization	Yes	Yes	Yes	Calc. Only	Yes	Yes	Yes
Clinical data	No	Yes	Age	Yes	Yes	Yes	Yes
Abnormally Annotation	No	Yes	BI-RADS	Yes	BI-RADS	Yes	Yes
Classification type	Binary	BI-RADS	BI-RADS	BI-RADS	BI-RADS	Binary	Binary
Abnormally Contour	Radius	Few cases	Yes	Yes	Yes	Yes	Yes
Image Descriptors	No	No	No	No	No	Yes	Yes

Fonte: Augusto (2014).

Com base neste comparativo, podemos ressaltar que o *BancoWeb* evoluiu em vários aspectos com a nova versão. A maioria das bases de imagens citadas são apenas repositórios das imagens, algumas oferecem ferramentas para auxílio ao pesquisador, porém, as funcionalidades não são completas, vários dados necessários aos usuários não são apresentados. A tabela 3 apresenta alguns comparativos entre as bases citadas, incluindo ano, origem, quantidade de imagens, acesso, resolução, sistema de busca, classificação de CADx, imagens de *phantom* e se oferece manutenção/atualização.

Tabela 3: Comparativos (2018).

	MIAS	DDSM	UCSF / LLNL	MammoGrid (CALMa)	INBreast	BCDR	<i>BancoWeb</i>
Ano	1994	1999	2010	2001	2013	2013	2010
Origem	UK	US	USA	IT	Portugal	Portugal	BR
Quantidade de imagens	322	10480	198	3369	410	1247	Reais 1900 <i>Phantom</i> 611
Acesso aos arquivos	Livre	Livre	Pago (US\$ 100)	Fechada	Livre	Livre	Livre após cadastro
Resolução de contraste (bits/pixel)	8	12 (80%) 16 (20%)	12	12	12	8	12
Sistema de busca	Não	Sim, mas não funcional	*	*	*	*	Sim
Classificação com CADx	Não	Não	Não	Não	Não	Não	Sim
Imagens de <i>Phantom</i>	Não	Não	Não	Não	Não	Não	Sim
Manutenção e atualização	Não	Não	Sim	Sim	Sim	Sim	Sim

Fonte do autor.

A base DDSM supera todas as outras em termos de quantidade de imagens, a segunda maior é a do projeto CALMa. Considerando, porém, todas as imagens na nossa base interna (isto é, incluindo as não disponíveis em rede) o *BancoWeb* tem por volta de 7.400 imagens digitais, pouco mais que a metade das imagens disponíveis na DDSM. Na base que está disponível *online* existem cadastradas 1900 imagens reais e 611 imagens de *phantom*. No entanto, um módulo inédito foi incorporado neste trabalho – associado a uma configuração exclusiva e adaptada do seu sistema de gerenciamento – que corresponde à base de imagens de *phantom* mamográfico (SOUSA, 2017). Dada a alta versatilidade de configurações que aquele *phantom* permite (e, conseqüentemente, a quantidade de imagens diferentes, com ou sem sinais

de interesse, que isso possibilita gerar), a estruturação organizada exclusiva para essas imagens e suas informações, em separado do conjunto das imagens reais (novas ou antigas) é não só importante para os principais objetivos da nossa base em si – apoio para desenvolvimento e testes de técnicas e esquemas computacionais de processamento de imagens em mamografia – como representa uma ferramenta inédita nessa área.

Em termos de acesso, a maioria das bases de imagens existentes são de acesso livre; entretanto, não são todas que apresentam uma interface amigável, e algumas proporcionam apenas o *download* das imagens.

A base MIAS era a única que apresentava imagens com menos de 10 bits de resolução de contraste. Atualmente, existe a base INBreast e BCDR com imagens de 12 bits.

5.3 Sistemas de busca

No momento, o *BancoWeb* é ainda a única base que apresenta um sistema de buscas amplo e funcional. Tal sistema é eficiente, exhibe informações da paciente, da imagem, ferramentas de recorte, possibilita classificação por esquema CADx associado à base, e permite *download* completo do exame, entre outras funcionalidades. Além disso, oferece variadas opções para realização de busca, propicia ao usuário a busca por uma imagem ou por característica específica.

A reestruturação do sistema de busca do *BancoWeb*, possibilitou uma interação mais clara e ampla, apresentando as imagens e suas informações na forma horizontal; os detalhes podem ser visualizados com um recurso de expandir e retrain, além de apresentar o resultado em 1 segundo (considerando a configuração do servidor atual).

Embora a base DDSM afirme que fornece um sistema de busca, ele não funciona atualmente, o projeto está concluído e os autores não mais respondem a questionamentos técnicos da base. Apesar de ser a base mais utilizada na literatura, a sua última atualização remota foi em 2003. A falta de um bom sistema de busca torna difícil a procura por características específicas, considerando a quantidade de imagens disponíveis, pois é a maior base de imagens atualmente. As demais bases comparadas não apresentam sistema de busca, contudo, pela quantidade de imagens que possuem, é fácil procurar diretamente a característica desejada.

5.4 Classificação com CADx

De acordo com a literatura, existem vários autores (e mesmo empresas) que desenvolvem esquemas de CADx para auxílio ao diagnóstico, e para os testes destes esquemas, são utilizadas as imagens de bases mamográficas como as comparadas neste trabalho. Entretanto, nenhuma base apresenta acoplado a ela um classificador CADx – o que era uma proposta de trabalho futuro no projeto original do sistema (MATHEUS, 2010), com objetivo de ampliar as capacidades de busca por conteúdo, analisando a imagem em si e exibindo o resultado para análise do usuário. O projeto do atual protótipo de esquema CADx completo e funcional para uso por radiologista é mais recente (MATHEUS, 2015).

O presente trabalho também integrou, a partir do desenvolvimento e disseminação do protótipo de esquema CADx de nosso grupo (MATHEUS, 2015), sua respectiva biblioteca JAVA desenvolvida com nossa base de imagens. O classificador desse esquema permite selecionar uma ou até duas áreas das imagens para categorizar instantaneamente. O usuário pode, então, escolher a imagem que deseja na nossa própria base e obter a classificação de uma região de interesse selecionada na tela de resultado da busca. O resultado indica se a área selecionada é suspeita ou normal.

A ferramenta é de fácil manuseio e está totalmente funcional, garantindo o objetivo de em um único programa abranger várias funcionalidades essenciais aos pesquisadores e desenvolvedores da área. Com apenas uma imagem, o programa oferece variadas aplicações, o que é totalmente inédito, particularmente no comparativo entre as bases de imagens disponíveis consideradas nesse capítulo.

5.5 Imagens de *Phantom*

O objeto simulador (*phantom*) de mama que foi desenvolvido em nosso grupo (SOUSA, 2017) é capaz de gerar diversos padrões de imagens obtidos pela variação aleatória de tamanho, forma, contraste e distribuição de lesões. Estes objetos podem ser utilizados por exemplo, em testes de esquema CADx.

O sistema de gerenciamento para cadastro e busca dos conjuntos de imagens gerados pelo simulador e sua disponibilidade *on-line* também é inédito. As bases de imagens comparadas neste trabalho não apresentam recursos como estes. Como mencionado na descrição da literatura estudada, a maioria possui imagens mamográficas provenientes de filmes digitalizados.

As imagens de *phantom* atualmente armazenadas no *BancoWeb* correspondem a 611, sendo: 33% circunscrita, 33% espiculada, 27% de cluster de microcalcificações e 7% de microcalcificação.

5.6 Perfil estatístico

A base possui uma ferramenta para avaliação da distribuição estatística de imagens, que permite ao usuário uma melhor compreensão das imagens como um todo. Estas informações são obtidas em tempo real na base a cada pedido realizado. Conforme novas imagens sejam inseridas, os gráficos sempre estarão atualizados.

Com isso, podemos analisar a nova classificação dos tipos de achado, que foram apresentadas neste trabalho, e também, a classificação dos tipos de achado das imagens de *phantom*. A Figura 38 apresenta a nova distribuição do tipo de achado.

Figura 38: Estatística tipo de achado das imagens reais

Fonte do autor.

As imagens existentes na base foram reclassificadas em relação ao seu tipo de achado. Foram criados 17 tipos de achados e para cada um existem as características adicionais – por exemplo, para nódulo, pode ser bem definido, arredondado, dispersos, entre outros. Assim, conseguimos distinguir as imagens pelo tipo de achado, facilitando o desenvolvedor no momento da busca.

A Figura 39 apresenta a classificação dos tipos de achado de *phantom*. As imagens cadastradas foram classificadas em 4 tipo de achado de *phantom*, que foram aplicados durante o desenvolvimento e testes deste novo *phantom*; assim, o desenvolvedor poderá pesquisar por um determinado tipo e ter um resultado mais específico.

Figura 39: Estatística tipo de achado das imagens de *phantom*

Fonte do autor.

5.7 Outras ferramentas

O *BancoWeb* também apresenta outras ferramentas únicas, como informações sobre tipos de achado, informações da imagem, fonte da imagem (CR, DR e digitalizadas), recorte e obtenção do negativo da imagem.

Capítulo 6: Discussão e Conclusões

O grupo do LAPIMO tem trabalhado nessa linha de pesquisa (estruturação de base de dados) desde 1997, desenvolvendo uma base de imagens mamográficas como ferramenta de auxílio ao desenvolvimento de programas de processamento em esquemas CAD em mamografia (BENATTI, 2003; NUNES, SCHIABEL, *et al.*, 2004; NUNES, SCHIABEL e GOES, 2007). O sistema original de busca da base de dados/imagens mamográficas *BancoWeb* tornou-se disponível em 2010, sendo *on-line* e gratuito. A sua estrutura fundamental continua funcional até hoje.

Em anos de pesquisa na área, o LAPIMO acumulou uma grande quantidade de imagens de vários hospitais, digitalizadas de diferentes equipamentos. Esta biblioteca associada ao *BancoWeb* oferece o gerenciamento da base de dados com um conjunto amplo de imagens mamográficas de qualidade apropriada ao processamento digital e, principalmente, comparação entre diferentes esquemas CAD (MATHEUS, 2010).

A dificuldade de acesso aos arquivos, em geral confidenciais, de hospitais e a necessidade de equipamentos apropriados para digitalização das imagens mostra a necessidade de formar uma base pública ampla e de fácil acesso. A quantidade e variedade de imagens para testar de modo eficiente um esquema CAD é extremamente importante, pois nem sempre os desenvolvedores conseguem acesso fácil.

A primeira versão da base foi desenvolvida em DELPHI com interação com o usuário na base local (BENATTI, 2003), enquanto o padrão SQL foi empregado para a estruturação e administração da base. Nessa primeira versão, em 2003, a 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 contemplava um sistema de busca que permitia resgatar imagens pelo hospital de origem, idade da paciente e estrutura de interesse. Ao selecionar uma imagem, eram apresentadas as informações pertinentes ao caso, como laudo médico e equipamentos. Por ser um programa executável construído em DELPHI, ele funcionava apenas em plataformas Windows, além de dificultar a busca de imagens por estarem armazenadas em diversos CDs diferentes.

Com a base local em funcionamento, foi planejada a expansão dela para uma base *on-line* que permitisse a qualquer interessado o acesso às imagens. A nova versão foi desenvolvida em PHP e HTML para interação remota com o usuário, e MySQL como dialeto para a base de

dados. O programa e as imagens passaram a ser armazenados em um servidor Apache em uma plataforma LINUX.

As imagens oferecidas na base *on-line* foram selecionadas e inseridas de acordo com as exigências encontradas na literatura. Ela proporciona para *download* cerca de 1.900 imagens, provenientes de 3 hospitais e digitalizadas usando 2 *scanners* a laser diferentes. Por isso, têm 12 bits de resolução de contraste (3.500 a 3.800 níveis de cinza) e resolução espacial entre 0,150mm e 0,075mm.

Com os avanços em pesquisas durante todos estes anos, a necessidade e demanda de melhorias, mudanças e incrementos no sistema de gerenciamento e buscas foram surgindo. O trabalho desenvolvido teve como objetivo desenvolver estas demandas, proporcionando maior interação com o usuário. Todos os recursos estão na versão em português e inglês.

Uma delas, por exemplo, foi a carência de disponibilidade junto da base de um esquema CAD que permitisse ao usuário estabelecer comparações com outros esquemas de processamento, além de aplicar recursos de processamento de imagens nos arquivos pré-existent, recurso, aliás, inexistente em qualquer base de dados similar mundialmente disponível. Como um esquema desse tipo foi desenvolvido recentemente no grupo (MATHEUS, 2015), sua implementação junto à nossa base de imagens em questão se tornou uma consequência imediata para implantação e testes.

A base foi integrada com um aplicativo do esquema CADx que permite selecionar uma imagem do resultado da busca. Sendo ela de um lado específico e posição da mama, por exemplo, o sistema apresenta a imagem da mama direita sendo crânio-caudal e ao lado exibe a mama direita médio-lateral da mesma paciente; assim, é possível fazer um recorte de uma imagem e um recorte da outra e, ao classificar, o CADx avalia se a área é suspeita ou não e apresenta essa informação ao usuário.

Além disso, por questões de confidencialidade, informação sobre a origem das imagens foi ocultada do sistema, permanecendo apenas internamente na base. No cadastro das imagens, em “Tipo de Achado” pode-se indicar se há alguma suspeita de anormalidade na mama e, em caso positivo, é possível descrever o tipo de achado, como: nódulo, calcificações, densidades, linfonodos, entre outros. A base apresentava vários Tipos de Achado repetidos, pois, embora se tratasse do mesmo tipo (nódulo, por exemplo), mas com alguma característica adicional, era incluído nessa classe como um novo Tipo de Achado. Assim, foram unificados os cadastros apresentando a característica adicional como uma subclasse, facilitando, assim, a busca.

No cadastro de exame, então, há agora a possibilidade da inserção de imagens, considerando como dado-chave a sua fonte de aquisição, ou seja, a indicação se se trata de uma imagem digital gerada por sistema do tipo CR, DR ou se se trata de filme digitalizado.

Há ainda o recurso para recorte de imagem que permite selecionar uma área da mesma e, após recortar, salvar a região selecionada na própria base de dados ou efetuar o *download*. Em alguns casos, pesquisadores e médicos preferem adicionalmente analisar a imagem em seu negativo. Com isso, a inclusão desta ferramenta vai agora possibilitar a visualização da imagem nessas duas formas diferentes (original e negativo).

Segundo Schiabel et al. (2011), a partir de pesquisas na literatura, foram encontradas algumas bases de imagens com características que não atendem adequadamente às exigências para testes de desenvolvimento de esquemas de processamento de imagens. Por isso, o *BancoWeb*, nesta nova versão, traz um recurso inédito que é a integração com um programa CADx, para permitir aos usuários também classificar as mesmas a partir do esquema de processamento digital. Como foi habilitada a opção de “Recorte”, conforme já mencionado, selecionada uma ou duas determinadas áreas suspeitas, o sistema automaticamente envia o recorte ao esquema CADx associado, que analisa a região e retorna à informação de achado “Suspeito” ou “Normal”. Esta ferramenta permite enviar até duas áreas da imagem por vez para essa análise. Se a imagem escolhida é, por exemplo, a projeção crânio-caudal esquerda, será exibida ao lado imagem da mesma mama, porém, na projeção médio-lateral esquerda. Desta forma, o usuário terá a possibilidade de selecionar uma área em cada imagem das duas projeções da mesma mama – o que deve aumentar o grau de eficácia da avaliação do sistema de processamento.

A fim de melhorar a usabilidade e interação do usuário com a ferramenta de busca por imagens, foi implementada uma nova estruturação da apresentação do resultado dessa busca, mantendo as mesmas informações, por outro lado, exibidas em formato horizontal (de linha). Os dados que possuem uma quantidade maior de informações – como patologia, característica adicional do tipo de achado, entre outros – são listados quando o usuário seleciona o botão “expandir”, facilitando assim a visualização imediata da imagem ao abrir a tela. Desta forma, o usuário visualiza o resultado de forma mais clara e ampla, evitando sempre o uso do efeito de *scroll* (rolagem) que existia na versão original.

O *download* pode ser feito das imagens do exame e todos os seus dados, inclusive dados da paciente (não confidenciais). Deve-se ressaltar que, na primeira versão da base, este recurso permitia, no entanto, o *download* apenas das imagens. A fim de possibilitar aos pesquisadores o uso das informações para estudos, através do botão “Baixar Todos”, será possível o *download*

de um arquivo .RAR contendo todas as imagens e um arquivo .TXT com as informações da paciente e dos exames.

Nos últimos anos, como resultado de projeto de doutoramento concluído em 2017 (SOUSA, 2017) um novo conjunto extenso de imagens voltadas a testes de esquemas CAD foi incorporado a partir de exposições do novo *phantom* mamográfico elaborado (SOUSA, 2017). Estas imagens estão associadas ao *software* com busca e recuperação separada das imagens reais, ou seja, o usuário poderá ter as duas visões (imagens reais e simuladas) em uma mesma base.

Com vistas, então, à criação de um módulo da base exclusivo para o conjunto dessas imagens de *phantom*, foram criados menus para cadastro de equipamento, achado de *phantom*, cadastro de imagens de *phantom*, pesquisa por imagens de *phantom*, além das devidas adaptações nos cadastros de tipo de achado, e perfil de imagens. Para isto, foram criadas novas tabelas para armazenar os dados e também incluído um campo no cadastro de tipo de achado para distinguir os tipos de achado de imagens reais da base e das imagens provenientes de *phantom*.

O módulo de *phantom* conta com cadastros básicos para inserção das informações e também de uma busca completa das imagens e *download* completo com todos os detalhes pertinentes à imagem, incluindo o gabarito relativo às estruturas inseridas e demais características do objeto de teste exposto aos raios X. Este módulo, além de fornecer o armazenamento das imagens em uma base de dados, também fornece os recursos de interação com o usuário, que para pesquisadores e desenvolvedores é de grande utilidade, já que permite casos para comparação e testes na realização de seus projetos e pesquisas.

Desde a primeira versão do sistema gerenciador da base para interação com o usuário de forma local até essa nova versão, o foco tem sido não só fornecer uma ampla quantidade de imagens, mas também ferramentas únicas que facilitem a utilização dos arquivos armazenados. Por isso, a implementação do classificador do esquema CADx acoplado à base é um importante recurso, além de totalmente inédito considerando um comparativo com outros repositórios, de modo a oferecer um recurso suplementar ao diagnóstico médico.

Comparações do *BancoWeb* com outras bases mostraram que os requisitos citados pela literatura que são importantes para uma base de imagens médicas estão atendidos, mas com a vantagem de apresentar recursos que as demais bases não oferecem, até o momento, como a classificação com um esquema CADx. A busca de imagens também tem sido um grande foco na base, possibilitando interação com o usuário de forma clara e objetiva.

Em conclusão, com os avanços em pesquisas durante todos estes anos, a necessidade e demanda de melhorias, mudanças e incrementos no sistema de gerenciamento e buscas foram surgindo. Por isso, os objetivos deste trabalho foram em desenvolver e aperfeiçoar recursos na busca de imagens, tipos de achado, permitir a classificação pelo sistema CAD, além da implementação um novo sistema de gerenciamento e armazenamento de imagens de um novo *phantom*, desenvolvido como resultado de pesquisa ainda mais recente no grupo (SOUZA, MATHEUS, SCHIABEL, 2018).

Assim, embora o sistema já tivesse uma variedade de imagens com qualidade, oferecendo um conjunto básico de ferramentas computacionais para facilitar o estudo das imagens e localização das informações pertinentes a cada caso, este trabalho apresenta avanços que irão complementar e/ou aprimorar processos pré-existentes, e adicionar uma nova base de imagens (as de *phantom*) no sistema, além de oferecer novos recursos de gerenciamento e também processamento digital dessas imagens.

As implementações realizadas tiveram como objetivo oferecer aos usuários, pesquisadores, desenvolvedores e interessados da área novos recursos e ferramentas que contribuam na busca por imagens, a fim de realizar testes e comparativos. Finalmente, num comparativo com outras bases análogas que disponibilizam imagens mamográficas, o *BancoWeb* oferece uma ferramenta única e poderosa, com mais opções que as demais bases disponíveis atualmente.

Sugestões para pesquisas futuras

Além de todos os recursos desenvolvidos nesta versão do *BancoWeb*, outros passos futuros deste projeto incluiriam a estruturação da forma de aceitação de um novo usuário, e desenvolvimento de mecanismos de segurança tecnologicamente mais modernos.

E uma outra implementação futura seria permitir classificar as imagens do módulo de *phantom*, desenvolvido nesta versão, para classificação do esquema CADx.

Referências Bibliográficas

- 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 CANCER SOCIETY. **Breast Cancer Facts & Figures 2012-2014**. Breast Cancer Facts & Figures, p. 1-11, 2017.
- AUGUSTO, Gustavo Barbosa. **Multiple Kernel Learning for Breast Cancer Classification**. 2014. Trabalho de Mestrado – Departamento de Ciência de Computadores. Faculdade de Ciência da Universidade do Porto em Ciência de Computadores.
- BENATTI, R. H. **Elaboração de banco de imagens mamográficas digitalizadas**. 2003. Tese de Doutorado – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.
- 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. 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.
- BRASIL. **Curso de atualização em mamografia para técnicos e tecnólogos em radiologia**. BRASÍLIA MINISTÉRIO DA SAÚDE. **Secretaria de Gestão do Trabalho e da Educação na Saúde**. Departamento de Gestão da Educação na Saúde, 2014.
- BRASIL STORAGE. 2014. **Onde e por quanto tempo guardar prontuários médicos**. Disponível em: <<https://www.brasilstorage.com.br/blog/onde-e-por-quanto-tempo-guardar-prontuarios-medicos/>>. Acesso em: 30 nov. 2018.
- BRAY, Freddie et al. **Global cancer statistics 2018**: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians, 2018.
- CALAS, Maria Julia Gregorio; GUTFILEN, Bianca; DE ALBUQUERQUE PEREIRA, Wagner Coelho. **CAD e mamografia: por que usar esta ferramenta?**. **Radiologia Brasileira**, v. 45, n. 1, p. 46-52, 2012.

- CALDEIRA, C. P. **Introdução ao Modelo de Dados Relacional**. Universidade de Évora. Disponível em: <<http://www.portalwebmarketing.com/Tecnologia/Introdu%C3%A7%C3%A3oaoModelodeDadosRelacional/MDRDefini%C3%A7%C3%A3oBaseDadosRelacional/tabid/654/Default.aspx>> Acesso em: 11 set. 2017.
- CARVALHO, P. M. de S. **Classificação de tecidos da mama a partir de imagens mamográficas em massa e não massa usando índice de diversidade de McIntosh e máquina de vetores de suporte**. 2012. Trabalho de Mestrado. Universidade Federal do Maranhão.
- COWBURN, P. et al. **O que o PHP pode fazer?** 2018. Disponível em: <https://secure.php.net/manual/pt_BR/intro-whatcando.php>. Acesso em: 28 jan. 2018.
- DANCE, D. R. et al. **Diagnostic Radiology Physics A Handbook for Teachers and Students**. Vienna: International Atomic Energy Agency, 2014.
- DUARTE, *et al.* **Algoritmo de extração de textura baseado em Wavelet e CLBP para classificação de lesões em mamogramas**. In: XIII Workshop de Informática Médica. 2013. p. 174-183.
- DINIZ, J. O. B. **Detecção de Regiões de Massa por Análise Bilateral Adaptada à Densidade da Mama Utilizando Índices de Similaridade e Redes Neurais Convolucionais**. 2017. Trabalho de Mestrado. Universidade Federal do Maranhão.
- DHUNGEL, Neeraj; CARNEIRO, Gustavo; BRADLEY, Andrew P. **Deep structured learning for mass segmentation from mammograms**. In: Image Processing (ICIP), 2015 IEEE International Conference on. IEEE, 2015. p. 2950-2954.
- ELMASRI, Ramez; NAVATHE, Shamkant B. **Sistema de Banco de Dados**. Revisor técnico Luíz Ricardo de Figueiredo. 2005.
- ESCARPINARTI, M. C. **Investigação de formatos e compressão de imagens digitais para processamento de imagens mamográficas de mamas densas**. 2002. Trabalho de Mestrado – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.
- FANTACCI, M. E. et al. **Search of microcalcification clusters with the CALMA CAD station. Proceedings of SPIE**. [S.I]: [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.
- GALANTE, C. A.; MOREIRA, R. L. E.; BRANDÃO, C. F. **Banco de dados orientado a objetos: Uma realidade**. Faculdade Salesiana Maria Auxiliadora. 2007.
- GÓES, R. F.; SCHIABEL, H. ; ESCARPINATI, M. . **Qualidade do Processamento de Imagens Mamográficas Adquiridas em um Digitalizador com Tecnologia a Laser e um**

com **Tecnologia CCD**. In: USP - São Carlos - SP, 2006, São Carlos - SP. II Workshop de Visão Computacional, 2006.

GÔDA, R. R. da S. **Inteligência Computacional Aplicada na Classificação de Microcalcificações Mamárias**. 2016. Trabalho de Mestrado. Universidade Federal Rural do Rio de Janeiro Instituto de Ciências Exatas.

GOLÇALVES, C. V. et al. **O conhecimento de mulheres sobre os métodos para prevenção secundária do câncer de mama**. Faculdade de Medicina, Universidade Federal do Rio Grande. **Ciênc. saúde coletiva** vol.22 no.12. Rio de Janeiro Dec. 2017.

GOVERNO DO BRASIL; **INCA**. Câncer de mama pode ser detectado de forma precoce e tratado com eficácia. **Governo do Brasil, com informações do Ministério da Saúde e do Inca**, 2017. Disponível em: <<http://www.brasil.gov.br/saude/2017/10/cancer-de-mama-pode-ser-detectado-de-forma-precoce-e-tratado-com-eficacia>>. Acesso em: 22 jan. 2018.

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.

INCA. **Mama**. Rio de Janeiro: **Instituto Nacional do Câncer/ Ministério da Saúde**, 2018. Disponível em: <<http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/mama>> Acesso em 22 jan. 2018.

ISO/IEC 9075-1:2016. Disponível em: <<https://www.iso.org/standard/63555.html>>. Acesso em 30 nov. 2018.

KUZMIAK, C. M.; PISANO, E. D.; COLE, E. B.; ZENG, D.; BURNS, C. B.; ROBERTO, C.; PAVIC, D; LEE, Y.; SEO, B. K.; KOOMEN, M.; WASHBURN, D. **Comparison of full-field digital mammography to screen-film mammography with respect to contrast and spatial resolution in tissue equivalent breast phantoms**. *Medical Physics*, v. 32, p. 3144-3150, 2005.

LAZZARI, B. et al. **Physical characteristics of five clinical systems for digital mammography**. *Medical Physics*, v. 34, n. 7, p. 2730–2743, 12 jun. 2007.

Mammography Image Databases. Disponível em: <<http://marathon.csee.usf.edu/Mammography/Database.html%25>> . Acesso em: 2018.

MAGALHÃES, Sarah Braga. **ESTUDO PARA REDUÇÃO DE DOSE ABSORVIDA NA AQUISIÇÃO DE IMAGENS MAMOGRÁFICAS DIGITAIS DE ALTA RESOLUÇÃO**. 2010. Tese de Doutorado. Universidade Federal do Rio de Janeiro.

MARINARO, Maria; TAGLIAFERRI, Roberto (Ed.). **Neural Nets WIRN VIETRI-98: Proceedings of the 10th Italian Workshop on Neural Nets, Vietri sul Mare, Salerno, Italy, 21–23 May 1998**. Springer Science & Business Media, 2012.

MATHEUS, Bruno Roberto Nepomuceno. ***BancoWeb***: base de imagens mamográficas para auxílio em avaliações de esquemas CAD. 2010. Tese de Mestrado – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.

MATHEUS, Bruno Roberto Nepomuceno. **Sistema JAVA para gerenciamento de esquema CADx em mamografia**. 2015. Tese de Doutorado – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.

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.

MATHEUS, B. R. N.; SCHIABEL, H. – **Online mammographic images database for development and comparison of CAD schemes** – Journal of Digital Imaging (Publicado online em 18/05/2010 – doi: 10.1007/s10278-010-9297-2), v. 24, n. 3, p. 500-506, 2011.

MICROSOFT. **Transações (Mecanismo de Banco de Dados)**. 2018. Disponível em: <[https://technet.microsoft.com/pt-br/library/ms190612\(v=sql.105\).aspx](https://technet.microsoft.com/pt-br/library/ms190612(v=sql.105).aspx)>. Acesso em: 21 mar. 2018.

MIGOWSKI, Arn et al. **Diretrizes para detecção precoce do câncer de mama no Brasil**. II- Novas recomendações nacionais, principais evidências e controvérsias. Cadernos de Saúde Pública, v. 34, p. e00074817, 2018.

MOREIRA, I. C. et al. **INbreast**. Academic Radiology, v. 19, n. 2, p. 236–248, fev. 2012.

MOURA, D. C.; LÓPEZ, M. A. G.. **An evaluation of image descriptors combined with clinical data for breast cancer diagnosis**. International journal of computer assisted radiology and surgery, v. 8, n. 4, p. 561-574, 2013.

NASCIMENTO, M. Z. do. **Método computacional automático para correção do efeito “heel” nas imagens radiográficas**. 2005. Tese Doutorado - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.

NASCIMENTO, F. B. do; PITTA, M. G. da R.; RÊGO, M. J. B. de M. **Análise dos principais métodos de diagnóstico de câncer de mama como propulsores no processo inovativo**. Arquivos de Medicina, v. 29, n. 6, p. 153-159, 2015.

NETO, O. P. da S. **Deteção automática de massas em imagens mamográficas usando particle swarm optimization (PSO) e índice de diversidade funcional**. 2016. Trabalho de Mestrado. Universidade Federal do Maranhão.

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. Mammographic databases. **Breast Disease**, n. 10(3,4), p. 137-150, 1998.
- NUNES, F. de L. dos S. **Esquema de diagnóstico auxiliado por computador para detecção de agrupamentos de microcalcificações por processamento de imagens mamográficas**. 1997. Trabalho de Mestrado – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.
- 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. 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.
- ODEH, Mohammed et al. **A use-case driven approach in requirements engineering: the MammoGrid project**. arXiv preprint cs/0402008, 2004.
- ORACLE. **Oracle MySQL**. Disponível em: <<https://www.oracle.com/br/mysql/index.html>>. Acesso em: 28 jan. 2018.
- OSHIRO, M. L., et al. **Câncer de mama avançado como evento sentinela para avaliação do programa de detecção precoce do câncer de mama no Centro-Oeste do Brasil**. *Revista Brasileira de Cancerologia* 60.1 (2014): 15-23.
- PEREIRA, S. do L. **Estrutura de dados fundamentais: conceitos e aplicações**. Editora Érica, 2006.

- PERRY, Nicholas et al. **European guidelines for quality assurance in breast cancer screening and diagnosis**. —summary document. *Annals of Oncology*, v. 19, n. 4, p. 614-622, 2008.
- PIRES, S. R.; MEDEIROS, R. B.; ELIAS, S. **QualIM®**: software para treinamento na interpretação de imagens médicas digitais. *Radiologia Brasileira*, v. 41, n. 6, p. 391-395, 2008.
- ROCHA, A. R. R. et al. **Unified Modeling Language**. 2001. Disponível em: <http://www.dei.unicap.br/~almir/seminarios/2001.1/5mno/uml/>. Acesso em: 18 mar. 2018.
- 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. **Esquemas CAD**: uma análise dos seus aspectos e aplicações como ferramenta de auxílio ao diagnóstico em mamografia – Cap. do livro “Tecnologias, Técnicas e Tendências em Engenharia Biomédica” (ISBN 978-85-7917-289-2), Canal 6 Editora (Ed. Adriano O. Andrade; Alcimar B. Soares; Alexandre Cardoso; Edgard A. Lamounier), p. 34-54, 2014.
- SCHIABEL, H. N. F. L. S.; ESCARPINATI, M. C.; BENATTI, R. H. **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.
- SCHIABEL, H.; MATHEUS, B. R. N.; ANGELO, M. F.; PATROCINIO, A. C.; VENTURA, L. – **A prototype of mammography CADx scheme integrated to imaging quality evaluation techniques – SPIE Medical Imaging 2011: Computer-aided Diagnosis**, v. 7963, p. 796323-1 - 796323-12 – Orlando, FL (USA) – Feb, 12-17, 2011.
- SILVA, Ronaldo Corrêa Ferreira da; HORTALE, Virginia Alonso. **Rastreamento do câncer de mama no Brasil: quem, como e por quê?**. *Rev. bras. cancerol*, v. 58, n. 1, p. 67-71, 2012.
- STAIR, Ralph et al. **Principles of information systems**. Cengage Learning Australia, 2011.
- SOLID IT. DB Engines. 2018. Disponível em: <<https://db-engines.com/en/system/MySQL>>. Acesso em: 28 jan. 2018.
- SOLOMONIDES, Anthony Eugene. **Data, metadata, and workflow in healthcare informatics**. 2018. Tese de Doutorado. University of the West of England.

- SOUSA, M. A. Z. **Desenvolvimento de um objeto simulador de mama para avaliação de esquemas CAdE/CADx**. 2017. Tese (Doutorado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.
- SOUSA, M. A. Z.; MARCOMINI, K. D. ; BAKIC, P. R. ; MAIDMENT, A. D.A. ; SCHIABEL, H. **Observer study to evaluate the simulation of mammographic calcification clusters**. In: SPIE Medical Imaging, 2016, San Diego. v. 9787. p. 97871D. doi: 10.1117/12.2217257.
- SOUSA, M. A.; MATHEUS, B. R. N.; SCHIABEL, H. **Development of a structured breast phantom for evaluating CAdE/Dx schemes applied on 2D mammography** – Biomedical Physics & Engineering Express, v. 4, n. 4, 2018 (<https://doi.org/10.1088/2057-1976/aac2f2>)
- SOUZA, Johnatan Carvalho et al. **Diagnóstico de câncer de mama a partir de imagens de mamografia 2d utilizando descritores de forma 3d**. 2018.
- SUCKLING, J. et al. **The mammographic image analysis society digital mammogram database**. Excerpta Medica. International Congress Series. 1994.
- TAKAI, K. O; ITALIANO, C. I.; FERREIRA, E. J. **Introdução a banco de dados**. DCC-IME-USP. 2005.
- VALENTE, Nelma T. Zubeck; FUJINO, Asa. **Atributos e dimensões de qualidade da informação nas Ciências Contábeis e na Ciência da Informação: um estudo comparativo**. Perspectivas em Ciência da Informação, v. 21, n. 2, p. 141-167, 2016.
- WADE, B. W.; CHAMBERLIN, D. D. **IBM relational database systems: The early years**. *IEEE Annals of the History of Computing*, v. 34, n. 4, p. 38-48, 2012.
- WARREN R., et al. **MammoGrid: a prototype distributed mammographic database for Europe**. *Clin Radiol* 2007; 62: 1044–1051.
- W3SCHOOL Home. HTML Introdução**. 2016. Disponível em: <<http://www.escolaw3.com/tutoriais/html/introducao>>. Acesso em: 28 jan. 2018.
- YUAN, S. et al. **Abstract PD2-15: Effect of mammography screening frequency on false-positive biopsy rates and detection of local recurrence among breast cancer survivors**. 2018.