

**UNIVERSIDADE FEDERAL DO MARANHÃO  
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE ELETRICIDADE**

**LUIS CLAUDIO DE OLIVEIRA SILVA**

**IMPLEMENTAÇÃO DE UM SISTEMA DE TELEDIAGNÓSTICO PARA  
CLASSIFICAÇÃO DE MASSAS EM IMAGENS MAMOGRÁFICAS  
USANDO ANÁLISE DE COMPONENTES INDEPENDENTES**

São Luís  
2012

**LUIS CLAUDIO DE OLIVEIRA SILVA**

**IMPLEMENTAÇÃO DE UM SISTEMA DE TELEDIAGNÓSTICO PARA  
CLASSIFICAÇÃO DE MASSAS EM IMAGENS MAMOGRÁFICAS  
USANDO ANÁLISE DE COMPONENTES INDEPENDENTES**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Eletricidade da Universidade Federal do Maranhão, como parte dos requisitos para obtenção do título de Mestre em Engenharia de Eletricidade na área de concentração Automação e Controle.

Orientador: Prof. Ph.D. Allan Kardec Duailibe Barros Filho

Co-orientador: Prof. Dr. Ewaldo Éder Carvalho Santana

São Luís

2012

Silva, Luis Claudio de Oliveira.

Implementação de um sistema de telediagnóstico para classificação de massas em imagens mamográficas usando análise de componentes independentes/ Luis Claudio de Oliveira Silva – São Luís, 2012.

67 f.

Impresso por computador (fotocópia).

Orientador: Allan Kardec Duailibe Barros Filho.

Co-orientador: Ewaldo Éder Carvalho Santana.

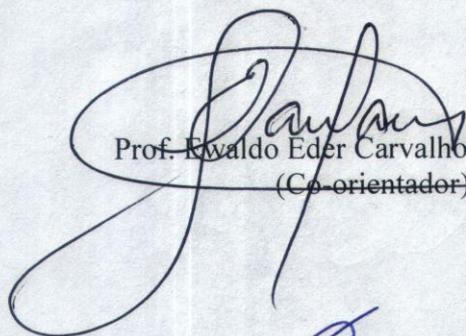
**IMPLEMENTAÇÃO DE UM SISTEMA DE TELEDIAGNÓSTICO  
PARA CLASSIFICAÇÃO DE IMAGENS MAMOGRÁFICAS  
USANDO ANÁLISE DE COMPONENTES INDEPENDENTES**

**Luis Cláudio de Oliveira Silva**

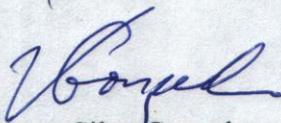
Dissertação aprovada em 24 de julho de 2012.



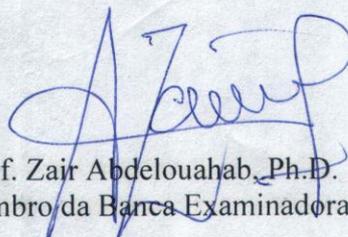
Prof. Allan Kardec Duailibe Barros Filho, PhD.  
(Orientador)



Prof. Eivaldo Eder Carvalho Santana, Dr.  
(Co-orientador)



Prof. Vicente Silva Gonçalves Neto, Dr.  
(Membro da Banca Examinadora)



Prof. Zair Abdelouahab, Ph.D.  
(Membro da Banca Examinadora)

## **AGRADECIMENTOS**

À Deus, pela vida.

À minha mãe, Maria Ivete, pelo amor, criação e pela paciência em me educar para trilhar os caminhos da vida. Aos meus irmãos, Keyla Maria e Cláudio Junior, pelo companheirismo, apoio e incentivo. À minha companheira Mayra Cassandra, pelo estímulo, apoio logístico, amor e principalmente paciência.

Aos professores Allan Kardec e Ewaldo Eder Santana, pela orientação, apoio e principalmente, por terem acreditado e investido neste projeto.

Aos amigos do Laboratório de Processamento da Informação biológica da UFMA.

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

*“Uma vez que não podemos ser universais e saber tudo quanto se pode saber acerca de tudo, é preciso saber-se um pouco de tudo, pois é muito melhor saber-se alguma coisa de tudo do que saber-se tudo apenas de uma coisa.”*

Blaise Pascal

## RESUMO

Este trabalho propõe a modelagem e implementação de um sistema de telediagnóstico para análise e detecção automática de lesões em imagens mamográficas, baseado em análise de componentes independentes e máquina de vetor de suporte. O sistema analisa imagens de mamografia digital enviadas pela Internet e fornece um diagnóstico da imagem, indicando a presença de regiões suspeitas, que podem ser confirmadas por um especialista em imagens mamográficas. Além de apresentar a metodologia para o desenvolvimento do sistema proposto, foi desenvolvido um protótipo para a realização de testes objetivando medir sua eficiência. A base de dados usada para treinamento e teste dos algoritmos foi a mini-MIAS, e foi empregada análise de componentes independentes para extrair os filtros usados na segmentação das regiões de interesse, bem como máquina de vetor de suporte para classificar as regiões de interesse em normais ou suspeitas. A partir de testes realizados com a base de dados utilizada, obteve-se média de acerto de 87,8% para imagens que contém lesões.

Palavras-chave: mamografia, câncer de mama, telediagnóstico, análise de componentes independentes, máquina de vetor de suporte.

## **ABSTRACT**

This thesis proposes the modeling and implementation of a telediagnostic system for analysis and detection of lesions in mammographic images based on independent component analysis and support vector machine. The system analyzes images from digital mammography sent over the Internet and provides a diagnostic, indicating the presence of suspicious regions, which can be confirmed by a specialist in mammographic images. Besides presenting the methodology for the development of the proposed system, a prototype was developed for testing and to measure its efficiency. The database used for training and testing of the algorithms is the mini-MIAS, and was employed independent component analysis to extract the filters used in segmenting the regions of interest, as well support vector machine to classify regions of interest in normal or suspicious. From tests with the database used, we obtained an average accuracy of 87.8% for images containing lesions.

**Keywords:** Mammogram, breast cancer, telediagnostic, independent component analysis, support vector machine.

## **LISTA DE TABELAS**

Tabela 1 - Funções de núcleo mais comuns. ....	40
Tabela 2 - Treinamento SVM com amostras de 30x30 pixels: escolha do número de IC's. ....	54
Tabela 3 - Treinamento SVM com amostras de 50x50 pixels: escolha do número de IC's. ....	55

## LISTA DE FIGURAS

Figura 1 - Tipos de neoplasias.....	17
Figura 2 - Neoplasias. (a) Benigna. (b) Maligna. Fonte: (SUCKLING, PARKER, <i>et al.</i> , 1994). .....	18
Figura 3 - Neoplasia do tipo calcificação. (a) Imagem original. (b) Imagem realçada. Identificação no MIAS: mdb209. Fonte: (SUCKLING, PARKER, <i>et al.</i> , 1994). .....	18
Figura 4 - Neoplasias do tipo calcificação. (a) Identificação no MIAS: mdb245. (b) Identificação no MIAS: mdb248. Fonte: (SUCKLING, PARKER, <i>et al.</i> , 1994).....	19
Figura 5 - Mamografia convencional obtida a partir do mini-MIAS. Fonte: (SUCKLING, PARKER, <i>et al.</i> , 1994). .....	20
Figura 6 - Neoplasias realçadas em imagens mamográficas obtidas a partir do mini-MIAS. (a) Benigna; identificação: mdb010. (b) Maligna; identificação mdb148. Fonte: (SUCKLING, PARKER, <i>et al.</i> , 1994). .....	20
Figura 7 - Exemplo de remoção de artefatos em mamografia. (a) mamografia com artefatos; (b) mamografia sem artefatos. ....	23
Figura 8 - Representação de uma imagem M x N .....	24
Figura 9 - Passos fundamentais para o processamento de imagens .....	25
Figura 10 - Amostra de uma imagem mamográfica como combinação linear de suas características.....	30
Figura 11 - Sinais artificiais originais.....	31
Figura 12 - Sinais artificiais misturados. ....	31
Figura 13- Sinais artificiais reconstruídos .....	31
Figura 14 - Um conjunto de treinamento de duas classes, os vetores de suporte e o hiperplano ótimo.....	36
Figura 15 - Distância entre os hiperplanos <b>H1</b> e <b>H2</b> . .....	37
Figura 16 - Atividades do processo de software no modelo cascata. Adaptado de (PRESSMAN, 2002). .....	41
Figura 17 - Visão geral do sistema proposto. ....	42
Figura 18 - Arquitetura do sistema proposto. ....	43
Figura 19 - Fluxograma do sistema proposto. ....	44
Figura 20 - Diagrama de casos de uso do sistema proposto (requisitos funcionais). .....	45
Figura 21 - Diagrama de classes do sistema proposto.....	46
Figura 22 - Modelo conceitual do banco de dados do sistema proposto. ....	47
Figura 23 - Modelo lógico do banco de dados do sistema proposto. ....	48
Figura 24 - Diagrama de atividades do módulo de processamento das imagens. ....	50

Figura 25 - Resultado do pré-processamento. (a) Imagem original. (b) Imagem resultante depois de retirados os artefatos.....	51
Figura 26 - Funções bases extraídas de regiões de interesses com lesões.....	52
Figura 27 - Resultado da segmentação. (a) Imagem segmentada. (b) Região com provável lesão.....	52
Figura 28 - Resultado da filtragem de convolução com as funções base na etapa de segmentação.....	53
Figura 29 - Região de interesse que será classificada. (a) Resultado da segmentação. (b) Agrupamento com maior energia. (c) Região recortada para ser classificada. ....	55
Figura 30 - Divisão da região segmentada em quadrantes para classificação.....	56
Figura 31 - Resultado da classificação em uma imagem com lesão. ....	56
Figura 32 - Tela de <i>login</i> do sistema proposto. ....	57
Figura 33 - Tela principal do usuário solicitante.....	58
Figura 34 - Tela de cadastro de pacientes. ....	58
Figura 35 - Tela de cadastro dos exames.....	59
Figura 36 - Tela de visualização dos exames. ....	60
Figura 37 - Tela de resultado do processamento do exame.....	60
Figura 38 - Tela de listagem de exames do usuário especialista. ....	61
Figura 39 - Tela de liberação do exame pelo especialista. ....	62
Figura 40 - Exemplos de imagens classificadas pelo sistema: regiões quadriculadas correspondem à classificação do sistema; o círculo corresponde à localização definida pelo mini-MIAS. ....	64

## LISTA DE SIGLAS

BSS	Separação Cega de Fontes
CAD	Diagnóstico Auxiliado por Computador
ER	Entidade-Relacionamento
HTML	Linguagem de Marcação de Hipertexto
ICA	Análise de Componentes Independentes
INCA	Instituto Nacional de Câncer
JEE	<i>Java Enterprise Edition</i>
JSP	<i>Java Server Pages</i>
MIAS	<i>Mammographic Image Analysis Society</i>
RHP	Real Hospital Português
ROI	<i>Region of Interest</i>
SGBDR	Sistema de Gerenciamento de Banco de Dados Relacional
SUS	Sistema Único de Saúde
SVM	<i>Support Vector Machine</i>
UFPE	Universidade Federal de Pernambuco
UML	<i>Unified Modeling Language</i>

# SUMÁRIO

1 INTRODUÇÃO.....	15
2 CÂNCER DE MAMA.....	17
2.1 Mamografia .....	19
3 DIAGNÓSTICO AUXILIADO POR COMPUTADOR.....	22
3.1 Sistema de auxílio à detecção/diagnóstico em mamografias.....	22
3.2 Processamento de imagens digitais .....	24
3.3 Informática médica.....	26
3.3.1 Aplicações da telemedicina .....	27
3.3.2 A telemedicina no Brasil .....	28
4 ANÁLISE DE COMPONENTES INDEPENDENTES.....	30
4.1 Definições.....	32
4.2 Independência estatística e não-correlação de variáveis .....	33
4.3 Estimação das componentes independentes .....	33
4.4 Negentropia como medida de não-gaussianidade .....	34
5 MÁQUINA DE VETOR DE SUPORTE .....	36
6 SISTEMA PROPOSTO.....	41
6.1 Visão geral do sistema.....	42
6.2 Modelagem e desenvolvimento do sistema.....	44
6.3 Implementação do sistema.....	49

6.3.1 Camada de processamento de imagens .....	50
6.3.2 Camada de apresentação.....	57
7 RESULTADOS E DISCUSSÃO .....	63
8 CONCLUSÃO E TRABALHOS FUTUROS .....	65
REFERÊNCIAS .....	67

## 1 INTRODUÇÃO

Pesquisas do Instituto Nacional do Câncer (INCA, 2007) revelam que são esperados para o Brasil em 2012, 52.680 novos casos de câncer de mama, sendo este o segundo tipo de câncer que mais acomete mulheres em todo o mundo, perdendo apenas para o câncer de pele do tipo não melanoma. Esta quantidade elevada de óbitos se deve, na maioria das vezes, pelo diagnóstico tardio da doença, o que dificulta o tratamento e a cura.

A mamografia é um dos exames realizados para diagnosticar a doença, permitindo revelá-la ainda em sua fase inicial, quando as lesões ainda são pequenas e possuem poucos milímetros. Consiste em um tipo de radiografia da mama que, quando analisada por um especialista, pode sugerir um diagnóstico ainda no seu estágio inicial. Embora seja um exame bastante utilizado e com uma efetividade de 46% a 88%, a identificação de lesões depende de fatores tais como: tamanho e localização da lesão, a densidade do tecido da mama, habilidade de detecção por parte do especialista, dentre outros (INCA, 2012).

Por esse motivo, a literatura científica vem propondo, a cada ano, métodos para identificação de anomalias em exames de mamografias, na forma de ferramentas de diagnóstico auxiliado por computador. A aplicação destas ferramentas tem contribuído para a localização e classificação das lesões, auxiliando especialistas e pacientes no processo de diagnóstico da doença (TANG, RANGAYYAN, *et al.*, 2009).

Dos trabalhos apresentados na literatura científica, muitos métodos de sugestão de diagnóstico baseiam-se em técnicas de processamento de imagem para procurar regiões de interesse em imagens mamográficas, através de análise de textura. Alguns destes métodos utilizam análise de componentes independentes para extração de características das regiões de interesse e assim realizar a segmentação destas regiões. Acredita-se que a aplicação deste método para detecção e segmentação de massas em imagens mamográficas possa trazer bons resultados que posteriormente podem ser aplicados em algoritmos de classificação, como por exemplo, a máquina de vetor de suporte, análise discriminante e redes neurais multicamadas *perceptron*.

No entanto, poucos trabalhos apresentam sistemas de auxílio ao diagnóstico, desenvolvidos a partir de técnicas de detecção automática e quantificação de anomalias em imagens mamográficas na forma de aplicações de telediagnóstico em tempo real,

possibilitando seu uso por usuários em hospitais e centros de tratamento, bem como especialistas da área médica, utilizando a Internet como ferramenta de comunicação e colaboração. Além de fornecer um resultado a partir da análise de imagens pela Web, um sistema de telediagnóstico pode fazer uso de bancos de dados para armazenar estes resultados, permitindo o acompanhamento da evolução da doença bem como o cruzamento destas análises com a *anamnese* do paciente.

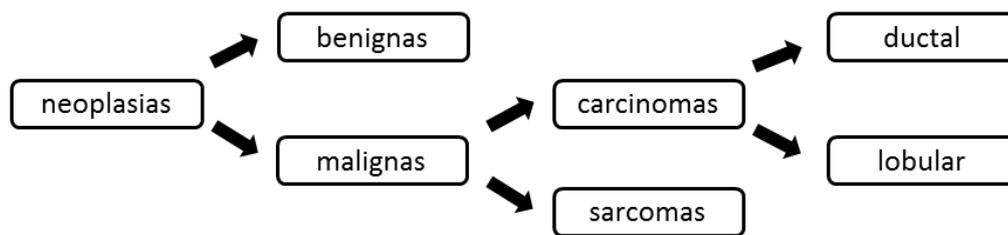
O objetivo deste trabalho é modelar e implementar um sistema de telediagnóstico, capaz de analisar imagens de mamografias digitais enviadas pela Internet e fornecer um diagnóstico destas imagens, indicando a presença de regiões suspeitas, que poderão ser analisadas, posteriormente, por um especialista em imagens mamográficas. Especificamente, pretende-se apresentar a metodologia para o desenvolvimento do sistema, incluindo os modelos de banco de dados e das classes existentes na aplicação, além de efetuar o treinamento dos algoritmos usados na localização e classificação das regiões suspeitas utilizando análise de componentes independentes, bem como desenvolver um protótipo do sistema para a realização de testes objetivando medir sua eficiência.

O trabalho está organizado em mais sete capítulos. O **Capítulo 2** aborda informações básicas sobre o câncer de mama, características das lesões e o uso da mamografia como método para detecção das lesões. O **Capítulo 3** aborda os conceitos de diagnóstico auxiliado por computador, os conceitos de processamento de imagens digitais e informática médica, incluindo as definições de telemedicina e telediagnóstico. No **Capítulo 4**, serão desenvolvidas as teorias necessárias para entendimento e aplicação de análise de componentes independentes para extração de características. Os conceitos e o funcionamento da máquina de vetor de suporte serão abordados no **Capítulo 5**. O desenvolvimento do sistema, incluindo a modelagem dentro dos paradigmas da engenharia de software, treinamento dos algoritmos e elaboração do protótipo do sistema será abordado no **Capítulo 6**. No **Capítulo 7** serão apresentados os resultados obtidos nos testes do sistema e o trabalho será finalizado no **Capítulo 8**, com a conclusão e proposta de trabalhos futuros.

## 2 CÂNCER DE MAMA

O câncer de mama pode ser definido como sendo uma disfunção nas células que compõem o tecido mamário, capaz de ocasionar multiplicação celular desordenada e causando o surgimento de estruturas benignas ou malignas. As estruturas malignas podem se espalhar para outras regiões do corpo, invadindo outros órgãos (INCA, 2012). Assim também ocorre com o câncer de mama. O termo *neoplasia* é utilizado para alterações celulares que acarretam no crescimento desordenado destas células, podendo ser benigna ou maligna.

As neoplasias malignas se dividem, por sua vez, em tumores epiteliais, ou carcinomas, de origem ductal ou lobular, e em sarcomas, originados a partir de tecidos conjuntivos. Os carcinomas ductais formam-se nos ductos que levam o leite do lóbulo para os mamilos. Já os carcinomas lobulares formam-se nos bulbos que produzem o leite materno. A Figura 1 simplifica os tipos de neoplasias.

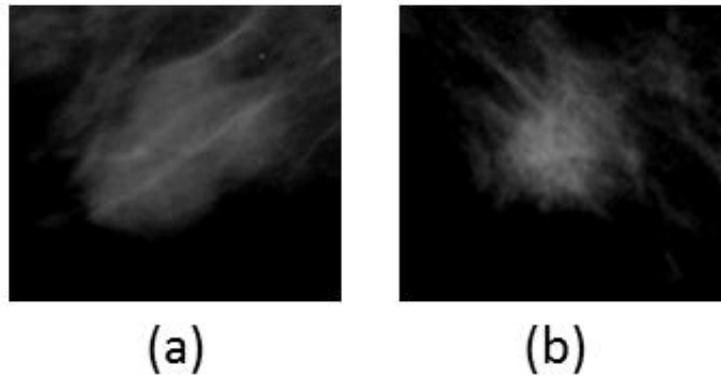


**Figura 1 - Tipos de neoplasias.**

Assim como qualquer tipo de câncer, o câncer de mama pode se espalhar para outras partes do corpo, processo este conhecido como metástase. Por isso é de vital importância que sua detecção seja a mais precoce possível, aumentando assim a chance de tratamento e de cura. As formas mais eficazes para detecção precoce do câncer de mama são o exame clínico e a mamografia (INCA, 2012).

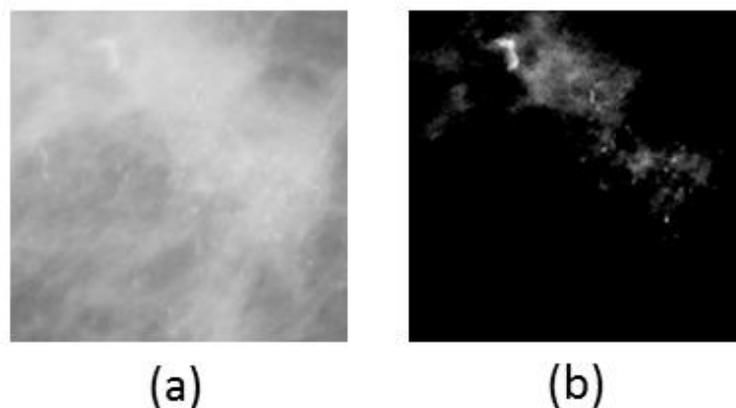
A manifestação do câncer pode se dar através do surgimento de massas e calcificações. As massas, ou nódulos, correspondem a um conjunto de células aglomeradas, tornando-se mais densas do que o tecido ao redor, podendo configurar neoplasias benignas ou malignas. As calcificações correspondem a depósitos de sais de cálcio desenvolvidos no tecido mamário, que podem inclusive ser consequências de processos inflamatórios, alterações degenerativas, processos tóxicos metabólicos, traumatismos ou resultados de processos secretores ativos de células tumorais (INCA, 2007).

Através da Figura 2, podem-se visualizar duas imagens de neoplasias: (a) benigna; (b) maligna. As imagens foram obtidas do banco de dados de imagens mamográficas mini-MIAS (*Mammographic Image Analysis Society*) database (SUCKLING, PARKER, *et al.*, 1994).



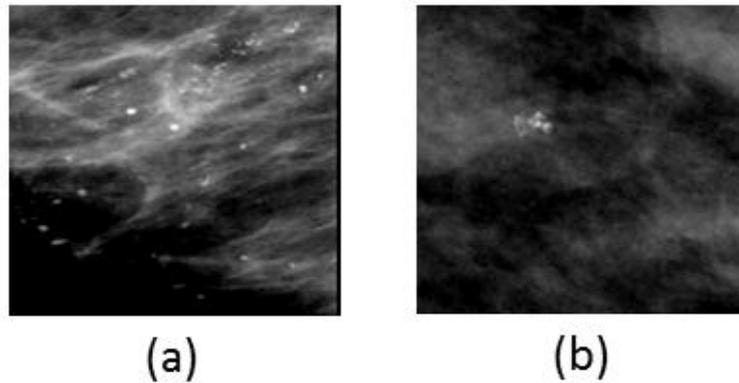
**Figura 2 - Neoplasias. (a) Benigna. (b) Maligna.**  
**Fonte: (SUCKLING, PARKER, *et al.*, 1994).**

Através da Figura 3, pode-se visualizar uma neoplasia do tipo calcificação: (a) corresponde à região de interesse, sem realce para melhorar visualização; (b) corresponde à mesma região de interesse, com tratamento de realce para melhorar a visualização das calcificações.



**Figura 3 - Neoplasia do tipo calcificação.**  
**(a) Imagem original. (b) Imagem realçada. Identificação no MIAS: mdb209.**  
**Fonte: (SUCKLING, PARKER, *et al.*, 1994).**

Através da Figura 4, podem-se visualizar duas neoplasias do tipo calcificação. Ambas foram recortadas de imagens contendo lesões. Após um tratamento nas imagens, é possível notar as calcificações, como sendo os pontos brancos dentro da mamografia.



**Figura 4 - Neoplasias do tipo calcificação.**  
(a) Identificação no MIAS: mdb245. (b) Identificação no MIAS: mdb248.  
Fonte: (SUCKLING, PARKER, *et al.*, 1994)

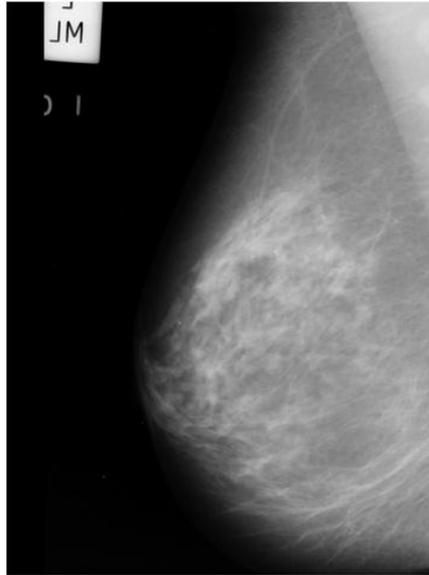
## 2.1 Mamografia

Basicamente, a mamografia é uma forma de radiografia da mama. Seu objetivo é rastrear a mama através de imagens de alta resolução, permitindo a detecção de anomalias na mama. Para a realização deste exame, é utilizado um aparelho de raios-X, a fim de se obter imagens radiográficas do tecido mamário.

A mamografia é realizada por um equipamento denominado *mamógrafo*, que comprime as mamas com o objetivo de permitir uma melhor visualização de pequenas alterações, muitas vezes difíceis de serem detectadas a olho nu (INCA, 2012). A detecção nos estágios iniciais aumenta a chance de tratamento e a radiografia da mama permite detectar alterações ainda neste estágio.

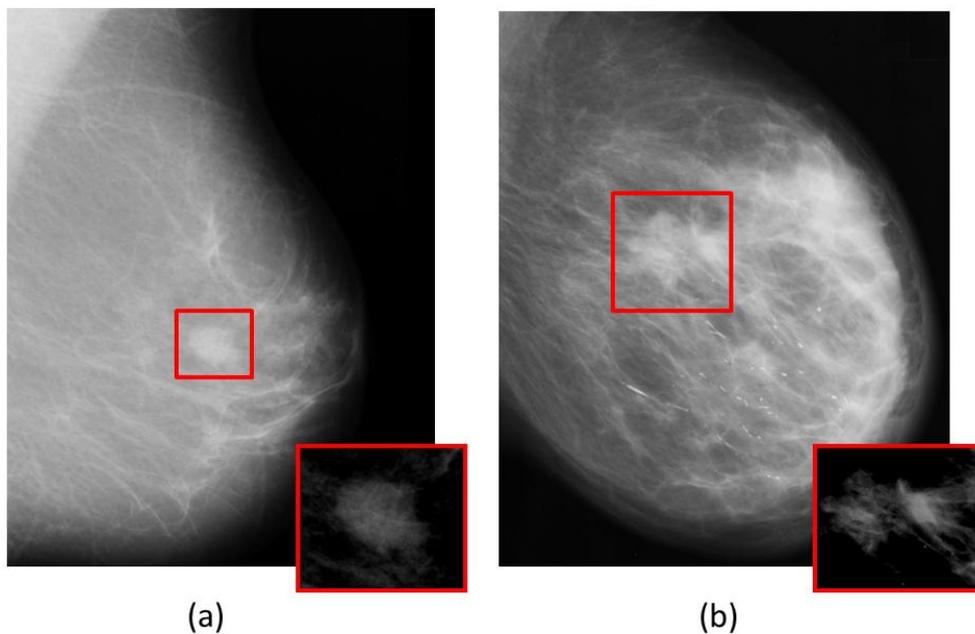
Existem dois tipos de mamografia: a convencional e a digital. Apesar dos dois sistemas utilizarem raios-X para produzir a imagem da mama, a diferença está em como a imagem é captada. A mamografia convencional utiliza um filme para a exposição da imagem após a exposição da mama aos raios-X. Já a mamografia digital possui um sensor que transforma os raios-X em sinais elétricos, transmitindo-os para um computador, possibilitando seu armazenamento e processamento a partir de programas específicos de análise de imagens mamográficas (MASTOLOGIA, 2012).

Observa-se na Figura 5 uma mamografia convencional obtida a partir da base de dados mini-MIAS.



**Figura 5 - Mamografia convencional obtida a partir do mini-MIAS.**  
**Fonte: (SUCKLING, PARKER, *et al.*, 1994).**

A Figura 6 mostra duas imagens mamográficas contendo, cada uma, uma neoplasia destacada e ampliada após um tratamento de imagem para realçar a região de interesse: (a) neoplasia benigna; (b) neoplasia maligna.



**Figura 6 - Neoplasias realçadas em imagens mamográficas obtidas a partir do mini-MIAS.**  
**(a) Benigna; identificação: mdb010. (b) Maligna; identificação mdb148.**  
**Fonte: (SUCKLING, PARKER, *et al.*, 1994).**

Até o presente momento, a mamografia prevalece como o melhor método por imagem para o diagnóstico do câncer de mama, possibilitando o diagnóstico em sua fase

inicial e possibilitando um tratamento não invasivo ou menos agressivo, e com maior qualidade de vida e menor mortalidade para o paciente (MASTOLOGIA, 2012). Além disso, permite detectar lesões pequenas que ainda não são palpáveis, por meio de várias alterações que podem indicar malignidade tais como microcalcificações agrupadas, assimetrias, distorções da arquitetura mamária e nódulos.

### 3 DIAGNÓSTICO AUXILIADO POR COMPUTADOR

Com o objetivo de fornecer ao especialista na área médica ferramentas para auxiliar na detecção e diagnóstico de lesões em imagens mamográficas, diversos trabalhos vêm propondo o desenvolvimento de sistemas computacionais capazes de fornecer uma sugestão de diagnóstico. Estas ferramentas são denominadas de Sistemas de Auxílio à Detecção/Diagnóstico ou simplesmente Sistemas CAD/CADx (do inglês, *Computer-Aided Detection/Diagnosis*) e são capazes de fornecer uma segunda opinião para o especialista, reduzindo erros de interpretação causados por má qualidade da imagem ou grande volume de exames a serem analisados para poucos profissionais disponíveis.

Estes trabalhos tiveram início em 1967 (WINSBERG, ELKIN, *et al.*, 1967), com a elaboração de um procedimento para analisar a densidade óptica de imagens mamográficas, e a partir desta análise identificar áreas suspeitas. Porém, foi com o desenvolvimento de sistemas radiológicos mais modernos e técnicas mais avançadas de processamento de imagens, que os sistemas CAD passaram a produzir resultados mais confiáveis, podendo até detectar regiões de microcalcificações (DANCE e DAVIES, 1990).

#### 3.1 Sistema de auxílio à detecção/diagnóstico em mamografias

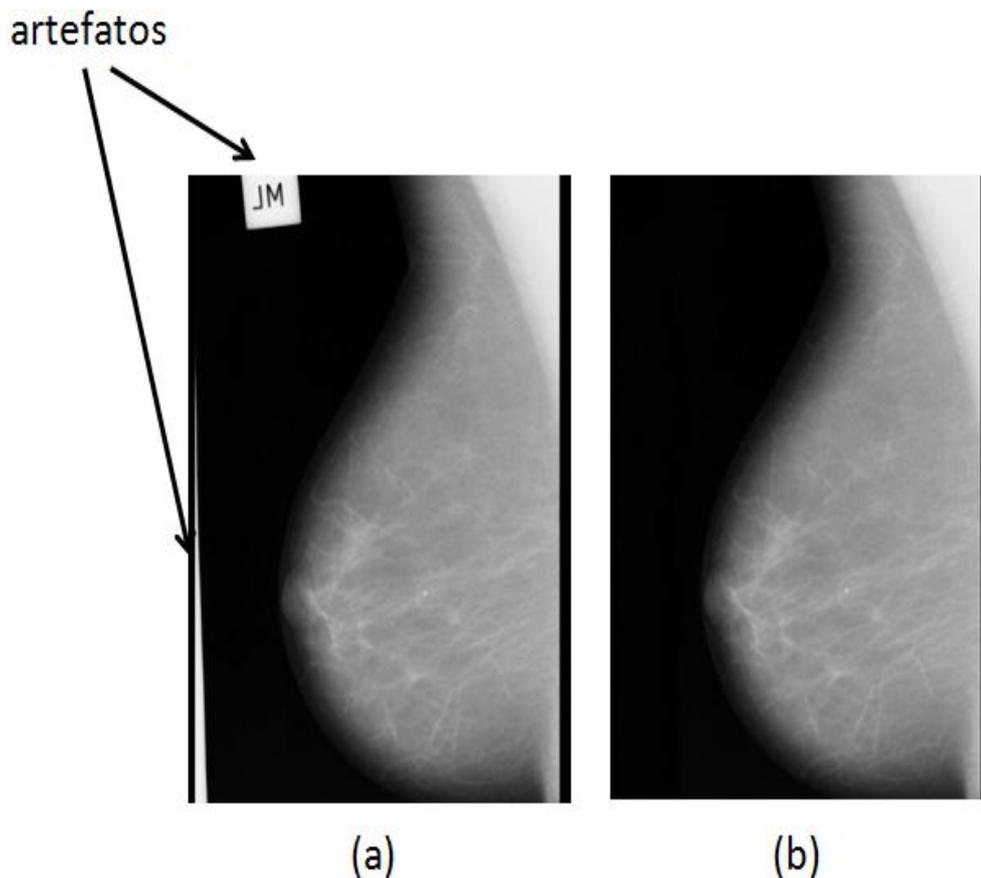
Os sistemas CAD/CADx são capazes de fornecer uma segunda opinião ao especialista, através de análises das informações médicas por meio de algoritmos e computadores. Em se tratando de análise de imagens mamográficas, por exemplo, estes sistemas podem detectar regiões suspeitas em uma mamografia, além de analisar estas regiões quanto às suas características de benignidade e malignidade, auxiliando no apoio à decisão do diagnóstico médico (TANG, RANGAYYAN, *et al.*, 2009).

Através de técnicas de processamento de imagem, reconhecimento de padrões e inteligência artificial, os sistemas CAD/CADx são capazes de fornecer uma *segunda opinião*, na detecção de lesões. Existem dois tipos de sistemas CAD para análise de mamografias: uma baseada em tecnologia de mamografia em filme radiográfico e a outra baseada em tecnologia de mamografia digital (TANG, RANGAYYAN, *et al.*, 2009).

Nos sistemas CAD/CADx baseados em tecnologia de mamografia em filme radiográfico, as mamografias são digitalizadas utilizando scanners de alta resolução óptica,

algo em torno de 3200 pontos por polegada (*dpi*, do inglês *dot per inch*), e armazenados em computadores para análise. Já nos sistemas baseados em tecnologia de mamografia digital, a imagem é adquirida através de um receptor digital e enviada diretamente para o computador, ao invés de usar um filme radiográfico. Desta forma, a mamografia digital possibilita maior rapidez na conclusão do exame, bem como mais eficiência na detecção de lesões ao passo que diminui a taxa de ruído acrescentado na imagem enviada para o computador.

Após a aquisição, a imagem mamográfica é sujeita a técnicas de análise baseadas em processamento de imagem para remoção de artefatos e redução de ruídos. A Figura 7 mostra o resultado da remoção de artefatos em uma imagem mamográfica. Em seguida, são aplicados métodos de detecção de massas ou lesões nas imagens, a fim de encontrar regiões suspeitas que possam ser, posteriormente, classificadas em benignas ou malignas.



**Figura 7 - Exemplo de remoção de artefatos em mamografia.**  
(a) mamografia com artefatos; (b) mamografia sem artefatos.

A seguir, serão apresentadas algumas técnicas de processamento de imagens empregadas em imagens mamográficas, algumas delas utilizadas neste trabalho.

### 3.2 Processamento de imagens digitais

O processamento de imagens digitais compreende um conjunto de procedimentos executados em uma imagem por meio de hardware e software (MARQUES FILHO e VIEIRA NETO, 1999), bem como fundamentos teóricos usados na elaboração de algoritmos usados em tarefas como filtragem, reconhecimento de padrões, representação, dentre outras (GONZALEZ, WOODS e EDDINS, 2003).

Uma imagem digital pode ser representada matematicamente por uma função  $f(x, y)$ , onde  $x$  e  $y$  representam as coordenadas espaciais e o valor de  $f$  em qualquer ponto, o brilho ou nível de cinza da imagem naquele ponto (MARQUES FILHO e VIEIRA NETO, 1999). Desse modo, uma imagem digital é uma imagem discretizada tanto em coordenadas espaciais quanto em brilho (GONZALEZ, WOODS e EDDINS, 2003), sendo comumente representada por uma matriz cujos índices de linha e coluna identificam um ponto na imagem, e o correspondente valor o brilho ou nível de cinza, naquele ponto. Os elementos desta matriz são comumente denominados *pixels*. A Figura 8 mostra a representação de uma imagem de tamanho  $M \times N$ .

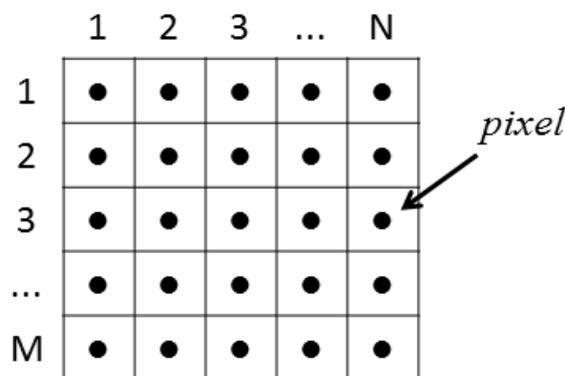


Figura 8 - Representação de uma imagem  $M \times N$

Os passos fundamentais para efetuar o processamento de uma imagem encontram-se ilustrados na Figura 9. De acordo com a figura, o primeiro passo é a *aquisição* da imagem, através de algum dispositivo capaz de capturar a imagem, transformando-a em uma imagem digital capaz de ser processada por um computador ou dispositivo com este fim. Este dispositivo pode ser, por exemplo, um *scanner* ou uma câmera digital.



**Figura 9 - Passos fundamentais para o processamento de imagens**

Logo após a *aquisição*, pode-se opcionalmente arquivar a imagem, em um dispositivo de memória temporário, tal qual a memória RAM do computador, ou em um dispositivo de memória permanente, como por exemplo, um disco rígido ou dispositivo de armazenamento baseado em memória *flash*.

Após, o próximo passo é o *pré-processamento*. Nesta fase, faz-se o melhoramento da imagem, aumentando a chance de sucesso nas etapas seguintes (GONZALEZ, WOODS e EDDINS, 2003). Estão incluídas nesta etapa, tarefas como realce de contraste, remoção de imperfeições ou ruídos, dentre outras.

A *segmentação* divide a imagem de entrada em partes ou regiões, com o objetivo de simplificar a tarefa de processamento através da localização de objetos de interesse e formas na imagem. Como resultado, pode-se obter um conjunto de regiões ou um conjunto de contornos, com alguma característica ou propriedade de interesse para o usuário.

A *descrição*, também denominada seleção de características, extrai características que resultem em informação quantitativa de interesse ou que permitam a discriminação entre classes de objetos, tais como forma e textura. Dessa forma, pode-se efetuar o *reconhecimento* destes objetos, através da atribuição de um rótulo, ou classe, a cada objeto baseado na informação dada pelo descritor. Dessa forma pode-se, por exemplo, identificar caracteres em uma imagem, objetivando reconhecer letras e palavras que ali possam existir.

Estes procedimentos podem ser aplicados em imagens mamográficas para:

- efetuar filtragem, removendo ruídos ou informação desnecessária incorporado no processo de digitalização;
- detecção de objetos, tais como calcificações, massas, distorções arquiteturais e assimetria bilateral;
- classificação de objetos encontrados através de descritores de forma ou textura;
- permitir registro e recuperação baseados em conteúdo, análogo à motores de busca textual.

Desta forma, abre-se um leque de possibilidades no que tange ao emprego do computador e de ferramentas de comunicação na área da saúde, apoiado no uso da informática médica, subsidiando um processo de diagnóstico precoce, capaz de aumentar a expectativa de vida de muitos pacientes.

### **3.3 Informática médica**

Informática Médica pode ser definida como sendo um campo de aplicação do computador, bem como de sistemas de comunicação, como ferramenta de auxílio na área da saúde (SHORTLIFFE e BLOIS, 2001). Esta terminologia é derivada de Computação Médica, cuja terminologia refere-se, inicialmente, à interdisciplinaridade entre ciência da computação e medicina ou ciência da informação médica.

Nas últimas décadas, a medicina vem passando por várias mudanças em todo o mundo, principalmente por conta da disseminação da informação por meio de tecnologias de comunicação e por conta da aplicação de sistemas computacionais que auxiliem no diagnóstico de doenças dos mais variados tipos.

A Internet contribuiu muito para essas mudanças, permitindo que a informação seja compartilhada entre os diversos grupos de pesquisa espalhados pelo mundo, possibilitando a rápida distribuição dos resultados para os colaboradores do estudo e demais interessados.

Vários aplicativos já foram desenvolvidos objetivando registro e recuperação de dados de pacientes, análise de imagens médicas, processamento de sinais biológicos, monitoramento de sinais vitais, dentre outros. Padrões como DICOM (*Digital and*

*Communications in Medicine*) foram estabelecidos para permitir o intercâmbio eletrônico de dados e imagens, independente do equipamento onde ocorreu o processamento ou dos computadores onde serão armazenados os dados. Ferramentas de registro e recuperação de imagens e sinais também foram propostas permitindo que pesquisadores possam criar bases de dados para análise e treinamento de algoritmos, mesmo à distância através da Internet.

A disponibilização de serviços e atividades médicas à distância tem se consolidado como importante recurso aplicado não somente à gestão da saúde, mas também na disponibilização de serviços de auxílio ao diagnóstico e à tomada de decisão. Estes serviços caracterizam, segundo a Organização Mundial de Saúde (OMS), o que se conhece por *Telemedicina* (WHO, 2010).

### 3.3.1 Aplicações da telemedicina

Segundo a OMS, as aplicações de telemedicina podem envolver a troca de informações entre os indivíduos envolvidos, seja de profissionais para profissionais de saúde ou de profissionais de saúde para pacientes, de forma síncrona ou assíncrona (WHO, 2010). A exemplo de transmissão assíncrona de informações, uma equipe médica pode avaliar remotamente um caso clínico em diferentes momentos através de sistemas de informações de saúde. Por outro lado, um sistema de videoconferência pode prover comunicação em tempo real, de maneira síncrona, permitindo que equipes médicas possam discutir e avaliar casos médicos à distância.

As aplicações de telemedicina podem ser classificadas em:

- telemonitoramento;
- teleterapia;
- teledidática;
- telefonia social e
- telediagnóstico.

O telediagnóstico compreende o envio remoto de dados, podendo ser estes imagens médicas ou sinais vitais, bem como exames clínicos ou laboratoriais. Estes dados podem ser analisados por especialistas médicos, permitindo a tomada de decisão diagnóstica. Além disso, os sistemas de telediagnóstico podem fornecer um auxílio ao diagnóstico, uma

espécie de segunda opinião, de forma a facilitar a tomada de decisão por parte do profissional envolvido no caso (SIZILIO, LEITE, *et al.*, 2011).

### 3.3.2 A telemedicina no Brasil

No Brasil, alguns trabalhos envolvendo telemedicina estão sendo desenvolvidos objetivando a implantação de serviços de telediagnóstico, bem como registro e recuperação de imagens médicas (ROSA, ISOLDI, *et al.*, 2006).

Como exemplo, pode ser citado o projeto *HealthNet*, cujo objetivo é dar suporte ao telediagnóstico através de uma segunda opinião médica (BARBOSA, NOVAES, *et al.*, 2001). Fazem parte deste projeto os grupos de pesquisa em telemedicina da Universidade Federal de Pernambuco (UFPE) em parceria com outros centros de pesquisas, bem como com o Hospital das Clínicas da UFPE e o Real Hospital Português (RHP) de Beneficência de Pernambuco, criando uma rede denominada *Rede Integrada de Cooperação em Saúde*.

Este projeto provê a elaboração de um sistema de telediagnóstico, permitindo que médicos residentes em diferentes locais e com recursos limitados possam interagir com médicos especialistas para chegarem a um diagnóstico sobre seus pacientes.

Além de ser uma ferramenta de uso colaborativo, onde os médicos envolvidos podem trocar informações acerca do caso a ser analisado, tais como informações textuais, imagens e vídeos, o sistema faz o registro dos casos em bases de dados, permitindo consultas posteriores.

Também como exemplo de trabalhos envolvendo telemedicina no Brasil, pode-se citar o *Programa Telessaúde Brasil Redes*, um projeto criado em 2007 que objetiva melhorar a qualidade do atendimento no Sistema Único de Saúde (SUS), através da integração de ensino e serviço por meio de ferramentas de tecnologia da informação, promovendo desta forma a *Teleassistência* e a *Teleducação* (SAÚDE, 2012). O Telessaúde Brasil Redes é formado por gestores da saúde, instituições responsáveis pela formação de profissionais em saúde e o serviço de saúde do SUS. Está atualmente implementado em 11 estados e oferece teleconsultorias a aproximadamente 1.500 Unidades Básicas de Saúde.

Entre os serviços oferecidos pelo Telessaúde Brasil Redes, destacam-se:

- teleconsultoria, uma consulta registrada e realizada entre profissionais e gestores da área de saúde, por meio de instrumentos de comunicação bidirecional, para esclarecimento de dúvidas sobre procedimentos clínicos, ações de saúde e questões relativas ao processo de trabalho.
- telediagnóstico, um serviço que utiliza as tecnologias de informação e comunicação para realizar serviços de apoio ao diagnóstico remotamente;
- tele-educação, através de conferências, aulas e cursos ministrados à distância;

Dessa forma, percebe-se que o uso das tecnologias de informação e comunicação na área de saúde, seja para troca de informações ou auxílio ao diagnóstico médico, vem contribuindo para prestar melhores serviços de saúde à distância de forma cooperativa e segura, possibilitando aos especialistas ferramentas de apoio à decisão, capazes de diminuir a probabilidade de erros no diagnóstico e reduzir o tempo de espera na obtenção de uma segunda opinião médica.

## 4 ANÁLISE DE COMPONENTES INDEPENDENTES

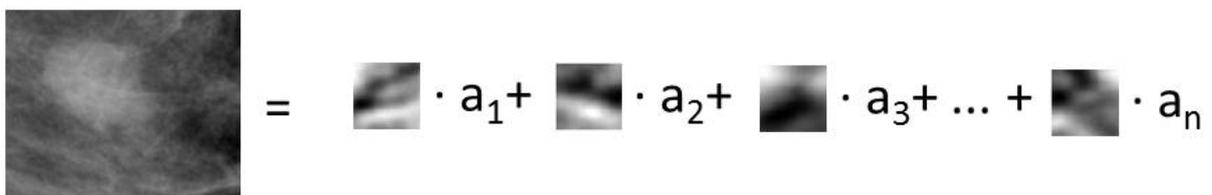
A análise de componentes independentes (ICA, do inglês *Independent Component Analysis*) é um método baseado em estatística computacional desenvolvido, inicialmente, para resolver problemas de separação cega de fontes (BSS, do inglês *Blind Source Separation*) (HYVÄRINEN, KARHUNEN e OJA, 2001). Seu objetivo é separar um conjunto de sinais obtidos a partir de sinais misturados, sem para isso, conhecer alguma informação acerca dos sinais originais, ou fontes, e do processo de mistura (CHOI, CICHOCKI, *et al.*, 2005).

Trata-se de um modelo generativo, pois os sinais misturados são a combinação linear dos sinais originais (componentes independentes) com uma matriz de mistura, possuindo aplicações em diversas áreas: áudio, sinais de SONAR, instrumentação médica, comunicação móvel, engenharia biomédica, dentre outras.

Neste modelo, a análise ou separação das fontes ou sinais originais que são estatisticamente independentes é efetuada a partir de um determinado modelo de mistura das fontes (HYVÄRINEN, KARHUNEN e OJA, 2001).

Outra aplicação para ICA é a extração de características. Em processamento de imagem, as componentes podem fornecer uma representação para uma imagem. Tal representação permite executar tarefas como compressão ou reconhecimento de padrões (HYVÄRINEN, KARHUNEN e OJA, 2001).

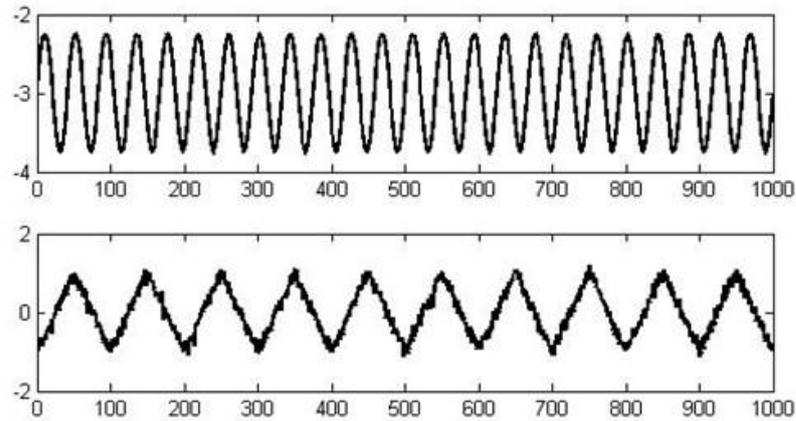
Neste trabalho, uma imagem mamográfica será considerada uma mistura de imagens-base, ou seja, uma combinação linear de algumas imagens-base, comuns a todas as mamografias empregadas na estimação destas imagens. A Figura 10 ilustra a representação de uma imagem mamográfica como combinação linear de suas imagens-base.



$$\text{Imagem Mamográfica} = \text{Imagem}_1 \cdot a_1 + \text{Imagem}_2 \cdot a_2 + \text{Imagem}_3 \cdot a_3 + \dots + \text{Imagem}_n \cdot a_n$$

**Figura 10 - Amostra de uma imagem mamográfica como combinação linear de suas características.**

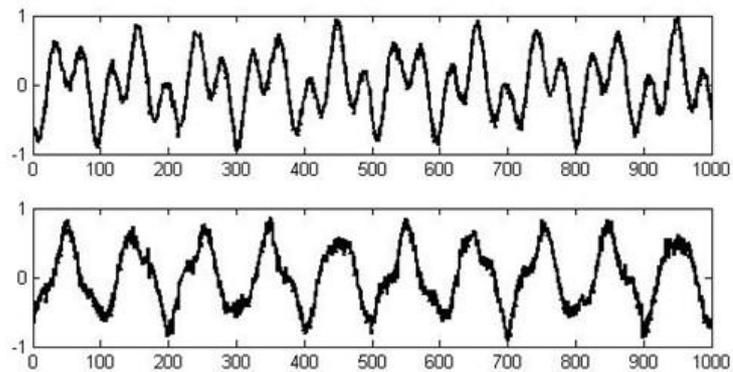
A Figura 11 mostra, a título de exemplo, dois sinais estatisticamente independentes que, depois de misturados, produzem os sinais apresentados na Figura 12.



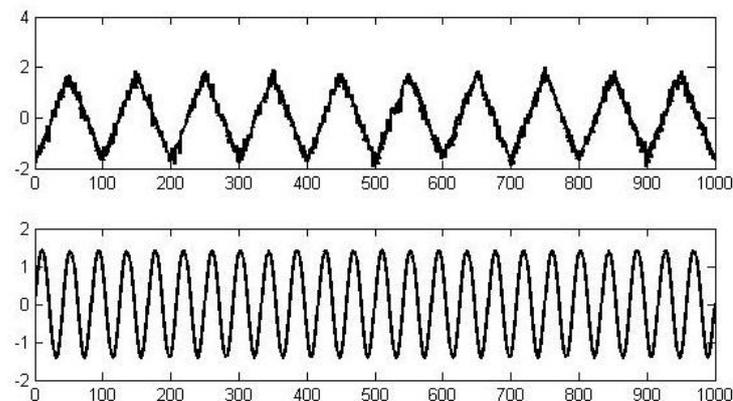
**Figura 11 - Sinais artificiais originais.**

A mistura foi efetuada com a matriz disposta na Eq. (1). Esta matriz foi gerada aleatoriamente, apenas para mostrar o processo de mistura com os sinais originais.

$$A = \begin{bmatrix} -0,6 & 0,5 \\ 0,2 & 0,7 \end{bmatrix} \quad (1)$$



**Figura 12 - Sinais artificiais misturados.**



**Figura 13- Sinais artificiais reconstruídos**

Através da ICA, as fontes puderam ser estimadas, produzindo os sinais apresentados na Figura 13. Convém notar que o sinal e a ordem dos sinais originais estimados foram diferentes dos sinais originais, fato este que será explicado nas seções seguintes.

#### 4.1 Definições

Sejam dadas observações de  $n$  sinais misturados, modelados como combinações lineares de  $n$  funções base

$$x_i = a_{i1}s_1 + a_{i2}s_2 + a_{i3}s_3 + \dots + a_{in}s_n \quad (2)$$

onde cada sinal  $x_n$ , bem como cada componente independente  $s_n$ , é uma variável aleatória.

Utilizando notação matricial, podemos reescrever esta equação da seguinte forma:

$$\mathbf{X} = \mathbf{AS} \quad (3)$$

O modelo apresentado na Eq. (3) é chamado de análise de componentes independentes (HYVÄRINEN e OJA, 2000). Este modelo descreve como os dados são gerados a partir do processo de mistura com as componentes independentes  $s_n$ . O objetivo deste modelo é permitir que se possa estimar a matriz de mistura  $\mathbf{A}$ , bem como a matriz de componentes independentes  $\mathbf{S}$ , somente observando  $\mathbf{X}$ .

A estimação das componentes é baseada em algumas pressuposições, a saber:

- as componentes independentes  $s_i$  são estatisticamente independentes;
- as componentes possuem distribuição não-gaussianas.

O modelo de ICA apresenta, no entanto, algumas ambiguidades no que diz respeito às componentes independentes:

- não se pode determinar suas variâncias;
- não se pode determinar sua ordem.

Tais ambiguidades se devem ao fato de  $\mathbf{A}$  e  $\mathbf{S}$  serem desconhecidas. Como consequências destas ambiguidades, não é possível determinar as energias ou as amplitudes dos sinais, nem tão pouco os sinais ou a ordem de  $s_n$ .

## 4.2 Independência estatística e não-correlação de variáveis

Duas variáveis são consideradas independentes quando o valor de uma não fornece informação acerca do valor da outra. Consideremos duas variáveis  $x_1$  e  $x_2$ . Estas variáveis são ditas independentes se, e somente se,  $x_1$  não fornece nenhuma informação de  $x_2$ , e vice-versa. Matematicamente,

$$p(x_1, x_2) = p(x_1) \cdot p(x_2) \quad (4)$$

ou em outras palavras, a probabilidade conjunta de  $x_1$  e  $x_2$  é igual ao produto das densidades marginais  $p(x_1)$  e  $p(x_2)$ .

Duas variáveis  $x_1$  e  $x_2$  são descorrelacionadas se a sua covariância for igual a zero:

$$cov_{x_1, x_2} = E[(x_1 - \mu_1) \cdot (x_2 - \mu_2)] = 0 \quad (5)$$

sendo  $\mu_1$  e  $\mu_2$  as médias das variáveis  $x_1$  e  $x_2$ , respectivamente.

## 4.3 Estimação das componentes independentes

A estimação das componentes independentes  $s_n$  pode ser obtida através da matriz de mistura  $\mathbf{A}$ , da seguinte forma:

$$\mathbf{S} = \mathbf{A}^{-1}\mathbf{X} \quad (6)$$

Sendo a matriz  $\mathbf{A}$  desconhecida, a ideia principal por trás da análise de componentes independentes consiste em considerar que os sinais observáveis  $x_n$  estão relacionados com os sinais originais, através de uma transformação linear. Assim, os sinais originais podem ser obtidos a partir de uma transformação inversa. Supondo, dessa forma, uma combinação linear de  $x_i$ , de modo que:

$$\mathbf{y} = \mathbf{b}^T\mathbf{X} \quad (7)$$

sendo  $\mathbf{X} = \mathbf{AS}$ , pode-se escrever:

$$\mathbf{y} = \mathbf{b}^T\mathbf{AS} \quad (8)$$

onde  $b$  deve ser determinado. A partir da Eq. (8), pode-se observar que  $\mathbf{y}$  é uma combinação linear se  $s_i$ , com coeficientes dados por  $\mathbf{q} = \mathbf{b}^T \mathbf{A}$ . Desta forma, obtemos:

$$\mathbf{y} = \mathbf{q}^T \mathbf{S} \quad (9)$$

Se  $\mathbf{b}$  corresponde a uma das linhas da inversa de  $\mathbf{A}$ , então  $\mathbf{y}$  será uma das componentes independente, e neste caso, apenas um dos elementos de  $\mathbf{q}$  será igual a um, e todos os outros serão iguais a zero. No entanto, sendo  $\mathbf{X}$  conhecido,  $\mathbf{b}$  não pode ser determinado exatamente, porém pode-se estimar seu valor.

Uma forma de determinar  $\mathbf{b}$  é variar os coeficientes em  $\mathbf{q}$  e verificar como a distribuição de  $\mathbf{y} = \mathbf{q}^T \mathbf{S}$  muda. Como pelo Teorema do Limite Central (PAPOULIS e PILLAI, 2002), a soma de duas variáveis aleatórias independentes é mais gaussiana que as variáveis originais,  $\mathbf{y}$  é mais gaussiana que qualquer uma das  $s_i$  e menos gaussiana quando se iguala a umas das  $s_i$ . Assim, apenas um elemento  $q_i$  de  $\mathbf{q}$  é diferente de zero (HYVÄRINEN, KARHUNEN e OJA, 2001). Como, na prática, os valores de  $\mathbf{q}$  são desconhecidos, e através da Eq. (7) e da Eq. (9) tem-se que

$$\mathbf{b}^T \mathbf{X} = \mathbf{q}^T \mathbf{S} \quad (10)$$

pode-se variar  $\mathbf{b}$  e observar a distribuição de  $\mathbf{b}^T \mathbf{X}$ .

Dessa forma, pode-se tomar como  $\mathbf{b}$  um vetor que maximiza a não-gaussianidade de  $\mathbf{b}^T \mathbf{X}$ , sendo  $\mathbf{q} = \mathbf{A}^T \mathbf{S}$ , contendo apenas uma de suas componentes diferente de zero. Isso significa que  $\mathbf{y}$  na Eq. (7) é igual a uma das componentes independentes, e a maximização da não-gaussianidade de  $\mathbf{b}^T \mathbf{X}$ , permite encontrar uma das componentes.

#### 4.4 Negentropia como medida de não-gaussianidade

A entropia de uma variável aleatória está relacionada com a quantidade de informação que essa variável possui, sendo maior quanto mais for imprevisível a variável. Em se tratando de variáveis aleatórias, é denominada entropia diferencial. Se  $\mathbf{y}$  é um vetor aleatório com função densidade de probabilidade  $f(\mathbf{y})$ , a sua entropia diferencial é dada por:

$$H(\mathbf{y}) = - \int f(\mathbf{y}) \log f(\mathbf{y}) d\mathbf{y} \quad (11)$$

Sabendo-se que uma variável gaussiana tem a maior entropia dentre todas as variáveis aleatórias de igual variância (HYVÄRINEN, KARHUNEN e OJA, 2001) (PAPOULIS e PILLAI, 2002), tem-se que uma versão modificada da entropia diferencial pode ser usada como medida de não-gaussianidade. Tal medida é denominada negentropia, definida por

$$H(y) = H(y_{gauss}) - H(y) \quad (12)$$

sendo  $y_{gauss}$ , uma variável aleatória de mesma matriz de covariância que  $y$ . A negentropia é sempre não-negativa, e pode assumir zero se, e somente se,  $y$  tem distribuição gaussiana e é invariante para transformações lineares inversíveis.

Apesar de permitir que se possa medir não-gaussianidade, a negentropia é de difícil estimação, sendo necessária sua estimação por aproximações através de momentos de alta ordem. Assim,

$$J(y) \approx \frac{1}{12} E\{y^3\}^2 + \frac{1}{48} kurt(y)^2 \quad (13)$$

sendo  $kurt(y)$  a curtose de  $y$ , definida como o momento de quarta ordem da variável aleatória  $y$ , expressa por

$$kurt(y) = E\{y^4\} - 3(E\{y^2\})^2 \quad (14)$$

## 5 MÁQUINA DE VETOR DE SUPORTE

A máquina de vetores de suporte (SVM, do inglês *Support Vector Machine*) é uma máquina de aprendizagem supervisionada capaz de classificar a partir de  $n$  indivíduos observados pertencentes a diversos subgrupos, a que classe um indivíduo que deve ser classificado pertence (BISHOP, 2006).

Para simplificar, será considerado o caso de classificação utilizando duas classes usando um modelo linear na forma

$$y(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b = 0 \quad (15)$$

onde  $\mathbf{w}$  é um vetor de pesos ajustados,  $b$  é um viés e  $\mathbf{x}$  é um vetor de treinamento de características, com seus respectivos rótulos  $y_i \in Y$ , em que  $Y = \{-1, +1\}$ . O modelo definido na Eq. (15) define um hiperplano ótimo que classifica todos os vetores de treinamento, e  $\mathbf{x}$  é dito linearmente separável se é possível separar os dados das classes  $-1$  e  $+1$  por este hiperplano (SMOLA e SCHÖLKOPF, 2002). A Figura 14 ilustra a ideia de um hiperplano ótimo para padrões linearmente separáveis.

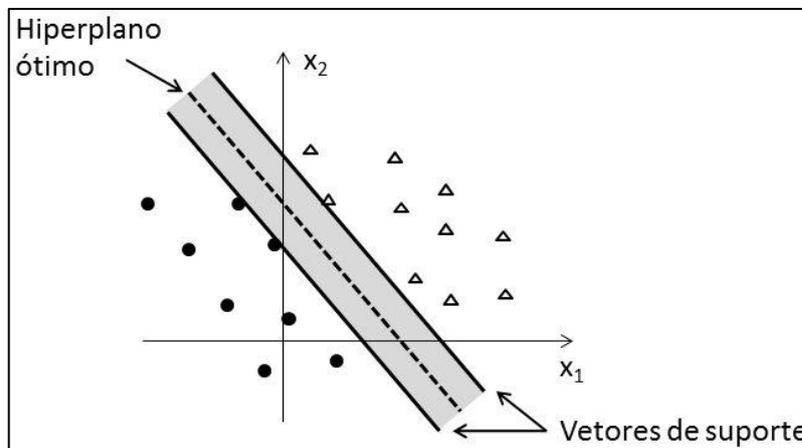


Figura 14 - Um conjunto de treinamento de duas classes, os vetores de suporte e o hiperplano ótimo.

De acordo com este modelo,

$$\mathbf{w}^T \mathbf{x} + b \geq 0 \quad \text{para } y(\mathbf{x}) = +1 \quad (16)$$

$$\mathbf{w}^T \mathbf{x} + b < 0 \quad \text{para } y(\mathbf{x}) = -1 \quad (17)$$

Seja  $x_1$  um ponto pertencente ao hiperplano  $H_1: \mathbf{w}^T \mathbf{x} + b = +1$  e  $x_2$  um ponto pertencente ao hiperplano  $H_2: \mathbf{w}^T \mathbf{x} + b = -1$ , conforme ilustrado na Figura 15.

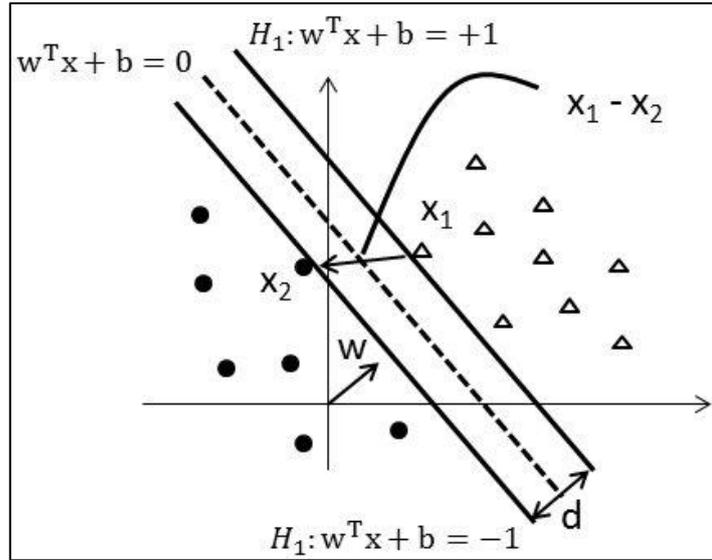


Figura 15 - Distância entre os hiperplanos  $H_1$  e  $H_2$ .

Ao se projetar  $x_1 - x_2$  na direção de  $w$ , perpendicular ao hiperplano ótimo, é possível obter a distância entre os hiperplanos  $H_1$  e  $H_2$ , dada pela (18).

$$(x_1 - x_2) \left( \frac{w}{\|w\|} \cdot \frac{(x_1 - x_2)}{\|x_1 - x_2\|} \right) \quad (18)$$

De acordo com a Eq. (16) e com a Eq. (17), tem-se que  $w^T x_1 + b = +1$  e  $w^T x_2 + b = -1$ . A diferença entre essas equações nos dá  $w \cdot (x_1 - x_2) = 2$ . Substituindo esse resultado na Eq. (18), tem-se:

$$\frac{2(x_1 - x_2)}{\|w\| \|x_1 - x_2\|} \quad (19)$$

Tomando-se a norma da Eq. (19), obtém-se:

$$\frac{2}{\|w\|} \quad (20)$$

Esta é a distância  $d$ , ilustrada na Figura 15, entre os hiperplanos  $H_1$  e  $H_2$ , paralelos ao hiperplano ótimo separados. Minimizando  $\|w\|$ , podemos maximizar a margem de separação dos dados em relação a  $w^T x + b = 0$ . Assim, tem-se o seguinte problema de otimização:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad (21)$$

Com as restrições:  $y_i(w^T x_i + b) - 1 \geq 0, \forall i = 1, 2, \dots, n$

Estas restrições são impostas para assegurar que não haja dados de treinamento entre as margens de separação das classes. Este é um problema de otimização quadrático, cuja função objetiva é convexa e os pontos que satisfazem as restrições formam um conjunto convexo, logo possui um único mínimo global.

Para solucionar este problema, aplica-se um operador Lagrangiano, capaz de englobar as restrições às funções objetivo, associadas aos multiplicadores de Lagrange, conforme Eq. (22).

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i (y_i (w^T x_i + b) - 1) \quad (22)$$

onde o operador Lagrangiano deve ser minimizado, implicando na maximização das variáveis  $\alpha_i$  e na minimização de  $w$  e  $b$ .

Igualando as derivadas de L em relação a  $b$  e a  $w$  a zero, obtemos as condições:

$$w = \sum_{i=1}^n \alpha_i y_i x_i \quad (23)$$

$$\sum_{i=1}^n \alpha_i y_i = 0 \quad (24)$$

Substituindo as equações (23) e (24) na equação (22), temos o seguinte problema de otimização exposto nas equações (25) e (26).

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (25)$$

$$\text{sujeito a: } \begin{cases} \alpha_i \geq 0, \forall i = 1, \dots, n \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{cases} \quad (26)$$

onde  $(x_i \cdot x_j)$  corresponde ao produto interno entre  $x_i$  e  $x_j$ .

Em se tratando de padrões não separáveis linearmente, não é possível construir um hiperplano ótimo de separação sem encontrar erros de classificação (HAYKIN, 1999). Para o caso de pontos de dados não separáveis, é introduzido um conjunto de variáveis escalares não negativas,  $\xi_i$ , na definição do hiperplano de separação:

$$d_i(w^T x_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, n \quad (27)$$

As variáveis  $\xi_i$  são chamadas variáveis de folga e medem o desvio de um ponto de dado da condição ideal de separabilidade de padrões. Dessa forma, o problema agora é:

$$\min_{w, \xi} \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i \quad (28)$$

$$\text{sujeito a: } y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i = 1, \dots, n \quad (29)$$

Novamente, trata-se de um problema de otimização quadrático, com as restrições lineares dadas pela Eq. (29). Aplicando o operador Lagrangiano e tornando suas derivadas parciais iguais a zero, tem-se o seguinte problema de otimização:

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (30)$$

sujeito às restrições

$$\begin{cases} 0 \leq \alpha_i \leq C, \quad \forall i = 1, \dots, n \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{cases} \quad (31)$$

onde  $C$  é um parâmetro positivo especificado pelo usuário.

Os problemas não lineares de classificação são resolvidos mapeando o conjunto de treinamento, saindo de seu espaço de entrada para um novo espaço com maior dimensão, denominado espaço de características. Seja  $\Phi(x)$ , uma função que mapeia o espaço de entrada sobre o espaço de características. Esta função é usada para mapear  $x_i$  e  $x_j$  para o espaço de características, antes da realização do produto interno entre eles:

$$k(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) \quad (32)$$

modificando, desta forma, o problema de maximização proposto na Eq. (32), para:

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j k(x_i, x_j) \quad (33)$$

sob as restrições da Eq. (33). A função definida na Eq. (32) é chamada de função de núcleo (*kernel*).

A **Erro! Fonte de referência não encontrada.** apresenta as funções de núcleo ara três tipos bem comuns de máquina de vetor de suporte.

**Tabela 1 - Funções de núcleo mais comuns.**

<b>Núcleo</b>	<b>Função <math>k(x_i, x_j)</math></b>	<b>Parâmetros</b>
Polinomial	$\delta(x_i \cdot x_j + k)^d$	$\delta, k$ e $d$
Função de base radial	$e^{-\gamma \ x_i - x_j\ ^2}$	$\gamma$
Sigmoidal	$\tanh(\delta(x_i, x_j) + k)$	$\delta$ e $k$

Os parâmetros utilizados nas funções de núcleo expostos na Tabela 1 **Erro! Fonte e referência não encontrada.** são definidos *a priori* pelo usuário.

## 6 SISTEMA PROPOSTO

O processo de construção de um sistema é uma atividade da Engenharia de Software, sendo necessário um conjunto de métodos e técnicas para a construção do produto final: o software. Neste capítulo serão descritos os métodos e as técnicas empregadas na construção do sistema.

Um sistema é uma coleção de componentes inter-relacionados capazes de trabalhar juntos para atingir alguns objetivos (SOMMERVILLE, 2008). Em se tratando de sistemas de software, estes componentes compõem os programas, arquivos de configuração e documentação. Qualquer software, ao ser concebido, passa por uma série de etapas, uma espécie de roteiro denominado *processo de software*. Através destes passos, os engenheiros e desenvolvedores podem controlar e organizar cada etapa, sem perder o foco naquilo que deve ser desenvolvido (PRESSMAN, 2002).

Diversos modelos de processo de software foram desenvolvidos para dar suporte à tarefa de elaboração de sistemas de software. Dentre estes, o *modelo cascata*, também chamado de *ciclo de vida clássico*, é o modelo mais antigo e talvez o mais simples de se aplicar, principalmente em ocasiões em que os requisitos de um problema estão razoavelmente bem compreendidos ou quando o trabalho flui da comunicação até a implantação de modo linear (PRESSMAN, 2002).

O modelo em cascata recebe este nome em virtude das atividades envolvidas na elaboração do software serem sequenciais, e cada atividade depender da execução da atividade anterior. A figura mostra as etapas deste modelo.

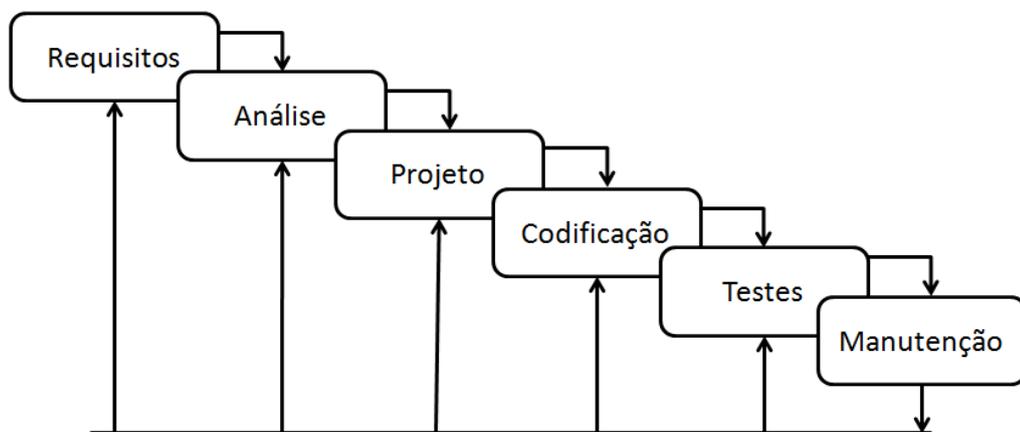


Figura 16 - Atividades do processo de software no modelo cascata. Adaptado de (PRESSMAN, 2002).

As seções seguintes irão abordar a construção do sistema utilizando o modelo em cascata. A escolha por esse modelo se deu em função de sua simplicidade.

## 6.1 Visão geral do sistema

O sistema de software proposto neste trabalho é um tipo de sistema de telediagnóstico, capaz de analisar remotamente imagens de mamografia digital enviadas pela Internet e fornecer um diagnóstico da imagem, indicando a presença de regiões suspeitas, que poderão ser confirmadas, posteriormente, por um especialista em imagens mamográficas.

O sistema foi projetado para ser um WebIS (*Web based Information System*). A Figura 17 apresenta um esquema resumido do funcionamento do sistema proposto. De um lado, tem-se o usuário que, pela Internet, enviará os dados do paciente juntamente com a imagem da mamografia digital a ser analisada. Do outro lado, um usuário especialista que irá conferir o resultado da análise feita pelo sistema, confirmando ou não o resultado do pré-diagnóstico gerado de forma automática. Entre essas partes, encontra-se o sistema que receberá os dados a serem processados, fornecerá o pré-diagnóstico ao especialista, e dará o diagnóstico final ao usuário solicitante.

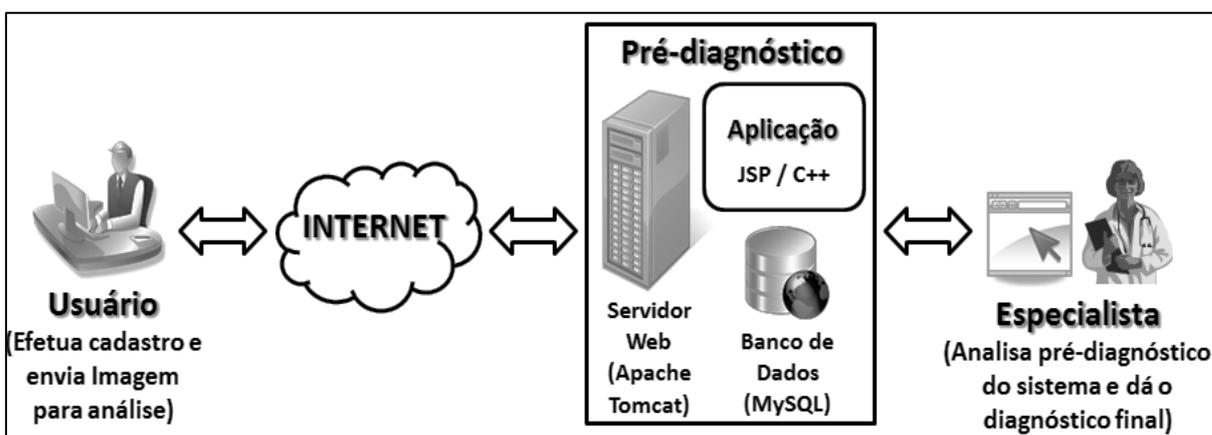
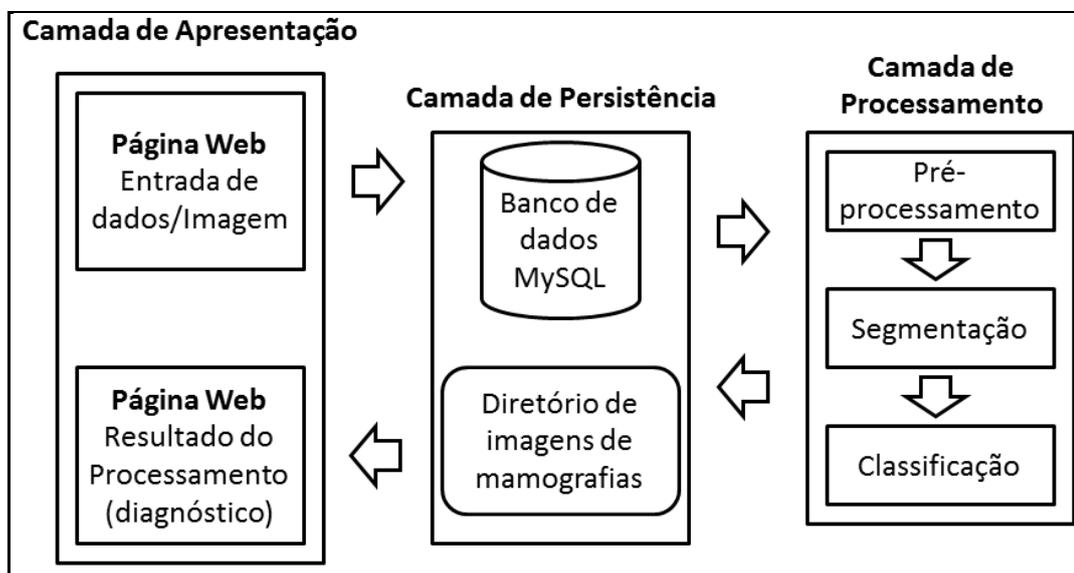


Figura 17 - Visão geral do sistema proposto.

Na Figura 18 pode-se observar um esquema da arquitetura do sistema. A transferência e a visualização das imagens são de responsabilidade da camada de apresentação. Nesta camada as imagens de mamografia digital são obtidas e transferidas juntamente com alguns dados clínicos do paciente para o servidor Web, onde serão armazenadas, juntamente com dados anônimos do paciente, para o posterior processamento. No servidor Web, os dados do paciente são gravados em um banco de dados enquanto que a

imagem é armazenada em uma estrutura de diretórios em disco, com suas referências sendo gravadas no banco junto com o caso clínico.



**Figura 18 - Arquitetura do sistema proposto.**

A etapa de processamento entra em operação efetuando a análise da imagem e gerando o pré-diagnóstico que será validado pelo especialista. Uma vez que os dados clínicos e a imagem a ser processada estejam no servidor, uma *thread* responsável por verificar se novas informações estão disponíveis para análise dispara um conjunto de rotinas que irão efetuar o processamento da imagem. Optou-se por implementar a rotina de verificação e processamento em uma *thread*, em virtude da possibilidade de instanciar várias *threads*, caso haja a necessidade de analisar várias imagens ao mesmo tempo, tornando a aplicação escalável.

Após a etapa final do processamento, serão geradas as imagens que o especialista irá avaliar, juntamente com os dados clínicos, para confirmação ou não do pré-diagnóstico gerado pelo sistema. Com o aval do especialista o sistema gera uma página Web contendo o resultado da análise feita sobre a mamografia, de forma que o usuário que enviou os dados para processamento possa ter acesso ao diagnóstico.

Este processo é ilustrado no fluxograma da Figura 19.

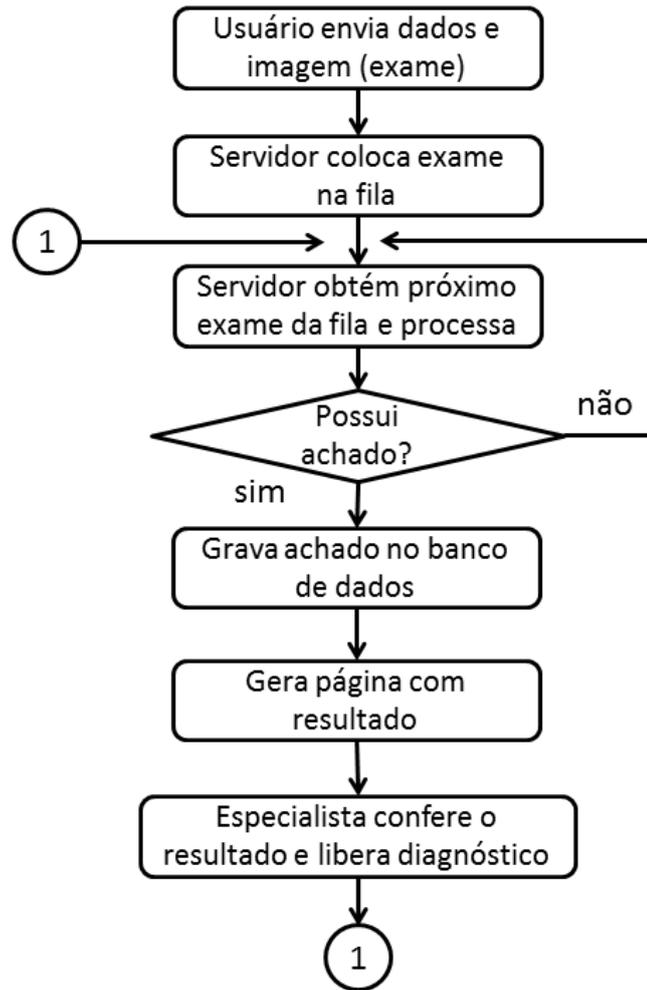


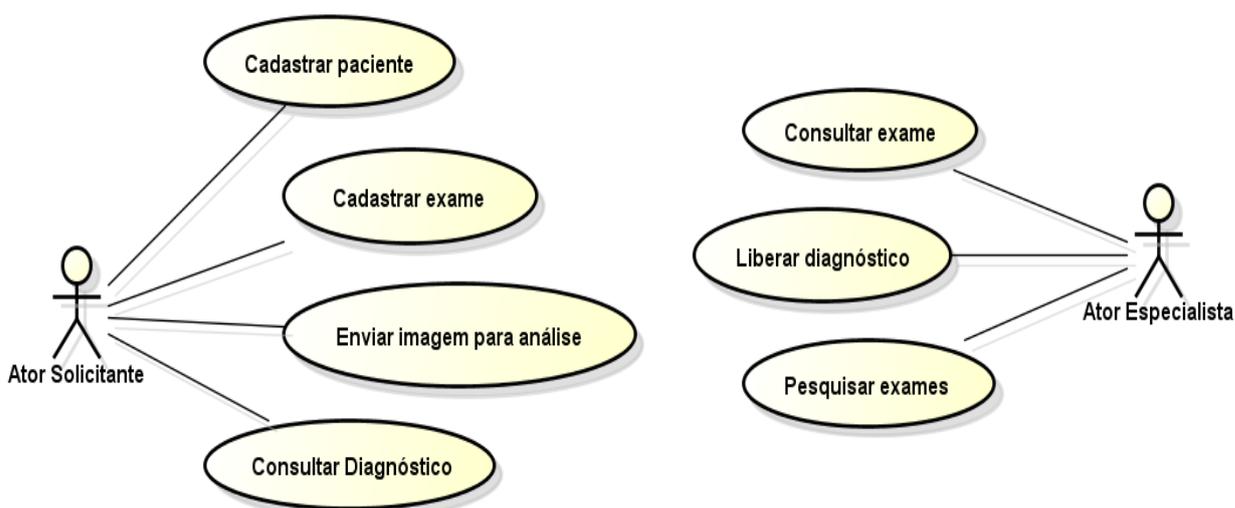
Figura 19 - Fluxograma do sistema proposto.

## 6.2 Modelagem e desenvolvimento do sistema

A primeira fase na elaboração de um sistema no modelo cascata, após se ter uma ideia geral de como este deve funcionar e que resultados deve produzir, diz respeito ao *levantamento de requisitos*. Nesta etapa são levantados os requisitos que o produto deve atender, tanto *funcionais* quanto *não funcionais* (SOMMERVILLE, 2008). Os requisitos funcionais são exatamente as funções que o sistema deve fornecer, ou seja, como deve reagir às entradas específicas do usuário e como deve se comportar em determinadas situações. Os requisitos não funcionais dizem respeito a tarefas intrínsecas a qualquer software desenvolvido com qualidade: desempenho, usabilidade, confiabilidade, segurança, disponibilidade e tecnologias envolvidas (PRESSMAN, 2002).

Para o sistema proposto, foram levantados seus requisitos funcionais, e a partir destes, foram elaborados os *casos de uso*. Os casos de uso definirão as funções da aplicação

que o sistema deverá oferecer para o usuário (JACOBSON, CHRISTERSON, *et al.*, 1992). Na Figura 20 é apresentado o diagrama de casos de uso, desenvolvido utilizando a linguagem de modelagem UML (*Unified Modeling Language*). Neste diagrama, os casos de uso correspondem aos balões, contendo cada uma das atividades que serão executadas pelos atores no sistema, aqui representados pelo *Ator Solicitante* e pelo *Ator Especialista*, ambos representando respectivamente, o usuário que requisita o processamento de uma imagem mamográfica e o usuário que libera o diagnóstico a partir do processamento executado pela aplicação.



**Figura 20 - Diagrama de casos de uso do sistema proposto (requisitos funcionais).**

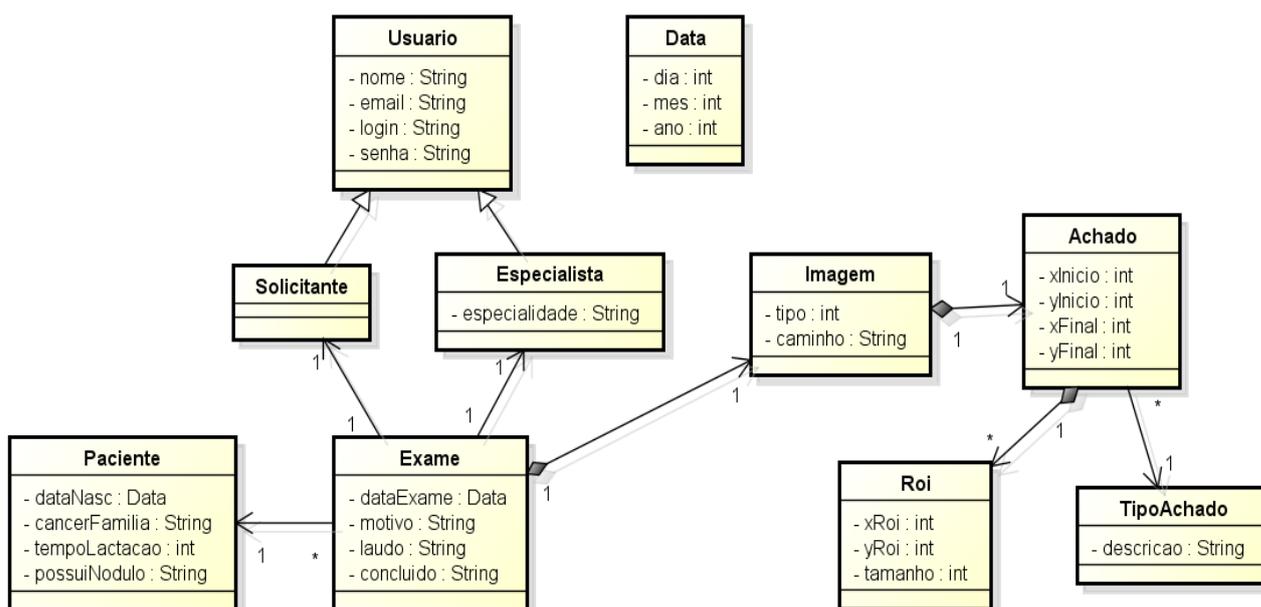
A análise de requisitos e a elucidação dos casos de uso são importantes etapas iniciais no desenvolvimento de qualquer software. Após essas etapas, os desenvolvedores podem prosseguir com o desenvolvimento do sistema.

Após o levantamento dos requisitos e sua formalização através do diagrama de casos de uso em UML, as etapas seguintes são a análise e o projeto do sistema. É nestas etapas, também chamadas de *modelagem do sistema*, que são desenvolvidos os modelos, capazes de esclarecer *como* o sistema funcionará para atender aos requisitos, mesmo antes do sistema estar pronto.

Neste trabalho, o processo de desenvolvimento será orientado a objetos, paradigma de programação baseado na composição e interação de unidades diversas denominadas objetos. A funcionalidade externa de um sistema se dá através da colaboração entre os objetos envolvidos na aplicação (BEZERRA, 2001). Os objetos colaboram entre si

para produzir resultados visíveis ao mundo exterior. A descrição estrutural da colaboração entre os objetos do sistema é feita através do diagrama de classes, pois são as classes que definem estes objetos.

Através da Figura 21, pode-se ver o diagrama de classes, utilizado para explicitar quais os tipos de objetos que o sistema irá manipular, bem como suas características e funcionalidades. Este diagrama foi desenvolvido na linguagem UML.



**Figura 21 - Diagrama de classes do sistema proposto.**

Como se pode observar na figura, foi proposto um conjunto mínimo de classes de modo a contextualizar o processamento de imagens mamográficas na aplicação. A ideia é que o sistema gere não somente os dados anônimos do paciente, incluindo o usuário solicitante, o usuário especialista e o arquivo de imagem a ser analisado, mas também os achados radiológicos encontrados durante o processamento da imagem.

O armazenamento dos dados enviados para a aplicação é feito através de um Sistema de Gerenciamento de Banco de Dados Relacional (SGBDR). Optou-se por utilizar o MySQL, em virtude de ser um SGBDR livre, prático e de fácil gerenciamento. No entanto, a aplicação desenvolvida neste trabalho poderá, com algumas alterações, funcionar com outros sistemas de bancos de dados.

Uma das tarefas essenciais a ser desenvolvida ainda no processo de análise da aplicação é a modelagem do banco de dados da aplicação, onde serão descritas as entidades

do sistema, seus atributos e os relacionamentos entre estas. O *modelo conceitual* do banco de dados da aplicação proposta neste trabalho pode ser visualizado na Figura 22.

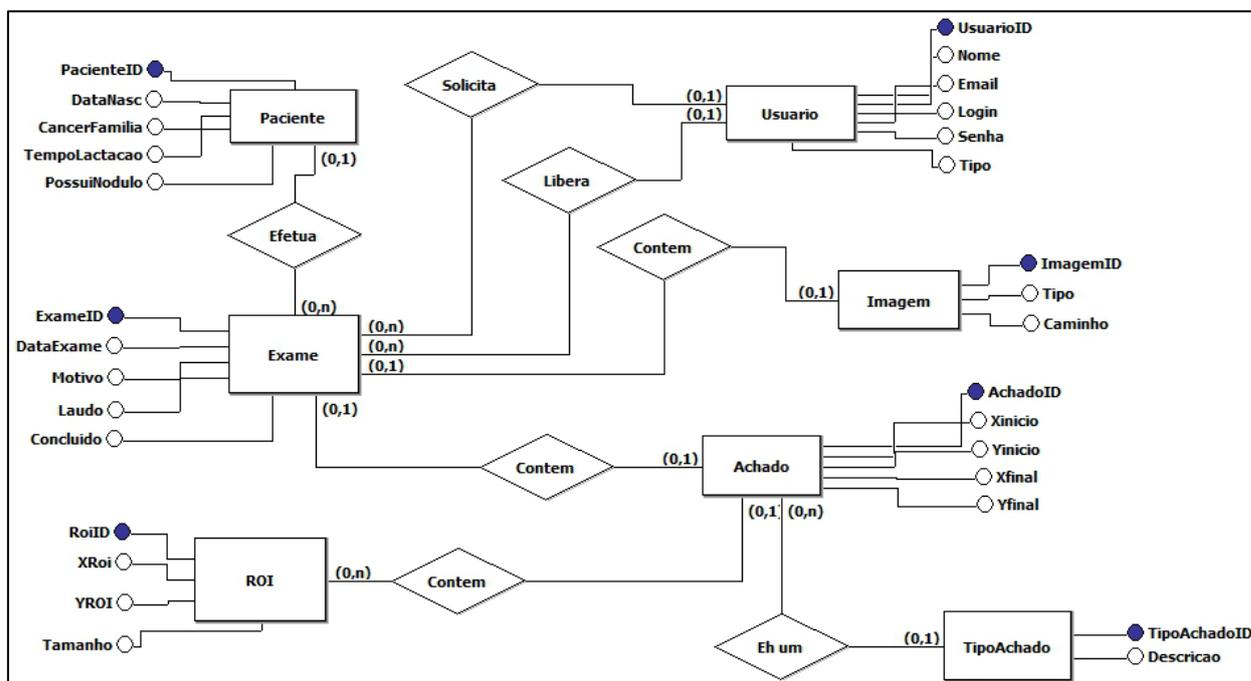


Figura 22 - Modelo conceitual do banco de dados do sistema proposto.

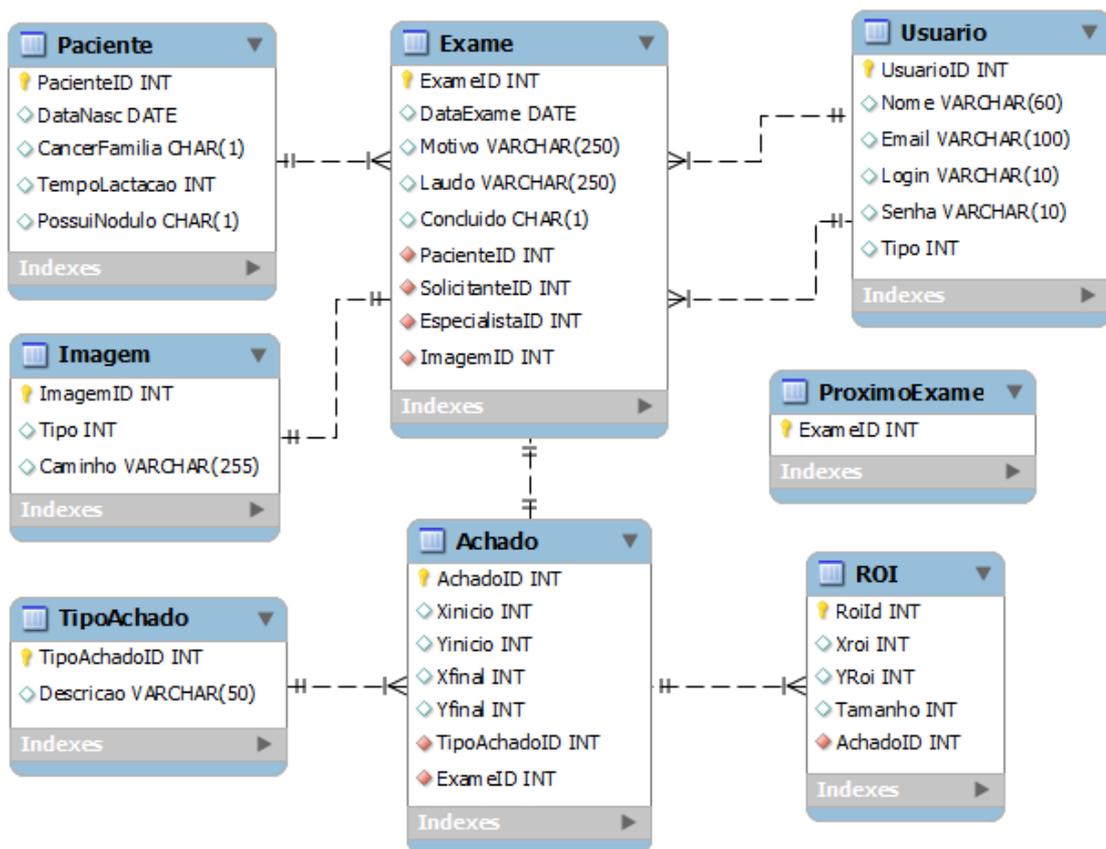
O modelo conceitual é frequentemente utilizado para descrever aplicações de bancos de dados. O resultado deste modelo é o diagrama Entidade-Relacionamento (ER). Neste diagrama são descritas as *entidades*, seus relacionamentos e atributos. As entidades correspondem aos objetos do mundo real, enquanto que seus atributos correspondem às características ou propriedades de cada entidade (ELMASRI e NAVATHE, 2004).

No diagrama ER, as entidades são representadas pelos retângulos com seus respectivos nomes dentro, enquanto que os atributos são representados por pequenos círculos ligados por linhas às entidades e seus respectivos nomes ao lado. As entidades servem para efetuar a *persistência* dos dados armazenados temporariamente nos objetos criados na aplicação. Os relacionamentos são as ligações entre as entidades, representadas por losangos com seu nome ou descrição em seu interior. Estes relacionamentos correspondem a associações que as diversas entidades possuem dentro da aplicação.

Como exemplo, a entidade *Paciente* está relacionada com a classe *Exame*, cujo nome da associação é *Efetua*. Esta associação implica dizer, na prática, que no mundo real um paciente efetua um ou vários exames, e que cada exame é efetuado por um paciente. Esta

quantificação é denominada cardinalidade, e serve para indicar o grau de associação entre as entidades: um-para-um, um-para-vários e vários-para-vários. No caso de *Paciente* e *Exame*, a cardinalidade é *um-para-vários*, implicando em dizer que um paciente está associado com vários exames. Esta cardinalidade entre as entidades *Paciente* e *Exame* pode ser vista na Figura 22, representada no diagrama pelas definições “(0,1)” ao lado da entidade *Paciente* e “(0,n)” ao lado da entidade *Exame*.

A partir da modelagem conceitual do banco de dados, foi possível elaborar o *modelo lógico*, responsável por apresentar as estruturas de armazenamento do banco de dados, denominadas *tabelas*, seus campos e os relacionamentos entre estas. A Figura 23 apresenta este modelo, onde é possível perceber as informações que serão cadastradas no banco de dados.



**Figura 23 - Modelo lógico do banco de dados do sistema proposto.**

Após a etapa de modelagem, a próxima etapa é a implementação do sistema. É na implementação que ocorre a codificação do projeto em uma linguagem ou plataforma de programação, a geração dos arquivos de configuração, a criação das interfaces de usuários,

dentre outras tarefas que irão permitir a construção do software. A próxima seção dará detalhes da implementação, indicando as ferramentas, linguagens e recursos necessários para a construção do protótipo do sistema.

### 6.3 Implementação do sistema

Para a construção do sistema, desde a geração dos arquivos de configuração até a elaboração das interfaces de usuário, foi necessária a integração de várias ferramentas. O protótipo do sistema foi desenvolvido na plataforma JEE 6.0 (*Java Enterprise Edition*) com alguns módulos desenvolvidos em C++, utilizando a biblioteca de processamento de imagem OpenCV (BRADSKI, 2000).

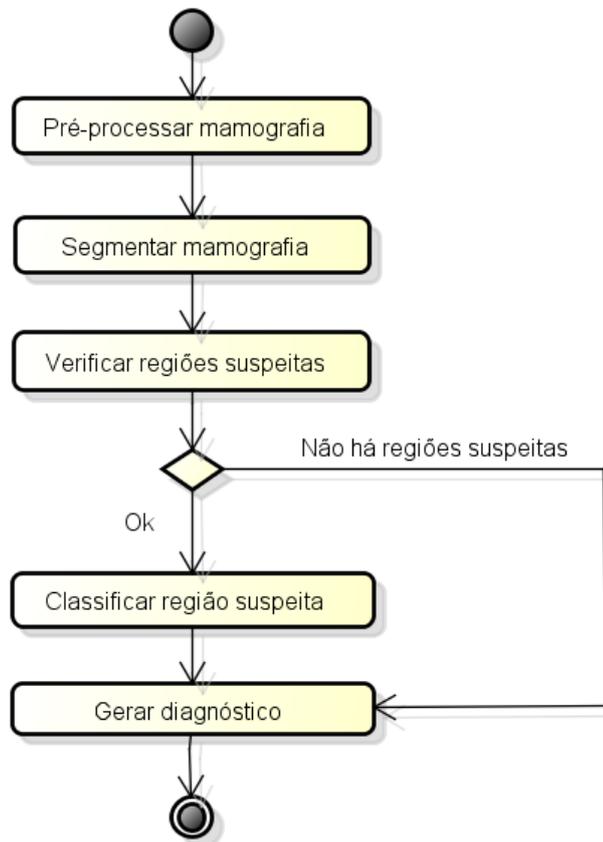
Pelo fato do sistema proposto basear-se na Web, a interface com o usuário é composta de páginas Web, desenvolvidas em HTML (*Hyper Text Markup Language*), a linguagem de marcação utilizada para desenvolver as páginas Web que servirão de interface de usuários, que permitirão a captação dos dados e da imagem mamográfica do paciente, bem como a visualização do resultado do processamento. Além disso, empregou-se JSP (*Java Server Pages*), uma tecnologia de elaboração de páginas Web dinâmicas baseada na plataforma/linguagem de programação Java. A combinação HTML/JSP permitiu a elaboração das *telas* do sistema, que compõem a camada de apresentação, apresentada na Figura 18.

A camada de persistência, responsável por permitir a comunicação com o banco de dados da aplicação, lendo e gravando dados, foi desenvolvida em Java. Optou-se por usar essa linguagem de programação por conta de sua portabilidade entre sistemas operacionais e plataformas de hardware, bem como a facilidade de integração com os vários sistemas de bancos de dados disponíveis no mercado.

Na camada de processamento, foram desenvolvidos programas em C++ juntamente com a biblioteca OpenCV. Estas rotinas efetuam o processamento das imagens, composta pelas tarefas de pré-processamento, segmentação e classificação das regiões segmentadas em *normais* ou *suspeitas*. A seção seguinte aborda em detalhes a camada responsável pelo processamento das imagens mamográficas.

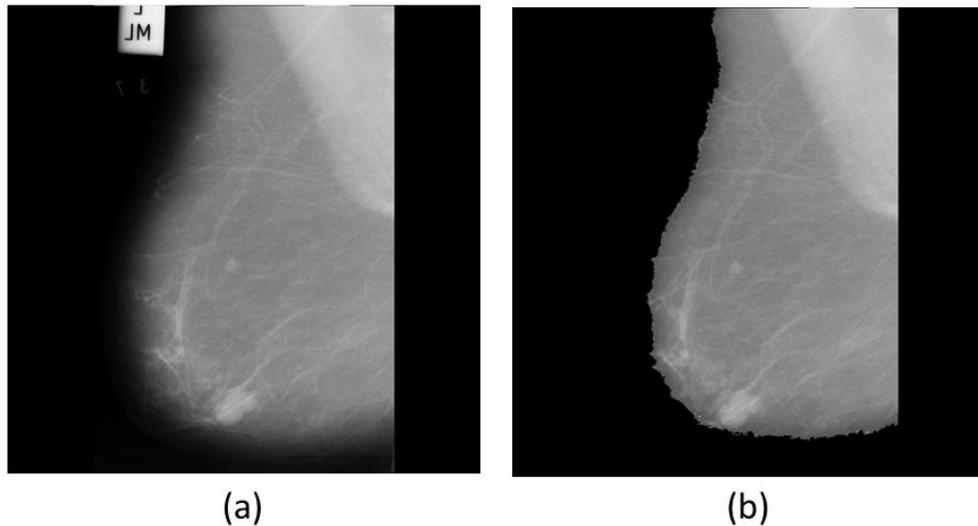
### 6.3.1 Camada de processamento de imagens

A Figura 24 apresenta um diagrama de atividades, em UML, que mostra o funcionamento do módulo de processamento das imagens mamográficas.



**Figura 24 - Diagrama de atividades do módulo de processamento das imagens.**

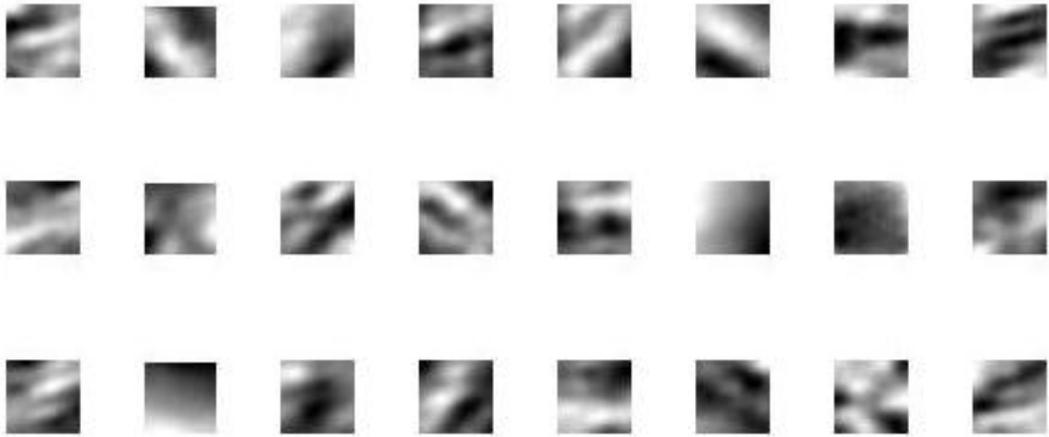
O primeiro programa do módulo de processamento das imagens mamográficas é responsável pela tarefa de *pré-processamento*. Nesta tarefa, são retirados da imagem os artefatos que não fazem parte da região da mama, reduzindo o tempo de processamento, já que tais artefatos não devem ser analisados. A Figura 25 mostra o resultado gerado pelo pré-processamento em uma das imagens analisadas nos testes efetuados.



**Figura 25 - Resultado do pré-processamento.**  
(a) Imagem original. (b) Imagem resultante depois de retirados os artefatos.

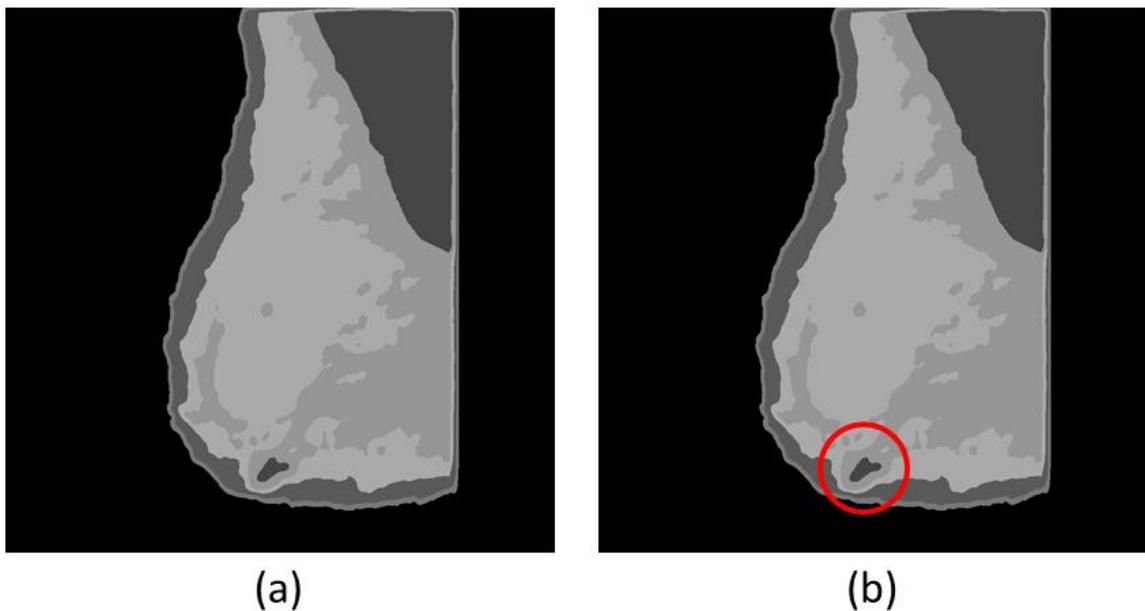
A etapa seguinte compreende a *segmentação* da imagem pré-processada. Esta etapa objetiva localizar os objetos de interesse, neste caso regiões suspeitas, nas imagens mamográficas. Dentre os vários algoritmos de segmentação existentes, o algoritmo utilizado neste trabalho faz uso de uma técnica conhecida como *filtragem* e em seguida outra técnica chamada agrupamento (*clustering*). A filtragem utilizada objetiva realçar a região de interesse (ROI) aplicando uma operação de convolução, usando como máscara, funções base de regiões selecionadas da base de dados mini-MIAS (DUAN, MA, *et al.*, 2009).

Para esta etapa, efetuou-se primeiramente a aquisição das imagens mamográficas a partir da base de dados mini-MIAS, e a seleção manual das regiões de interesse correspondente a tecidos lesionados, a partir de 93 imagens mamográficas. A partir desta seleção, foi criado um banco de imagens com estas regiões de interesse para treinamento do algoritmo. Todas as imagens selecionadas foram redimensionadas para 32x32 *pixels*, objetivando tornar todas do mesmo tamanho. A partir deste banco de treino, foram obtidas 24 funções base das regiões lesionadas, utilizando o algoritmo *FastICA* (HYVÄRINEN, 1999), baseado em análise de componentes independentes. A Figura 26 mostra as funções base extraídas.



**Figura 26 - Funções bases extraídas de regiões de interesses com lesões.**

De posse destas funções base, foi gerado um arquivo de configuração utilizado pelo programa de segmentação. O programa recebe como entrada a imagem pré-processada, efetua a filtragem de convolução usando como máscara as funções base selecionadas no processo de treinamento, ou aprendizagem. Após esta operação, caso a imagem pré-processada apresente regiões suspeitas, estas serão realçadas e utilizando o algoritmo de agrupamento denominado *Kmeans*, efetuou-se a segmentação da região de interesse que será posteriormente classificada como normal ou suspeita. O resultado da segmentação efetuada em uma das imagens de teste pode ser visto na Figura 27.

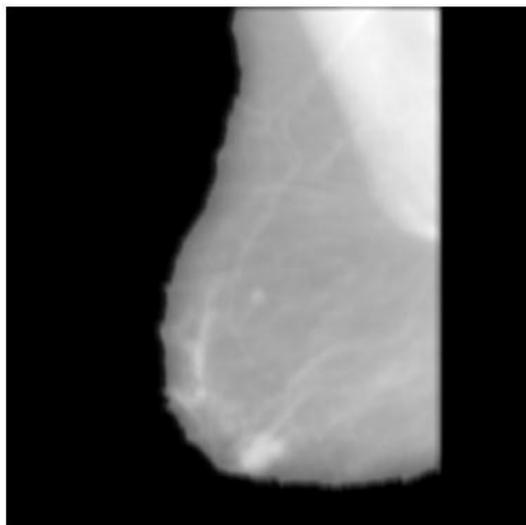


(a)

(b)

**Figura 27 - Resultado da segmentação.**  
(a) Imagem segmentada. (b) Região com provável lesão.

Como se pode ver na Figura 27, a segmentação efetua o agrupamento em classes baseado na proximidade das intensidades dos *pixels* da imagem resultante da filtragem de convolução com as funções base. A região circulada representa uma possível lesão, a ser confirmada pelo classificador. A imagem resultante já é o resultado do agrupamento utilizando *Kmeans*. O resultado da filtragem consiste em uma imagem com a região de interesse realçada, de acordo com o que se pode verificar na imagem da Figura 28.



**Figura 28 - Resultado da filtragem de convolução com as funções base na etapa de segmentação.**

A imagem resultante é então processada pelo programa que efetua a *classificação*. Nesta etapa, músculo peitoral é retirado da imagem resultante da segmentação, e região realçada é separada do restante da imagem, utilizando um critério baseado na energia dos agrupamentos. O agrupamento com maior energia é selecionado para a fase de classificação, pois corresponde à região realçada pela filtragem. Neste processo, foi utilizada a máquina de vetor de suporte (SVM) para classificar as regiões segmentadas em normal ou suspeito. Foi utilizada a biblioteca LibSVM (CHANG e LIN, 2011) para treinamento e teste das regiões de interesse.

Para efetuar o treinamento e os testes no módulo de classificação, novamente foram selecionadas, da base de dados mini-MIAS, 93 imagens contendo lesões e 93 imagens sem lesões, de onde foram recortadas apenas as regiões de interesse, no caso das lesões, e regiões de tecido normal, no caso das imagens normais. Estas regiões recortadas foram armazenadas em diretórios diferentes para compor o banco de imagens de treinamento e o banco de imagens de teste, na proporção 50% para treino e 50% para testes. Todas as regiões de interesse recortadas possuíam tamanho 100x100 *pixels*. Em seguida, elaborou-se um

algoritmo para extrair amostras destas regiões de interesse e compor uma matriz de treinamento contendo regiões normais e suspeitas, juntamente com um rótulo indicando a que classe de imagem a amostra corresponde: 1 para normais e 2 para suspeitas.

Utilizando novamente o ICA, foram extraídas as funções base das amostras selecionadas, para em seguida efetuar a estimação das componentes independentes, as fontes que geram as classes de imagens normais e suspeitas, a serem utilizadas no treinamento da SVM. A escolha do tamanho das amostras e do número de componentes a extrair foi feita a partir de testes realizados no software MATLAB, através de um algoritmo que aumentava o número de componentes independentes a serem extraídas gradativamente, iniciando com 6 e indo até 24 componentes, em amostras de 30x30 *pixels*, e iniciando com 6 e indo até 50 componentes, em amostras de 50x50 *pixels*. Para cada quantidade de componentes, foram extraídas as funções base do banco de imagens de treino, estimadas as componentes independentes, efetuado o treinamento da SVM e realizado os testes com o banco de testes, 10 vezes.

Para cada teste, foi registrada a porcentagem de acertos com as imagens de teste normais e suspeitas, e foram calculadas no final de cada bateria de testes, a média de acerto e respectivo desvio padrão. O resultado dos testes com amostras de 30x30 *pixels* encontra-se na Tabela 2. O resultado dos testes com amostras de 50x50 *pixels* encontra-se na Tabela 3. Realizando uma análise de variância (ANOVA) observou-se que o melhor resultado foi utilizando amostras de tamanho 50x50 *pixels* e 24 componentes independentes, obtendo taxa de acerto médio de 89,13% e desvio padrão de 2,17% para regiões normais, e taxa de acerto médio de 92,17% e desvio padrão de 2,42% para regiões suspeitas. Podem-se visualizar as linhas em destaque com os resultados selecionados na Tabela 3.

**Tabela 2 - Treinamento SVM com amostras de 30x30 pixels: escolha do número de IC's.**

IC's	ROI's	Teste 1	Teste 2	Teste 3	Teste 4	Teste 5	Teste 6	Teste 7	Teste 8	Teste 9	Teste 10	Média	Desv. Pad.
6	Normal	78,26	89,13	76,09	78,26	86,96	80,43	89,13	84,78	82,61	82,61	82,83	4,40
	Suspeito	89,13	86,96	82,61	84,78	84,78	82,61	86,96	89,13	84,78	89,13	86,09	2,42
12	Normal	80,43	76,09	82,61	86,96	91,3	82,61	80,43	71,74	76,09	84,78	81,30	5,43
	Suspeito	93,48	89,13	86,96	84,78	86,96	86,96	89,13	86,96	84,78	93,48	88,26	2,95
18	Normal	80,43	84,78	84,78	78,26	76,09	73,91	80,43	86,96	76,09	86,96	80,87	4,54
	Suspeito	86,96	80,43	91,3	93,48	91,3	84,78	91,3	89,13	93,48	93,48	89,56	4,10
24	Normal	84,78	78,26	69,57	82,61	84,78	78,26	80,43	78,26	73,91	84,78	79,56	4,78
	Suspeito	91,3	84,78	91,3	93,48	89,13	93,48	84,78	84,78	86,96	91,3	89,13	3,37

Tabela 3 - Treinamento SVM com amostras de 50x50 pixels: escolha do número de IC's.

IC's	ROI's	Teste 1	Teste 2	Teste 3	Teste 4	Teste 5	Teste 6	Teste 7	Teste 8	Teste 9	Teste 10	Média	Dev. Pad.
6	Normal	91,3	84,78	91,3	84,78	86,96	89,13	91,3	89,13	93,48	89,13	89,13	2,75
	Suspeito	89,13	86,96	89,13	91,3	91,3	89,13	93,48	84,78	84,78	86,96	88,70	2,71
12	Normal	91,3	86,96	84,78	78,26	82,61	91,3	84,78	97,83	80,43	78,26	85,65	6,01
	Suspeito	84,78	82,61	86,96	91,3	89,13	89,13	89,13	89,13	91,3	89,13	88,26	2,61
18	Normal	84,78	82,61	95,65	80,43	91,3	86,96	93,48	89,13	82,61	93,48	88,04	5,08
	Suspeito	95,65	93,48	93,48	93,48	84,78	91,3	84,78	86,96	91,3	91,3	90,65	3,64
24	<u>Normal</u>	<u>86,96</u>	<u>84,78</u>	<u>91,3</u>	<u>89,13</u>	<u>91,3</u>	<u>89,13</u>	<u>91,3</u>	<u>86,96</u>	<u>89,13</u>	<u>91,3</u>	<u>89,13</u>	<u>2,17</u>
	<u>Suspeito</u>	<u>95,65</u>	<u>91,3</u>	<u>93,48</u>	<u>89,13</u>	<u>89,13</u>	<u>95,65</u>	<u>93,48</u>	<u>89,13</u>	<u>93,48</u>	<u>91,3</u>	<u>92,17</u>	<u>2,42</u>
30	Normal	93,48	80,43	80,43	93,48	89,13	89,13	93,48	91,3	84,78	93,48	88,91	5,00
	Suspeito	89,13	91,3	93,48	93,48	91,3	95,65	89,13	93,48	89,13	89,13	91,52	2,27
50	Normal	95,65	93,48	89,13	82,61	84,78	86,96	91,3	86,96	78,26	86,96	87,61	4,87
	Suspeito	93,48	91,3	89,13	95,65	91,3	91,3	84,78	93,48	91,3	89,13	91,09	2,83

Acredita-se que os resultados obtidos com janelas de 50x50 *pixels* e 24 componentes independentes tenham gerado melhores resultados nos testes, em razão da estrutura das imagens mamográficas serem melhor representadas com essa quantidade de componentes, extraídas a partir de janelas com esta dimensão.

De posse destas componentes independentes, foi gerado um arquivo de configuração utilizado pelo programa de classificação. A Figura 29 apresenta a região cujo agrupamento possui maior energia e que será utilizada, após ser recortada da imagem original, no processo de classificação.

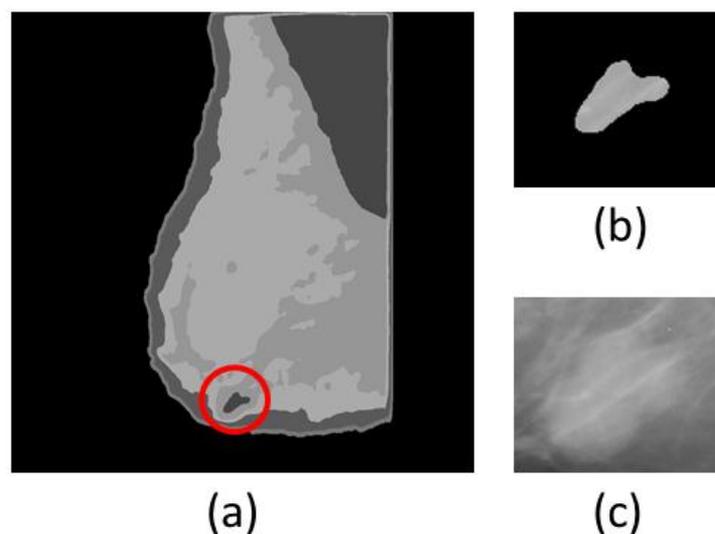
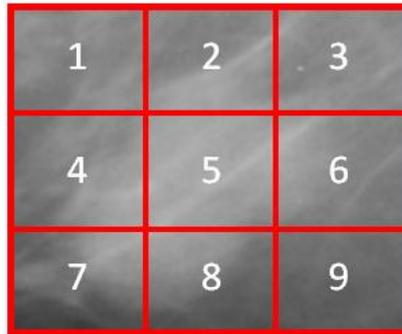


Figura 29 - Região de interesse que será classificada. (a) Resultado da segmentação. (b) Agrupamento com maior energia. (c) Região recortada para ser classificada.

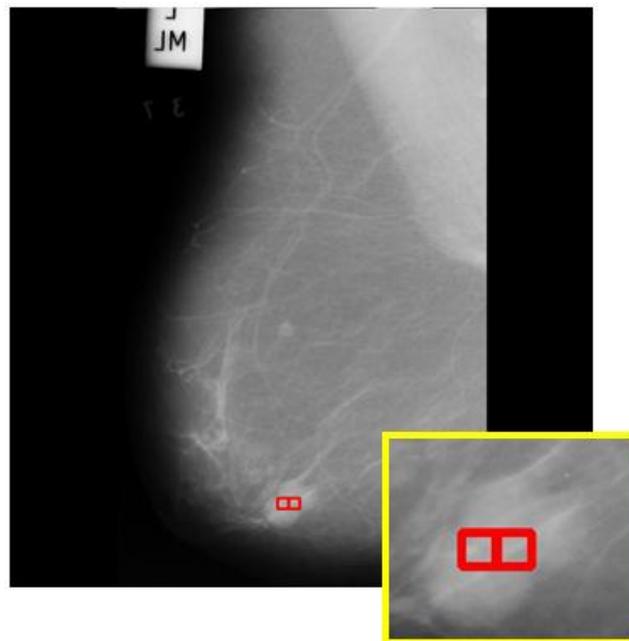
Ao receber a imagem segmentada, o programa de classificação divide a região de interesse em quadrantes de 50x50 *pixels* e, para cada quadrante, retira amostras da região

recortada, armazena em um vetor de testes e executa a LibSVM para realizar a classificação. Através da figura, pode-se observar um exemplo de uma região que será classificada, dividida em quadrantes.



**Figura 30 - Divisão da região segmentada em quadrantes para classificação.**

O resultado da classificação em uma das imagens usadas para teste no sistema pode ser visualizado na Figura 31.



**Figura 31 - Resultado da classificação em uma imagem com lesão.**

Após a classificação de cada quadrante, é armazenada no banco de dados a localização de cada quadrante, permitindo desta forma, um registro das posições de cada lesão para acompanhamento futuro do quadro do paciente.

### 6.3.2 Camada de apresentação

A camada de apresentação, responsável pela interface de usuário, permite a interação dos usuários com o sistema. É através dela que o usuário solicitante efetua o cadastro do paciente, o envio da imagem a ser processada e visualiza o resultado do processamento, bem como a confirmação do diagnóstico dado pelo usuário especialista.

Esta camada é composta de páginas Web, desenvolvidas em HTML, juntamente com a tecnologia JSP, baseada na plataforma Java. As páginas corresponderão às telas do sistema, ou seja, à interface de interação com o sistema. Para ter acesso ao sistema, o usuário acessará a tela de *login*, onde informará seu nome de usuário e sua senha. Através deste *login* o sistema reconhecerá o perfil do usuário, solicitante ou especialista, e montará uma página com um *menu* de opções específico para cada perfil. A página de *login* pode ser visualizada na Figura 32.

A imagem mostra a interface de login de um sistema web. No topo, há um cabeçalho azul escuro com o texto "Sistema de Análise de Imagens Mamográficas" em branco. Abaixo, o conteúdo principal tem um fundo branco e contém o seguinte:

- Um subtítulo azul: "Sistema de análise de imagens mamográficas pela Web".
- Um formulário de login com um cabeçalho azul contendo o texto ":: Acesso ao sistema".
- Dentro do formulário, há dois campos de entrada: "Login:" e "Senha:", cada um com um campo de texto adjacente.
- Abaixo dos campos, há dois botões: "Ok" e "Limpar".

Na base da interface, há uma barra azul com o texto "Desenvolvido por Luis Claudio (2012) versão 1.0" em branco.

Figura 32 - Tela de *login* do sistema proposto.

A partir da tela principal do sistema, o usuário solicitante terá acesso um menu contendo as opções:

- *Cadastrar Paciente*, onde poderá ser efetuado o cadastro dos pacientes cujas imagens serão analisadas;
- *Cadastrar Exame*, onde serão fornecidas as imagens que serão analisadas, podendo ser cadastrado mais de uma imagem para processamento;

- *Visualizar Exame*, onde o solicitante poderá acessar o resultado do processamento da imagem, bem como o laudo fornecido pelo especialista com a confirmação do diagnóstico.

Esta tela pode ser visualizada na Figura 33.

The screenshot shows the main interface of the 'Sistema de Análise de Imagens Mamográficas'. At the top, there is a dark blue header with the system name in white. Below the header is a navigation bar with four links: 'Cadastrar paciente', 'Cadastrar exame', 'Visualizar exame', and 'Sair'. The user is logged in as 'Usuário: Solicitante de Teste 1'. The main content area is titled 'Sistema de análise de imagens mamográficas pela Web' and contains a welcome message: 'Bem vindo ao SismamaWeb!!! Você está acessando com o perfil: **Usuário Solicitante**.' Below this, it says 'Para utilizar o sistema:' followed by three numbered instructions: 1. 'Acesse o menu Cadastrar paciente e efetue o cadastro dos dados do paciente.', 2. 'Acesse o menu Cadastrar exame para efetuar o cadastro dos exames de cada paciente.', and 3. 'Acesse o menu Visualizar exame para ter acesso ao resultado do processamento das imagens de cada exame.' At the bottom, there is a footer: 'Desenvolvido por Luis Claudio (2012) versão 1.0'.

**Figura 33 - Tela principal do usuário solicitante.**

Ao selecionar a função *Cadastrar Paciente*, o usuário poderá efetuar o cadastro anônimo do paciente, informando a priori os dados: data de nascimento, se há casos de câncer na família do paciente, o tempo de lactação em meses e se o paciente possui nódulo palpável.

The screenshot shows the 'Cadastrar paciente' form. The header and navigation bar are identical to Figure 33. The user is still logged in as 'Usuário: Solicitante de Teste 1'. The main content area is titled 'Sistema de análise de imagens mamográficas pela Web'. A sub-header for the form reads ':: Cadastrar paciente'. Below this, it says 'Digite os dados do paciente e clique no botão Ok para efetuar o envio dos dados.' The form contains four fields: 'Data de Nascimento:' with a text input field; 'Há casos de câncer na família?' with a dropdown menu; 'Tempo de lactação:' with a text input field containing '0' and the label 'meses'; and 'Possui nódulo:' with a dropdown menu. At the bottom right of the form are two buttons: 'Ok' and 'Limpar'. At the bottom of the page, there is a footer: 'Desenvolvido por Luis Claudio (2012) versão 1.0'.

**Figura 34 - Tela de cadastro de pacientes.**

Estes dados serão armazenados no banco de dados e pode ajudar o usuário especialista a inferir sobre o resultado da na análise da imagem ao analisar o histórico do paciente. A tela que permite efetuar este cadastro encontra-se na Figura 34.

Ao efetuar o cadastro do paciente, o usuário solicitante poderá cadastrar o exame deste paciente. No cadastro do exame, deverá ser informado o motivo da análise da imagem bem como será selecionado o arquivo de imagem que será processado no servidor. Após a confirmação do envio, a imagem será colocada na fila de imagens a serem processadas, e posteriormente o usuário solicitante poderá visualizar o resultado do processamento em outra tela. A tela de cadastro do exame pode ser visualizada na Figura 35.

A imagem mostra a interface de usuário de um sistema web. No topo, há um cabeçalho azul com o título "Sistema de Análise de Imagens Mamográficas" em branco. Abaixo dele, uma barra de navegação cinza contém os links "Cadastrar paciente | Cadastrar exame | Visualizar exame | Sair". No canto superior direito, o texto "Usuário: Solicitante de Teste 1" é exibido. O conteúdo principal da página é "Sistema de análise de imagens mamográficas pela Web". Centralizado, há um formulário com o título "Cadastrar exame" em uma barra azul. O formulário contém o texto "Digite os dados do exame e clique em Ok para efetuar o envio dos dados." e os seguintes campos: "Paciente: 1", "Motivo:" seguido de um campo de texto grande, "Imagem:" seguido de um campo de texto e um botão "Selecionar arquivo". Na base do formulário, há dois botões: "Ok" e "Limpar". Na parte inferior da tela, uma barra azul contém o texto "Desenvolvido por Luis Claudio (2012) versão 1.0".

**Figura 35 - Tela de cadastro dos exames.**

A visualização do resultado do exame é feita a partir da tela exibida na Figura 36. Após informar o código do paciente, o sistema exibirá a lista de exames cadastrados para este paciente. Na coluna *Concluído*, caso o sistema tenha encerrado o processamento do exame e o especialista tenha liberado o diagnóstico, o usuário terá acesso a outra página onde poderá visualizar o resultado da análise da imagem efetuada pelo sistema, bem como o laudo informado pelo especialista. A partir da tela da Figura 36, é possível visualizar dois exames de teste cadastrados para um paciente de teste. Ao clicar no link *Sim*, o usuário terá acesso ao resultado deste exame, conforme se observa na Figura 37.

## Sistema de Análise de Imagens Mamográficas

Cadastrar paciente | Cadastrar exame | Visualizar exame | Sair

Usuário: Solicitante de Teste 1

Sistema de análise de imagens mamográficas pela Web

**:: Visualizar exame**

Paciente: 1

Cód. Exame	Data Exame	Concluído
1	15/04/2012	<a href="#">Sim</a>
2	17/04/2012	Não

Desenvolvido por Luis Claudio (2012) versão 1.0

Figura 36 - Tela de visualização dos exames.

## Sistema de Análise de Imagens Mamográficas

Cadastrar paciente | Cadastrar exame | Visualizar exame | Sair

Usuário: Solicitante de Teste 1

Sistema de análise de imagens mamográficas pela Web

**:: Resultado do exame**

Paciente: 1

Código do Exame: 1

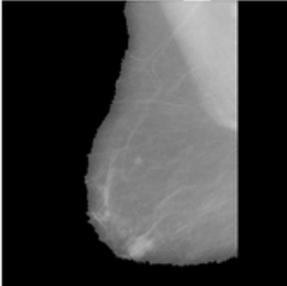
Data do Exame: 15/04/2012

Laudo: A imagem apresenta achados radiológicos suspeitos.

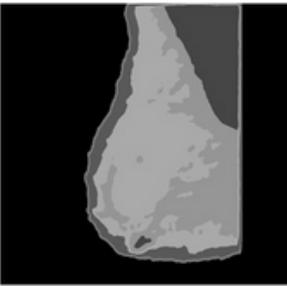
Resultado: **Suspeito**

Imagens resultante do processamento:

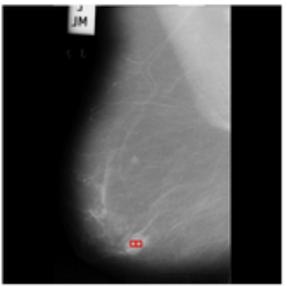
Pré-segmentação



Segmentação



Classificação



Clique na imagem para aumentar o zoom.

Desenvolvido por Luis Claudio (2012) versão 1.0

Figura 37 - Tela de resultado do processamento do exame.

O usuário especialista, ao efetuar o *login* terá acesso a uma tela que lista todos os exames cadastrados no servidor para processamento, conforme se pode observar na Figura 38.

The screenshot shows a web application interface. At the top, there is an orange header with the text 'Sistema de Análise de Imagens Mamográficas'. Below this is a grey navigation bar with 'Listar exames | Sair'. The main content area is white and displays 'Usuário: Especialista 1' in the top right. Below that, the title 'Sistema de análise de imagens mamográficas pela Web' is shown in orange. A blue box contains the text ':: Listar exames'. Below this is a table with four columns: 'Cód. Exame', 'Cód. Paciente', 'Data Exame', and 'Concluído'. The table contains three rows of data. The first row has '1' for exam code, '1' for patient code, '15/04/2012' for the date, and a blue link 'Sim' for 'Concluído'. The second row has '2' for exam code, '1' for patient code, '17/04/2012' for the date, and 'Não' for 'Concluído'. The third row has '3' for exam code, '2' for patient code, '17/04/2012' for the date, and 'Não' for 'Concluído'. At the bottom of the screenshot, there is an orange footer with the text 'Desenvolvido por Luis Claudio (2012) versão 1.0'.

Cód. Exame	Cód. Paciente	Data Exame	Concluído
1	1	15/04/2012	<a href="#">Sim</a>
2	1	17/04/2012	Não
3	2	17/04/2012	Não

Figura 38 - Tela de listagem de exames do usuário especialista.

Caso o processamento do exame tenha sido concluído pelo servidor, e o usuário especialista dê um clique nos *links* apresentados na coluna *Concluído*, será apresentada a tela da Figura 39, para que o especialista possa verificar o resultado do exame, fornecer o diagnóstico e liberar o exame para visualização por parte do usuário solicitante.

## Sistema de Análise de Imagens Mamográficas

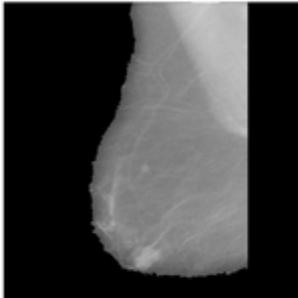
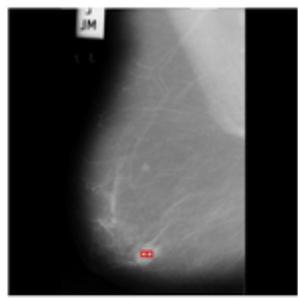
[Listar exames](#) | [Sair](#)

Usuário: Especialista 1

Sistema de análise de imagens mamográficas pela Web

:: Resultado do exame

Código do Exame: 1  
Código do Paciente: 1  
Há casos de câncer na família? **Sim**  
Tempo de lactação: **0 meses**  
Possui nódulo: **Sim**  
Data do Exame: **15/04/2012**  
Motivo do exame: **A paciente apresenta queixa de nódulo no seio.**  
Imagens resultante do processamento:

Pré-segmentação	Segmentação	Classificação
		

Clique na imagem para aumentar o zoom.

**Laudos:**

Desenvolvido por Luis Claudio (2012) versão 1.0

Figura 39 - Tela de liberação do exame pelo especialista.

## 7 RESULTADOS E DISCUSSÃO

Para a realização dos testes do sistema, um protótipo do sistema foi desenvolvido e instalado em uma máquina com processador Intel Core i5 contendo dois núcleos de 2,75GHz e 4 Gigabytes de memória RAM. Nesta máquina, foi instalado o servidor Web Apache Tomcat, para que a aplicação Web pudesse ser instalada, simulando um servidor Web de produção. O sistema operacional desta máquina de teste é o Windows 7 para arquitetura de 64 bits, no entanto a aplicação foi desenvolvida utilizando linguagens e ferramentas livres e multiplataformas, o que possibilita a implantação do sistema em um servidor com sistema operacional Linux, se necessário.

Uma vez efetuado o treinamento dos algoritmos, e concluído o protótipo do sistema, foi criado um usuário solicitante, bem como um usuário especialista para os testes, a fim de avaliar o desempenho do sistema e validar a classificação. Foram selecionadas aleatoriamente 41 imagens, das 330 imagens disponíveis no mini-MIAS, contendo algum tipo de lesão. As imagens foram submetidas ao protótipo para processamento e o sistema obteve sucesso em detectar 36 imagens que contém lesão, considerando que qualquer imagem que contenha algum tipo de achado maligno ou benigno seja considerada suspeita.

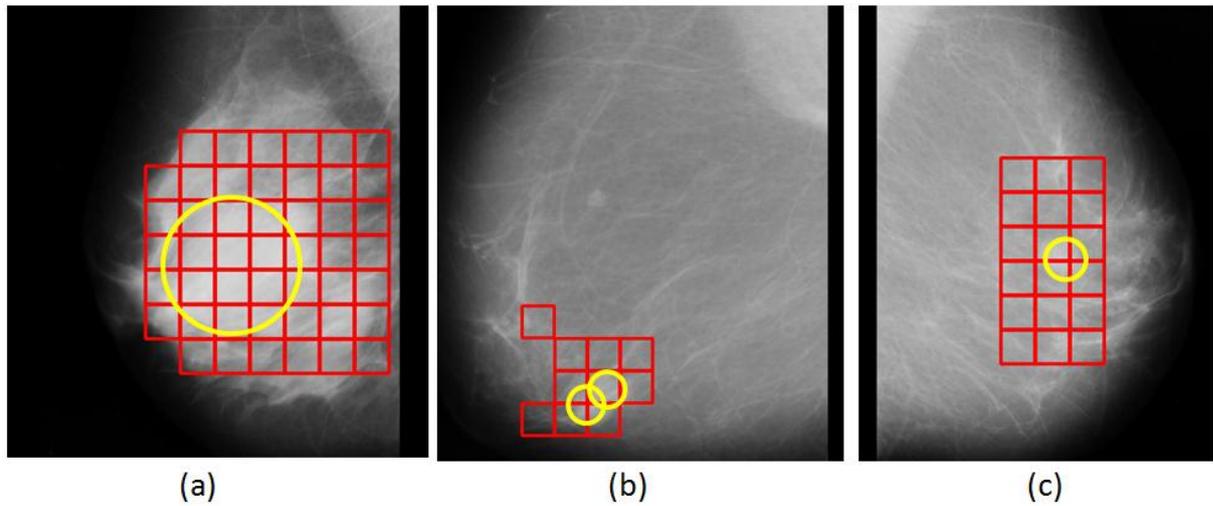
Sensibilidade é uma das medidas utilizadas para descrever um classificador, sendo definida como sendo a proporção de verdadeiros positivos classificados corretamente e identificados pelo classificador, dada pela Eq. (34). A sensibilidade do sistema proposto foi de 87,8%.

$$S = \frac{VP}{VP + FN} \quad (34)$$

As imagens aplicadas no teste possuem característica de tecido predominantemente adiposo (*Fatty*, na classificação do mini-MIAS), e observou-se que as imagens classificadas de forma errônea possuíam região lesionada com tecido muito semelhante ao tecido normal, o que pode gerar um erro de classificação. Outro detalhe importante é que as imagens cujo tecido é denso ou predominantemente denso, ocasionam um erro na classificação em virtude do tecido mamário se assemelhar a tecidos lesionados.

Através da Figura 40, pode-se ver o resultado do processamento em três imagens que contém lesão. Os retângulos correspondem à região segmentada e classificada pelo

sistema, enquanto que os círculos correspondem à região da lesão segundo os registros do mini-MIAS.



**Figura 40 - Exemplos de imagens classificadas pelo sistema: regiões quadriculadas correspondem à classificação do sistema; o círculo corresponde à localização definida pelo mini-MIAS.**

De acordo com a figura, observa-se que a região definida pela base de dados mini-MIAS está dentro da região classificada pelo sistema.

## 8 CONCLUSÃO E TRABALHOS FUTUROS

Neste trabalho, foi desenvolvido um sistema Web para processamento de imagens mamográficas, baseado em análise de componentes independentes e máquina de vetor de suporte. Para este desenvolvimento, foi efetuado a análise do sistema proposto e a implementação utilizando tecnologias de programação para aplicações Web, objetivando desenvolver uma ferramenta de telediagnóstico que possa auxiliar no processo de detecção de regiões suspeitas em imagens mamográficas.

A partir do levantamento de requisitos, foi efetuada a modelagem do sistema, de onde puderam ser obtidos os diagramas de classes e do banco de dados da aplicação. Em seguida, foi efetuada a construção do software partindo da implementação dos algoritmos de classificação de regiões lesionadas em imagens mamográficas baseados em análise de componentes independentes para extração de características de regiões de interesse em imagens mamográficas e análise de vetor de suporte para efetuar a classificações destas características. Estes algoritmos foram escritos em C++, utilizando a biblioteca de uso em aplicações gráficas OpenCV. Além dos algoritmos de detecção e classificação, foi criada a interface gráfica, baseado em padrões da Web, pois a aplicação se propõe a ser uma ferramenta de telediagnóstico, utilizando a plataforma/linguagem de programação Java.

Um protótipo do sistema foi construído para a realização dos testes em 41 imagens mamográficas escolhidas aleatoriamente da base de dados mini-MIAS, onde foi obtido 87,8% de acerto nos achados de regiões suspeitas.

Além de servir como ferramenta de diagnóstico, o sistema se propõe também a fazer um registro de imagens mamográficas contendo informações da imagem analisada e dos achados radiológicos encontrados pelo sistema, o que pode contribuir para estudos futuros tais como:

- desenvolvimento de algoritmos de aprendizagem a partir de novos casos registrados, supervisionados ou automatizados, objetivando melhorar a acurácia do classificador;
- aplicação de outras técnicas de classificação em conjunto à técnica atualmente implementada, permitindo dessa forma a elaboração de um classificador híbrido, que possa combinar resultados na obtenção de uma sugestão de

diagnóstico mais segura e eficiente tanto para o especialista como para o paciente;

- aprimoramento do algoritmo de detecção para ser utilizado em mamas densas, de difícil decisão diagnóstica, em função do tipo de tecido mamográfico presentes nestas imagens dificultar a localização de lesões;
- criação de uma interface mais rica, para uso por parte do especialista, dentro da aplicação web, permitindo seleção de regiões específicas, visualização das imagens em negativo, ou melhoramento gráfico da imagem, a partir de comandos adicionados na interface da aplicação.

As principais contribuições deste trabalho são:

- apresentação da arquitetura do sistema de telediagnóstico proposto, explicitando os módulos que o compõem;
- modelagem do sistema apresentando os diagramas de classes do sistema, bem como os diagramas do banco de dados da aplicação;
- implementação do protótipo do sistema para realização de testes com imagens mamográficas.

Este trabalho deu origem ao seguinte artigo, publicado no XI Workshop de Informática Médica:

Lucio Flávio A. Campos, Emanuel C. M. Lemos, Luis C. O. Silva, Daniel D. Costa, Allan Kardec Barros: “Segmentation and Classification of Breast Cancer Using Independent Component Analysis, Texture Features and Neural Networks”, XI Workshop de Informática Médica, Natal, 2011.

## REFERÊNCIAS

- BARBOSA, A. K. et al. HealthNet: um sistema integrado de telediagnóstico e segunda opinião médica. **Boletim bimestral sobre tecnologia de redes**, 5, n. 4, 2001.
- BEZERRA, E. **Princípios de análise e projeto de sistemas com UML**. Rio de Janeiro: Campus, 2001.
- BISHOP, C. M. **Pattern Recognition And Machine Learning**. New York: Springer, 2006.
- BRADSKI, G. The OpenCV Library. **Dr. Dobb's Journal of Software Tools**, 2000.
- CHANG, C.-C.; LIN, C.-J. LIBSVM: A library for support vector machines. **ACM Transactions on Intelligent Systems and Technology**, 2, n. 3, 2011. 1-27.
- CHOI, S. et al. Blind Source Separation and Independent Component Analysis: A Review. **Neural Information Processing - Letters and Reviews**, 6, n. 1, January 2005.
- CHRISTOYIANNI, I. et al. Computer aided diagnosis of breast cancer in digitized mammograms. **Computerized Medical Imaging and Graphics**, 26, 2002. 309-319.
- DANCE, D. R.; DAVIES, D. H. Automatic computer detection of clustered calcifications in digital mammograms. **Physics in Medicine and Biology**, 35, 1990.
- DUAN, L. et al. A Texture Images Segmentation Method Based on ICA Filters. **Fifth International Conference on Natural Computation**, 6, August 2009. 484-487.
- ELMASRI, R.; NAVATHE, S. B. **Fundamentals of database systems**. 4. ed. [S.l.]: Addison Wesley, 2004.
- GONZALEZ, R. C.; WOODS, R. E.; EDDINS, S. L. **Digital Image Processing Using MATLAB**. Upper Saddle River: Prentice-Hall Europe, 2003.
- HAYKIN, S. **Neural Networks: A comprehensive foundation**. 2. ed. [S.l.]: Prentice Hall, 1999.
- HYVÄRINEN, A. Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. **IEEE Transactions on Neural Networks**, 10, 1999.

HYVÄRINEN, A.; KARHUNEN, J.; OJA, E. **Independent component analysis**. New York: Wiley, 2001.

HYVÄRINEN, A.; OJA, E. Independent Component Analysis: Algorithms and Application. **Neural Networks**, 13, n. 4-5, 2000.

INCA. Instituto Nacional do Câncer. **Instituto Nacional do Câncer**, Janeiro 2012. Disponível em: <<http://www.inca.gov.br>>. Acesso em: 15 Janeiro 2012.

INCA, I. N. D. C. **Mamografia: da prática ao controle**. Rio de Janeiro: INCA, 2007.

JACOBSON, I. et al. **Object-Oriented Software Engineering - A use case driven approach**. New York: Addison-Wesley/ACM Press Books, 1992.

MARQUES FILHO, O.; VIEIRA NETO, H. **Processamento Digital de Imagens**. Rio de Janeiro: Brasport, 1999.

MASTOLOGIA, S. B. D. **Sociedade Brasileira de Mastologia**, Julho 2012. Disponível em: <<http://www.sbmastologia.com.br>>.

PAPOULIS, A.; PILLAI, S. U. **Probability, Random Variables and Sthocastic Processes**. 4. ed. New York: McGraw-Hill, 2002.

PRESSMAN, R. S. **Engenharia de Software**. 5. ed. [S.l.]: McGraw Hill, 2002.

ROSA, R. B. et al. **Avaliação do Crescimento da Telemedicina Brasil e no Mundo**. Anais e Programação do CBIS'2006. Florianópolis: [s.n.]. 2006.

SAÚDE, S. Ú. D. **Programa Nacional de Telessaúde**, Julho 2012. Disponível em: <<http://www.telessaude.or.br>>.

SHORTLIFFE, E. H.; BLOIS, M. S. The Computer Meets Medicine and Biology: Emergence of a Discipline. In: \_\_\_\_\_ **SHORTLIFFE EH, PERREAULT LE, WIEDERHOLD G, FAGAN LM. Medical Informatics: Computer Applications in Health Care and Biomedicine**. New York: Springer, 2001. Cap. 1.

SIZILIO, G. R. M. A. et al. Ambiente de telediagnóstico colaborativo utilizando plataforma inteligente de auxílio à tomada de decisão. **Revista Brasileira de Engenharia Biomédica**, 27, n. 3, Setembro 2011.

SMOLA, A. J.; SCHÖLKOPF, B. **Learning with Kernels**. Cambridge: MIT Press, 2002.

SOMMERVILLE, I. **Engenharia de software**. 8. ed. [S.l.]: Addison Wesley Bra, 2008.

STONE, J. V. **Independent component analysis: a tutorial introduction**. Cambridge,: MIT Press, 2004.

SUCKLING, J. et al. The mammographic images analysis society digital mammogram database. **Experta Medica International Congress Series**, 1069, 1994. 375-378.

TANG, J. et al. Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. **Transactions on Information Technology in Biomedicine**, 13, n. 2, Março 2009.

TANG, J. et al. Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. **Trans. Info. Tech. Biomed**, v. 13, n. 2, p. 236-251, January 2009.

WHO, W. H. O. **Telemedicine: opportunities and developments in Member States: report on the second global survey on eHealth**. 2. ed. Geneva: WHO Press, 2010.

WINSBERG, F. et al. Detection of Radiographic Abnormalities in Mammograms by Means of Optical Scanning and Computer Analysis. **Radiology**, 2, 1967.