

# **Método para Avaliação dos Algoritmos Utilizados no Processamento de Imagens Médicas**



**Silvia Cristina Martini Rodrigues**

Dissertação apresentada à Escola de Engenharia de São Carlos, Universidade de São Paulo, como parte dos requisitos para obtenção do Título de Mestre em Engenharia Elétrica.

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Annie Francè Frère Slaets

São Carlos

1999

**FOLHA DE APROVAÇÃO**

Candidata: Licenciada **SILVIA CRISTINA MARTINI RODRIGUES**

Dissertação defendida e aprovada em 24-09-1999  
pela Comissão Julgadora:



---

Profª. Titular **ANNIE FRANCE FRÈRE SLAETS (Orientadora)**  
(Escola de Engenharia de São Carlos - Universidade de São Paulo)



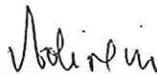
---

Prof. Associado **HOMERO SCHIABEL**  
(Escola de Engenharia de São Carlos - Universidade de São Paulo)



---

Prof. Titular **HENRIQUE KRIEGER**  
(Instituto de Ciências Biomédicas - Universidade de São Paulo)



---

Profª. Associada **VILMA ALVES DE OLIVEIRA**  
Coordenadora da Área de Engenharia Elétrica



---

**JOSÉ CARLOS A. CINTRA**  
Presidente da Comissão de Pós-Graduação

Dedico este trabalho ao meu marido e amigo Evandro, às minhas filhas Camila e Marília, pelo incentivo interminável, compreensão e privações. Muito obrigada por vocês existirem.

## AGRADECIMENTOS

Gostaria de agradecer a todas as pessoas que de uma forma ou de outra, me ajudaram na realização deste trabalho.

- A Profa. Dra. Annie France Frere Slaets, minha orientadora e amiga pela orientação, apoio e incentivo nas horas difíceis.
- Aos meus amigos de laboratório: Renata, Márcia, Márcio, Irita, Henrique, Chico Bento, Prof. Paulo, Jacaré, Sérgio, Lucas, Marcelo e Leandro pela convivência diária tão agradável.
- As meninas da biblioteca, principalmente a Eleninha, pelo auxílio na procura dos artigos e principalmente pelo bate papo tão gostoso.
- As secretárias e funcionários do Departamento de Engenharia Elétrica, pelo auxílio e amizade.
- Ao prof. Dr. Henrique Krieger pelo esclarecimento das dúvidas que surgiram durante o trabalho.
- A Capes pelo apoio financeiro.

Muito obrigada,

Aos meus pais Luiz e Geni a quem eu amo muito, por terem me dado meu bem maior: A VIDA; pois sem ela não teria conseguido nada disto.

Aos meus sogros, Luiz e Leontina, aos meus cunhados Rose e Hamilton, Giza e Dami, Ketty e Samir. Sobrinhos: Lya, Rafael, Caio, Karla, Karina (afilhada), Gabriel e Karolina por estarmos juntos nesta vida.

A Deus por ter me permitido tudo isto!

## RESUMO

Este trabalho apresenta como parte de resultados, uma ampla pesquisa que permitiu identificar os grupos de pesquisas mais importantes do mundo, os quais possuem em comum o processamento de imagens médicas, mais especificamente o processamento de imagens que busca a identificação de microcalcificações mamárias.

O vasto levantamento, a seleção e organização culminou na reunião de mais de cem artigos, publicados nos mais importantes periódicos da área, que mostram claramente as formas utilizadas pelos grupos de pesquisa para apresentação dos resultados encontrados pelos seus algoritmos. Esses resultados devem auxiliar o médico no diagnóstico do câncer de mama.

Demonstramos neste trabalho porque as técnicas utilizadas para apresentação dos resultados são insatisfatórias e propusemos um novo método de avaliação desses resultados.

O método proposto no trabalho baseia-se no teste do  $\chi^2$  (Qui-Quadrado), nas curvas ROC (Receiver Operating Characteristic) e no teste de concordância, que juntos permitem apresentar de forma clara e objetiva as relações entre verdadeiros positivos e falsos positivos, verdadeiros negativos e falsos negativos, sensibilidade e especificidade do algoritmo analisado.

O novo método é preciso e tem bases estatísticas conhecidas pelos médicos e pelos pesquisadores, facilitando sua aceitação.

## ABSTRACT

This work presents as part of results, a wide investigation that it allowed to identify the principal research groups of the world, which possess in common the processing of medical images, more specifically the processing of images that search for the identification of mammary microcalcifications.

The vast collection, selection and organization culminated in the meeting of more than a hundred articles, published in the most important newspapers of the area, that show the forms used by the research groups to present the results found clearly by its algorithms. Those results should assist the doctor in the diagnosis of the breast cancer.

We demonstrated in this work that the techniques used for presentation of the results are unsatisfactory and we proposed a new method of evaluation of those results.

The proposed method bases on the test of the  $\chi^2$  (Qui-square), in ROC curve (Receiver Operating Characteristic) and in the agreement test, that take together allow to present in a clear and objective way the relationships among true positive and false positive, true negative and false negative, sensibility and specificity of the analyzed algorithm.

The new method is precise and has statistical bases known by the clinicians and researchers, facilitating its acceptance.

## SUMÁRIO

<b>CAPÍTULO 1</b> .....	<b>1</b>
<i>INTRODUÇÃO</i> .....	<i>1</i>
1.1 MOTIVAÇÃO E OBJETIVO DESTE TRABALHO.....	1
1.2 APRESENTAÇÃO DA DISSERTAÇÃO.....	2
<b>CAPÍTULO 2</b> .....	<b>3</b>
<i>DADOS GERAIS SOBRE O CÂNCER DE MAMA</i> .....	<i>3</i>
<i>INTRODUÇÃO</i> .....	<i>3</i>
1.1 DADOS ESTATÍSTICOS DE CÂNCER DE MAMA.....	5
1.2 DIAGNÓSTICO AUXILIADO POR COMPUTADOR .....	8
<b>CAPÍTULO 3</b> .....	<b>11</b>
MÉTODOS CLÁSSICOS DE AVALIAÇÃO DO PROCESSO DIAGNÓSTICO POR PROCESSAMENTO DE IMAGENS .....	11
<i>INTRODUÇÃO</i> .....	<i>11</i>
3.1 CURVAS ROC.....	12
3.2 GRÁFICO DA CURVA ROC.....	16
3.3 “THRESHOLD” DA CURVA ROC .....	18
3.4 DESVIO PADRÃO .....	20
3.5 ÁREA SOB A CURVA ROC.....	22
3.6 COMPARAÇÕES ENTRE CURVAS ROC .....	23
3.7 PROCEDIMENTOS DE OBTENÇÃO DA CURVA ROC .....	25
3.7.1 MÉTODO DA TAXAÇÃO.....	25
3.7.2 FROC.....	26
3.7.3 FROCFIT .....	29
3.7.4 PORCENTAGEM.....	29
3.8 TESTE DO QUI-QUADRADO $\chi^2$ .....	29
<b>CAPÍTULO 4</b> .....	<b>33</b>
DESCRIÇÃO E ANÁLISE DOS RESULTADOS OBTIDOS.....	33
<i>INTRODUÇÃO</i> .....	<i>33</i>
4.1 LEVANTAMENTO DOS MÉTODOS UTILIZADOS PARA APRESENTAR O DESEMPENHO DE ALGORITMOS PARA DETECÇÃO DE MICROCALCIFICAÇÕES.....	33
4.1.1 GRUPOS QUE ESTUDAM A DETECÇÃO DE MICROCALCIFICAÇÃO .....	34
4.1.2 CLASSIFICAÇÃO DOS RESULTADOS .....	43
4.1.2.1 ROC CONVENCIONAL.....	43
4.1.2.2 ÁREA SOB A CURVA ROC .....	43
4.1.2.3 FROC CONVENCIONAL - (FREE-RESPONSE RECEIVER OPERATING CHARACTERISTIC).....	43

4.1.2.4 FROC – UTILIZANDO O MÉTODO FROCFIT .....	44
4.1.2.5 SENSIBILIDADE OU VERDADEIRO POSITIVO – VP .....	44
4.1.2.6 ESPECIFICIDADE –VERDADEIRO NEGATIVO – VN.....	44
4.1.2.7 FALSO POSITIVO – FP .....	44
4.1.2.8 FALSO NEGATIVO – FN .....	45
4.1.2.9 FALSO POSITIVO (FP) POR IMAGEM.....	45
4.1.2.10 PRECISÃO .....	45
4.1.2.11 PORCENTAGEM (%).....	45
4.1.3 RESUMO DOS MÉTODOS DE AVALIAÇÃO APRESENTADOS NA LITERATURA.....	50
4.2 EXEMPLOS DE ALTA E BAIXA SENSIBILIDADE.....	51
4.3 NOVO MÉTODO PARA APRESENTAR OS RESULTADOS DE DIAGNÓSTICOS .....	60
<b>CAPÍTULO 5 .....</b>	<b>71</b>
CONCLUSÕES.....	71
5.1 CONCLUSÕES GERAIS .....	71
5.2 CONTINUAÇÃO DESTA PESQUISA.....	72
<b>CAPÍTULO 6 .....</b>	<b>73</b>
REFERÊNCIAS BIBLIOGRÁFICAS .....	73

## LISTA DE FIGURAS

<i>Figura 1 Diagrama em Blocos do Sistema CAD para Mamografias [DOI-1995].....</i>	<i>9</i>
<i>Figura 2 - Sistema de detecção por imagem [Evans 1981].....</i>	<i>12</i>
<i>Figura 3 - Matriz de estímulo – resposta de um sistema de imagem [Evans 1981].....</i>	<i>13</i>
<i>Figura 4 - Curva ROC típica [METZ – 1986].....</i>	<i>16</i>
<i>Figura 5 – Informação contida na matriz de estímulos, usando as coordenadas do VP e FP no ponto A. Quando o critério do observador é mudado gera-se o ponto B. [EVANS,1981].....</i>	<i>17</i>
<i>Figura 6 - Distribuição das probabilidades de um sinal percebido X, quando introduzido o ruído e o sinal + ruído. <math>X_{TH}</math> é o valor do “threshold” do sinal percebido sobre qual valor o observador responde Sinal [EVANS, 1981].....</i>	<i>19</i>
<i>Figura 7 - Curva representando as probabilidades das distribuições da certeza de um radiologista em um diagnóstico positivo para um particular diagnóstico. O “threshold” é representado pela linha vertical, separando decisões negativas de positivas. [METZ-1986] .....</i>	<i>19</i>
<i>Figura 8 - Comparação entre duas curvas com diferença de detectabilidade [ EVANS, 1981].....</i>	<i>23</i>
<i>Figura 9 - Distribuição sinal – ruído das duas curvas A e B mencionadas anteriormente [Evans, 1981].....</i>	<i>24</i>
<i>Figura 10- a) duas curvas ROC A e B que se interceptam.....</i>	<i>24</i>
<i>Figura 11 - FROC - O corte do observador determina a probabilidade de detecção e localização da anormalidade (<math>v</math>) e a probabilidade de uma imagem falso positiva <math>\{P[FPI]\}</math>. [CHAKRABORTY - 1990].....</i>	<i>28</i>
<i>Figura 12 - Curva ROC com 100% de sensibilidade .....</i>	<i>53</i>
<i>Figura 13 - Curva ROC 100% de especificidade.....</i>	<i>55</i>
<i>Figura 14 - Gráfico da curva ROC dos exemplos do Teste A e B. ....</i>	<i>58</i>
<i>Figura 15 - Curva ROC com relação aos três pesquisadores.....</i>	<i>61</i>
<i>Figura 16 - Curvas ROC dos pesquisadores e do médico.....</i>	<i>66</i>

## LISTA DE TABELAS

<i>Tabela 1 - Número de óbitos e taxa bruta de mortalidade, como também casos novos e taxa bruta de incidência, por 100.000 habitantes no ano de 1997. [INCA,1997]</i> .....	6
<i>Tabela 2 - Número esperado de óbitos e taxa bruta de mortalidade, como também casos novos e taxa bruta de incidência, por 100.000 habitantes no ano de 1998. [INCA,1997]</i> .....	7
<i>Tabela 3 - Grupos de pesquisa analisados pelos tipos de métodos por eles utilizados</i> .....	46
<i>Tabela 4 - Grupos de pesquisa analisados pelos tipos de métodos por eles utilizados</i> .....	47
<i>Tabela 5 - Grupos de pesquisa analisados pelos tipos de métodos por eles utilizados</i> .....	48
<i>Tabela 6 - Grupos de pesquisa analisados pelos tipos de métodos por eles utilizados</i> .....	49
<i>Tabela 7 - Números de vezes que os métodos foram utilizados</i> .....	50
<i>Tabela 8 - 100% de Sensibilidade</i> .....	52
<i>Tabela 9 - 100% de Especificidade</i> .....	54
<i>Tabela 10 - Mudança no critério do observador (exemplo A)</i> .....	56
<i>Tabela 11 - Mudança no critério do observador (exemplo B)</i> .....	57
<i>Tabela 12 - Possibilidades obtidas com os exemplos A e B</i> .....	58
<i>Tabela 13 - Comparação dos resultados encontrados nos três exemplos anteriores</i> .....	59
<i>Tabela 14 - Resultados obtidos pela biopsia e pelos pesquisadores mudando os valores de "threshold" dos classificadores</i> .....	60
<i>Tabela 15 - Média dos resultados encontrados com diversos "threshold"</i> .....	61
<i>Tabela 16 - Média dos resultados obtidos pelo médico e pelos pesquisadores utilizando algoritmos diferentes</i> .....	63
<i>Tabela 17 - Tabela de contingência entre pesquisador P1 e médico</i> .....	64
<i>Tabela 18 - Tabela de contingência entre pesquisador P2 e médico</i> .....	64
<i>Tabela 19 - Tabela de contingência entre pesquisador P3 e médico</i> .....	64
<i>Tabela 20 - Resultados obtidos pela biopsia, pelo médico e pelos pesquisadores, mudando os valores de "threshold" dos classificadores para 200 imagens</i> .....	66
<i>Tabela 21 - Valor médio encontrado pelo médico, pela biopsia e pelos pesquisadores</i> .....	67
<i>Tabela 22 - A concordância para P1 está na tabela 21</i> .....	69
<i>Tabela 23 - A concordância para P2 está na tabela 21</i> .....	69
<i>Tabela 24 - A concordância para P3 está na tabela 21</i> .....	69

# CAPÍTULO 1

## INTRODUÇÃO

### 1.1 MOTIVAÇÃO E OBJETIVO DESTE TRABALHO

Acompanhando as pesquisas em processamento de imagens mamográficas, desenvolvidas pelo grupo @ladim (Laboratório de Imagens Médicas), percebemos os problemas enfrentados pelos pesquisadores que desenvolvem algoritmos para facilitar a detecção de microcalcificações mamárias. Além de terem que separar microcalcificações não muito maiores que ruídos e decidir se elas são associadas a tumores malignos ou benignos, estes pesquisadores enfrentam dificuldades na hora de mostrar seus resultados e os comparar com os apresentados na literatura. Os métodos utilizados convencionalmente impedem uma avaliação exata do desempenho dos algoritmos. Este fato incentiva o surgimento constante de novos algoritmos enquanto a aplicação à rotina dos ambulatórios não é contemplada.

O @ladim está desenvolvendo um sistema computacional que deve auxiliar o diagnóstico do câncer do HU (Hospital Universitário da Faculdade de Medicina de Ribeirão Preto). Este CAD deve ser utilizado principalmente como uma primeira triagem separando os casos evidentemente benignos e malignos dos casos duvidosos que deveriam ser examinados por um médico.

Neste tipo de aplicação os algoritmos devem ser capazes de separar com absoluta precisão os casos benignos. Isto porque é indispensável que, se uma paciente for diagnosticada como não tendo a doença, ela realmente não apresente nenhum índice suspeito.

O problema é evidentemente mais agudo para a determinação dos casos benignos do que para a determinação dos malignos, devido ao custo social que os falsos negativos representam na determinação do câncer mamário. Por outro lado o encaminhamento de todos os casos para biopsia gera um custo elevado para o sistema previdenciário.

É, portanto preferível que o algoritmo gere classes de indeterminadas quando ele não consegue separar com maior precisão as malignas das benignas, do que

classificar todas as imagens.

Entretanto os métodos de avaliação utilizados até agora não permitem que os algoritmos gerem uma classe de indeterminadas. Estes métodos somente analisam os resultados em função das porcentagens de malignas e benignas comparando-os com a biopsia.

O objetivo deste trabalho é, portanto, apresentar um método que permite comparar os algoritmos desenvolvidos pelos pesquisadores contemplando a classe das indeterminadas e possibilita a avaliação da confiabilidade em relação à detecção das benignas e das malignas.

Para alcançar este objetivo analisamos os métodos utilizados pelos grupos que processam imagens médicas para detecção de microcalcificações encontrados na literatura. Após esse levantamento as formas de demonstrarem os resultados foram separadas e classificadas, em seguida, através de exemplos, foram apresentadas as falhas encontradas na maneira de demonstrar os resultados. A nova proposta é a apresentação destes resultados com o auxílio do método estatístico  $\chi^2$  (Qui-quadrado), permitindo levar em conta a classe de imagens indeterminadas, que não era considerada anteriormente.

## **1.2 APRESENTAÇÃO DA DISSERTAÇÃO**

Após esta introdução a dissertação está dividida em cinco capítulos.

O capítulo 2 apresenta dados gerais sobre o câncer de mama, dados estatísticos sobre o câncer no Brasil e sobre o diagnóstico auxiliado por computador. O capítulo 3 traz um estudo sobre os métodos convencionalmente utilizados para avaliação do processo diagnóstico. O capítulo 4 mostra os resultados alcançados com este trabalho. O capítulo 5 traz as conclusões e as continuações desta pesquisa.

O capítulo 6 apresenta as referências bibliográficas.

## **CAPÍTULO 2**

### **DADOS GERAIS SOBRE O CÂNCER DE MAMA**

#### **INTRODUÇÃO**

O câncer de mama é provavelmente ainda hoje o mais temido pelas mulheres devido à sua alta ocorrência e, sobretudo segundo Davies [DAVIES - 1990] pelos seus efeitos psicológicos que afetam a percepção da sexualidade e a própria imagem pessoal da mulher.

Segundo o Ministério da Saúde (Brasil) o câncer de mama é relativamente raro antes dos 35 anos de idade, mas acima desta faixa etária sua incidência cresce rápida e progressivamente.

A detecção precoce é uma das maiores armas contra o câncer, essa detecção pode ser feita pelo auto-exame, o exame clínico e a mamografia. As pesquisas indicam um impacto significativo quando a paciente faz o auto-exame. Em 90% dos casos é a própria mulher quem descobre alterações em sua mama. Quando esse exame é feito regularmente consegue-se detectar tumores primários ainda pequenos e por consequência menor número de linfonodos axilares invadidos por células neoplásicas. Segundo Andrew F. Laine, Jian Fan e Wuhai Yang [LAINE - 1995], estudos feitos em mulheres americanas mostram que quando o câncer de mama é detectado no estágio inicial, a chance de cura é de 93%, mas quando se encontra num estágio intermediário essa taxa diminui para 72% chegando a apenas 18% de sobrevivência quando a doença está em estágio adiantado. Jean Hsu, David M. Chelberg, Charles F. Babbs, Zygmunt Pizlo, and Edward J. Delp [HSU - 1995] também afirmam que uma das maiores causas da morte por câncer de mama é a detecção tardia, devido à expansão das células cancerosas por todo o corpo, a metástase. Segundo o INCA – Instituto Nacional do Câncer, sessenta por cento dos casos identificados em estágios avançados, sendo tumores com mais de cinco centímetros de diâmetro, são incuráveis. Quando é possível remover o tumor, ainda são necessários dez anos de não incidência da doença para que a pessoa seja considerada

curada. A tendência é que pelo menos 10% das vítimas apresentem novamente o problema.

Até os dias de hoje não se sabe ao certo qual a principal causa do desenvolvimento do câncer de mama, mas após anos de estudos [INCA – 1998] constatou-se que alguns fatores contribuem para o desenvolvimento desta doença, como por exemplo, o histórico familiar, exposição à radiação ionizante antes dos 35 anos, a menopausa tardia (além dos 50 anos, em média), a primeira gravidez após os 30 anos, o uso de pílulas com alta dose de estrogênio e também a ingestão de álcool.

Documentou-se também o aumento no risco em mulheres migrantes de áreas de baixo risco para áreas de alto risco; não se sabe ao certo o motivo de algumas áreas possuírem maior número de casos do que em outra. Supõe-se que haja algum tipo de relação com a alimentação, as características genéticas, a qualidade de vida ou outros fatores.

Segundo Françoise Lefebvre, Habib Benali, René Gilles e Robert Di Paola [LEFEBVRE - 1994], a mamografia é considerada ainda hoje um dos mais importantes procedimentos para o rastreamento do câncer impalpável de mama. A sensibilidade da mamografia é alta ainda que, na maioria dos estudos feitos, sejam registradas perdas entre 10 a 15% dos casos de câncer detectáveis ao exame físico. O exame é pouco eficaz em mulheres com menos de 40 anos, devido à alta taxa de tecido parenquimal na mama, por isso não é recomendado em programas maciços de saúde pública, sendo somente indicada para mulheres jovens de alto risco ou com suspeitas de doenças mamárias [INCA - 1997].

O Instituto Nacional de Câncer recomenda que o Exame Clínico das Mamas seja realizado a cada três anos pelas mulheres com menos de 35 anos, a cada dois anos pelas mulheres entre 35 e 39 anos, anualmente pelas mulheres entre 40 e 49 anos. As mulheres na faixa etária entre 50 e 70 anos devem submeter-se ao exame anual ou semestral.

## 1.1 DADOS ESTATÍSTICOS DE CÂNCER DE MAMA

Algumas medidas devem ser tomadas para diminuir a mortalidade do câncer no Brasil como investir em ações de prevenção, detecção e no diagnóstico, para isso é necessário conhecer os dados estatísticos no país. Estes dados são levantados a partir do atestado de óbito do paciente, levando-se em conta a sua procedência ao invés do local de ocorrência do óbito. A partir 1992 estima-se que esse levantamento tenha uma cobertura de 80% do total de óbitos. Esses dados não são completos pela falta de informação das regiões Norte, Nordeste e Centro Oeste.

Os dados estatísticos são muito importantes quando se pretende fazer um combate à doença, por isso o Instituto Nacional do Câncer (INCA) tem como maior objetivo o aperfeiçoamento de um sistema nacional de informação sobre o câncer, que visa obter dados cada vez mais reais da sua distribuição no Brasil, incentivando e apoiando a implantação de novos registros. Para que se possa fazer um controle dos dados foram criadas bases nacionais em cinco cidades em estados diferentes, Belém, Fortaleza, Goiânia, Campinas e Porto Alegre, com o nome de Registros de Câncer de Base Populacional (RCBP). Os RCBP dedicam-se à coleta contínua e sistemática da ocorrência e das características de todos os casos novos de câncer, em uma população geograficamente definida, representando, portanto, registros de incidência de câncer.

Para se fazer um controle, algumas regras devem ser seguidas, como por exemplo, o número estimado de óbitos o qual baseia-se na suposição de que o padrão de mortalidade do período estudado se manterá o mesmo no período estimado. Esse cálculo é feito baseado nos dados do SIM (Sistema de Informação Sobre Mortalidade – Ministério da Saúde), referentes aos óbitos ocorridos no período de 1980 a 1995 e dados dos censos populacionais de 1980 a 1991, os dados do RCBP e complementando esses dados os dados dos 30 hospitais do câncer, ou seja, Registros dos Hospitais do Câncer (RHC). O mesmo raciocínio se dá com o número de casos novos, cujo cálculo se baseia no pressuposto de que a distribuição desses casos, no período divulgado pelo registro, também se manterá no ano da estimativa. Quanto maior for a extrapolação suposta, menor a precisão dos dados estimados, daí a necessidade de se continuar buscando ampliar a cobertura populacional dos registros

de câncer. À medida que estimativas anuais se sucederem, com dados mais completos, poder-se-á contar, no Brasil, com informações cada vez mais próximas da realidade.

Entre os tipos de câncer, o de mama é o tipo mais freqüente, aumentando em 32% a incidência em mulheres [BORING - 1993].

Segundo o American Cancer Society 1992, o esperado é que em cada nove mulheres uma irá desenvolver a doença, essa afirmação tem sido observada pelo National Institutes of Health (NIH).

Nos países ocidentais o câncer de mama representa uma das principais causas de morte entre as mulheres. As estatísticas indicam o aumento de sua freqüência tanto nos países desenvolvidos quanto nos países em desenvolvimento. Segundo a Organização Mundial da Saúde (OMS), nas décadas de 60 e 70, registrou-se um aumento de 10 vezes, em suas taxas de incidência ajustada por idade nos registros de câncer de diversos continentes.

Na década de 80, segundo o INCA, o câncer em geral representava a segunda causa de morte entre a população brasileira adulta. Desde 1987, ele passou a ser a terceira causa de óbitos (12%, em média), sendo que, de 1987 a 1995, aumentou em torno de 28% ao ano.

O número de óbitos estimados para o Brasil no ano de 1997, distribuídos pela localização primária do tumor entre Homens e Mulheres, foi de 97.705 óbitos e de 248.355 novos casos. (Fonte: Ministério da Saúde/INCA, 1997).

A tabela 1 mostra a relação do *câncer de mama* em mulheres no ano de 1997 com seus óbitos e novos casos em relação 100.000 habitantes.

**Tabela 1** - Número de óbitos e taxa bruta de mortalidade, como também casos novos e taxa bruta de incidência, por 100.000 habitantes no ano de 1997. [INCA,1997]

Óbitos	Taxa bruta mort.	Casos novos	Taxa bruta incid.
6.780	9,13	28.310	38,08

O Brasil está classificado entre os países com maior incidência de câncer, em todo o mundo. Porto Alegre apresenta as mais altas taxas brasileiras, entre os homens, seguindo de Fortaleza, Belém, Campinas e Goiânia. No sexo feminino, a

ordem é a mesma, exceto pela taxa maior em Goiânia do que em Campinas (INCA).

No Brasil, o câncer de mama é o que mais causa mortes entre as mulheres. Dos 269.000 novos casos de câncer em geral com previsão de serem diagnosticados em 1998, 140.705 ocorreriam em mulheres (52.3%) e 128.295 em homens (47.7%). Dos quais o câncer de mama será o principal a atingir a população feminina, sendo responsável por 32.695 novos casos, indo a óbito 7.165.

A tabela 2 representa os números esperados de câncer de mama no Brasil no ano de 1998.

**Tabela 2** - Número esperado de óbitos e taxa bruta de mortalidade, como também casos novos e taxa bruta de incidência, por 100.000 habitantes no ano de 1998. [INCA, 1997].

<b>MULHERES</b>			
<b>Óbitos</b>	<b>Taxa bruta mort.</b>	<b>Casos novos</b>	<b>Taxa bruta incid.</b>
7165	9.6	32.695	44.0

Quanto à mortalidade por câncer, estimada para este ano de 1999, é de 107.950 óbitos, dos quais 58.070 (53,8%), entre homens, e 49.880 (46,2%) entre mulheres.

Com base em todos esses dados fica a certeza do quanto é importante investir na melhoria do diagnóstico médico.

Essa melhora pode vir através de um diagnóstico preciso, para isso deve-se investir em sistemas de aquisição de imagens que representam com mais nitidez as estruturas importantes para o diagnóstico ou um sistema Diagnóstico Auxiliado por Computador (CAD), capaz de detectar as anomalias e as classificar em malignas e benignas, fornecendo mais informação aos médicos.

## 1.2 DIAGNÓSTICO AUXILIADO POR COMPUTADOR

### (O SISTEMA “CAD”)

No decorrer das atividades médicas, segundo [SABBATINI-1993] é constante o processo de tomada de decisões ocorrer em diversos pontos da atividade do médico. Algumas delas são bastante simples, como é o caso da interpretação de um resultado de laboratório. Do conjunto das decisões difíceis existem três que podem ser auxiliadas pelo computador. São elas, o diagnóstico, o planejamento terapêutico e o prognóstico. De todas as situações a mais complexa é o diagnóstico. A principal razão para isso é que a análise depende de dados e informações oriundas de fontes de naturezas muito diferentes, incluindo a experiência prévia do médico em realizar diagnósticos do mesmo tipo, bem como o senso comum e a intuição. Entretanto é bastante difícil formalizar esse conhecimento e representá-los através de um programa de computador.

Também os mecanismos mentais e o processo de raciocínio pelo qual os clínicos chegam ao diagnóstico são ainda mal conhecidos. Eles envolvem simultaneamente processos lógicos, avaliações probabilísticas e muitos outros mecanismos ainda entendidos parcialmente. É difícil, portanto tentar simular num computador um modelo incompleto de raciocínio.

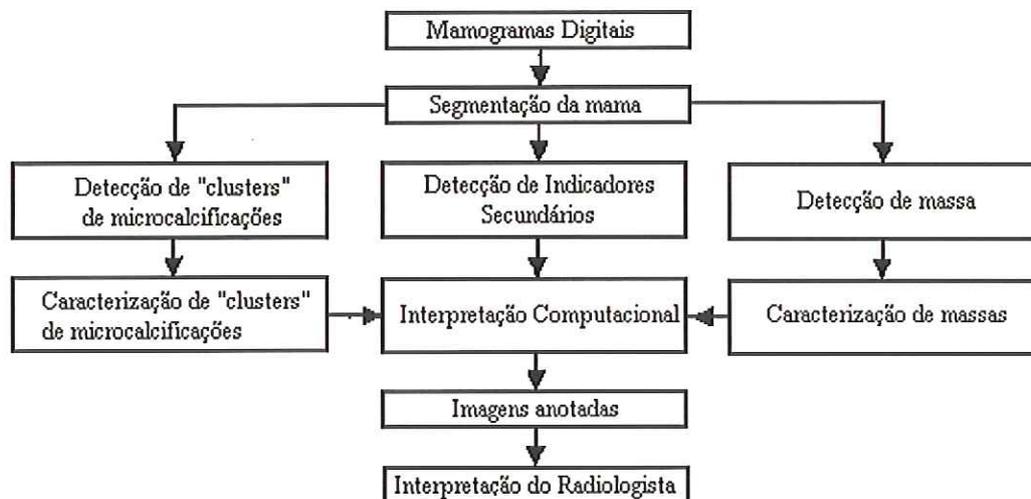
Apesar destas limitações alguns centros de pesquisas conseguiram desenvolver algoritmos para auxiliar médicos radiologistas no diagnóstico do câncer, inclusive o câncer de mama. Esses sistemas são conhecidos como diagnóstico auxiliado por computador (CAD). O primeiro sistema CAD para câncer de mama foi desenvolvido por Franklin e Angerman (1983), usando um sistema dedicado. Esta pesquisa mostrou as limitações da experiência e a importância de se ter uma ajuda no diagnóstico [DOI - 1991]. Segundo Kunio Doi, e outros [DOI - 1995] no sistema CAD a imagem é analisada por um sistema computacional oferecendo uma segunda opinião ao mastologista. O que se tem tentado é obter um sistema que consiga diminuir a taxa de falso positivo e falso negativo, dos exames normais.

Segundo [DOI - 1996], o sistema CAD que também pode ser chamado de “visão aproximada por computador”, para a obtenção dos resultados, emprega

técnicas de processamento de imagens tais como “threshold”, redes neurais, técnicas de transformação utilizando “wavelets”, análise de fractais, entre outras.

Esses métodos têm se mostrado de grande utilidade, mas o que se observa é que raramente os sistemas de auxílio ao diagnóstico têm sido usados rotineiramente na prática clínica. Segundo Renato Sabbatini [SABBATINI – 1993] um exemplo disso é o programa de diagnóstico da etiologia de uma dor abdominal aguda, desenvolvido na Inglaterra, que tem sido usado nas enfermarias de emergência de alguns hospitais. O seu uso ajudou na redução em 50% nas taxas de apendicite perfurada e na redução de 36% para 14% da incidência de cirurgias abdominais desnecessárias.

Na detecção do câncer de mama, o grupo do Dr. Kunio Doi tem se destacado pelos trabalhos publicados. A figura 1 mostra o diagrama (CAD) utilizado pelo grupo para indicar a localização de “clusters” de microcalcificações nas imagens e extrair o sinal verdadeiro de ruídos. [DOI-1991]



**Figura 1** Diagrama em Blocos do Sistema CAD para Mamografias [DOI-1995].

Este sistema CAD apresenta uma sensibilidade em torno de 85 %, com um tempo de processamento de 10 a 20 segundos quando usada uma “workstation”. Para se ter uma noção, num estudo feito pelo grupo, foram digitalizados 60 mamogramas e o CAD apresentou uma taxa de detecção de microcalcificações verdadeiras

positivas (VP) de 87%, com uma razão de detecção falso positivo (FP) de 4 “clusters” por imagem. Os resultados foram apresentados através de uma curva ROC (Receiver Operating Characteristic) que mostrou uma melhora significativa da capacidade dos radiologistas em detectar “clusters” de microcalcificações quando utilizaram o auxílio do computador.

Infelizmente a aplicação destes CAD ainda é limitada tanto pela pouca aceitação do pessoal médico quanto pela pouca disponibilidade dos programas. Muitos deles foram desenvolvidos para fins de pesquisas e não para aplicações práticas o que os tornam difíceis de conseguir. Por outro lado os médicos normalmente desconhecem os benefícios e as potencialidades de uso destes tipos de programas, e existem problemas éticos e legais relacionados ao uso de computadores na decisão médica.

## CAPÍTULO 3

### MÉTODOS CLÁSSICOS DE AVALIAÇÃO DO PROCESSO DIAGNÓSTICO POR PROCESSAMENTO DE IMAGENS

#### INTRODUÇÃO

Por muitos anos, a performance de um sistema diagnóstico foi medida e apresentada em termos de um tipo de “batting average”, ou seja, o percentual da decisão que prova ser correto o diagnóstico. Segundo Metz [METZ – 1978] historicamente, esta medida de performance tem sido chamada de precisão de diagnóstico na literatura médica. Esta precisão ou percentual correto medido, tem suas limitações óbvias que podem influenciar firmemente sobre a prevalência da doença.

Por exemplo, se somente 5% dos pacientes de uma população tem uma doença rara, o sistema pode ter uma precisão de 95% simplesmente se todos os exames forem chamados de negativos. Portanto a porcentagem não revela a relativa frequência de erros falsos positivos e falsos negativos os quais podem ter implicações substanciais no diagnóstico clínico.

A precisão diagnóstica pode ser caracterizada em um ambiente operacional. A curva ROC (Receiver Operating Characteristic) é uma das técnicas conhecidas para comprovar um índice de precisão desejado, em termos de custo benefício. Seu desenvolvimento inicialmente foi motivado por problemas surgidos em radares. Em 1960, as curvas ROC começaram a ser usadas com êxito em psicologia experimental e em psicofísica. Subseqüentemente, foi direcionada para o diagnóstico médico e começou a ser utilizada segundo [METZ - 1989], nos últimos quinze anos, em estudos de imagens médicas.

As curvas ROC são obtidas com base nos resultados dados por aparelhos médicos ou nas imagens que eles produzem, levando-se em conta a eficiência do

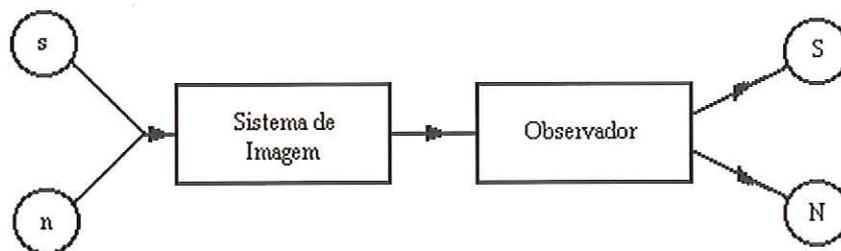
observador. Elas auxiliam a compreensão da exatidão do diagnóstico, porque indicam todas as combinações de sensibilidade, ou seja, o número de decisões verdadeiras positivas por número de casos verdadeiramente positivos. Elas indicam também a especificidade, o número de decisões negativas verdadeiras pelo número de casos verdadeiramente negativos [METZ - 1978]; que um teste de diagnóstico é capaz de fornecer.

Além disto às curvas ROC podem ser usadas como um método para controle de qualidade de imagens médicas, em relação as suas características físicas como contraste, resolução, saturação, etc.

### 3.1 CURVAS ROC

As curvas ROC resultam da apresentação de imagens ao observador, imagens estas que possuem ou não anormalidades e são obtidas por sistemas de registro como, raios X, ultra-som, ressonância magnética, etc. O observador pode ser o próprio médico ou um sistema CAD (Diagnóstico Auxiliado por Computador).

No sistema de imagem da Figura 2 temos:



**Figura 2** - Sistema de detecção por imagem [Evans 1981]

“s” - Existência da anormalidade – presença do sinal no paciente.

“n” - Ruído – inexistência do sinal.

“S” - Resposta do observador: há anormalidade.

“N” - Resposta do observador: não há anormalidade.

Segundo Evans [EVANS-1981] quando a imagem é apresentada ao observador espera-se que ele responda se a anormalidade está presente ou não. A resposta só poderá ser: *Sim* ou *Não*. Se sua resposta for *Sim* (S), significa que ele concluiu que tem anormalidade, se for *Não* (N) que a anormalidade não está presente. Na Figura 3 temos um diagrama envolvendo as respostas possíveis dadas pelos médicos com os estímulos recebidos do paciente:

		Resposta	
		S	N
Estímulo	s	p(S   s) Positivo Verdadeiro	p(N   s) Falso Negativo
	n	p(S   n) Falso Positivo	p(N   n) Negativo Verdadeiro

**Figura 3** - Matriz de estímulo – resposta de um sistema de imagem [Evans 1981]

Analisando o diagrama acima, perceberemos quatro situações possíveis:

**Verdadeiro Positivo (S/s)**; quando o observador responde (S) para uma imagem que contém uma anormalidade. Também podemos considerar que os verdadeiros positivos indicam a *sensibilidade*, do sistema, ou seja, número de decisões verdadeiras positivas por número de casos verdadeiramente positivos.

**Verdadeiro Negativo (N/n)**; quando o observador responde (N) para uma imagem que não mostra a doença. O verdadeiro negativo é um indicativo de *especificidade*, número de decisões verdadeiras negativas pelo número de casos verdadeiramente negativos.

**Falso positivo (S/n)**; quando o observador erradamente responde que o sinal está presente e na verdade a imagem não apresenta nenhum sinal de anormalidade, ou seja, número de falso positivo pelo número de casos negativos verdadeiros.

**Falso negativo (N/s)**; quando o observador responde (N) para uma imagem que apresenta uma anormalidade, ou seja, número de decisões falso negativo pelo número de casos positivos verdadeiros.

Talvez a medida mais simples de qualidade de decisão do diagnóstico seja a fração de casos para os quais o médico esteja correto, essa fração é freqüentemente chamada de precisão. A precisão como um índice de desempenho do diagnóstico é de utilidade limitada, porque a prevalência da doença afeta fortemente os resultados e nenhuma correção matemática pode redimir este índice de um modo significativo. Neste caso a precisão do diagnóstico fica sendo a sensibilidade (Eq. 3.1) e a especificidade (Eq. 3.2):

$$\text{Sensibilidade: } \frac{[\text{Números de decisões verdadeiras positivas}]}{[\text{Números de casos realmente positivos}]} \quad 3.1$$

$$\text{Especificidade: } \frac{[\text{Números de decisões verdadeiras negativas}]}{[\text{Números de casos realmente negativos}]} \quad 3.2$$

A fração de uma certa população estudada que é diagnosticada corretamente é relacionada pela sensibilidade e especificidade pela equação (Eq. 3.3) abaixo:

$$P = \text{sensibilidade} \times \left[ \begin{array}{l} \text{Fração da população que} \\ \text{é realmente positiva} \end{array} \right] + \text{especificidade} \times \left[ \begin{array}{l} \text{Fração da população que} \\ \text{é realmente negativa} \end{array} \right] \quad 3.3$$

A precisão é definida como:

$$P = \frac{\text{Número de decisões corretas}}{\text{Número de casos}} \quad 3.4$$

O pior caso é o do falso negativo, pois o observador responde que não há anormalidade e na verdade o sinal estava presente. Em consequência deste erro a doença pode não ser diagnosticada.

Se o sistema fosse ideal não existiria erro e, quando o observador analisasse uma imagem, ele responderia;

“S” toda vez que contivesse um sinal “s”;

“N” toda vez que não houvesse anormalidade presente na imagem, ou seja “n”.

Portanto, para um sistema ideal, as probabilidades de ocorrência das situações descritas seriam:

$p(S/s) = 1$ , ou seja, o verdadeiro positivo **VP**.

$p(N/n) = 1$ , ou seja, o verdadeiro negativo **VN**.

$p(S/n) = 0$ , ou seja, o falso positivo **FP**.

$p(N/s) = 0$ , ou seja, o falso negativo **FN**.

Se presumirmos que todos os casos analisados foram diagnosticados corretamente como positivos e negativos, o número de decisões corretas mais o número de decisões incorretas devem ser iguais ao número de casos reais. Assim segundo [METZ - 1978] teremos:

$$VP + FN = 1 \quad e \quad VN + FP = 1$$

No sistema real, os valores das probabilidades dessas ocorrências são números entre 0 e 1. O sistema será bom à medida que os valores de acerto ( $S/s$  e  $N/n$ ) estiverem próximos de 1 e os valores de erro ( $S/n$  e  $N/s$ ) estiverem próximos de zero.

Todas essas informações podem ser representadas em um gráfico bidimensional chamado **Curva ROC** (Receiver Operating Characteristic). Uma típica curva ROC é mostrada na Figura 4.

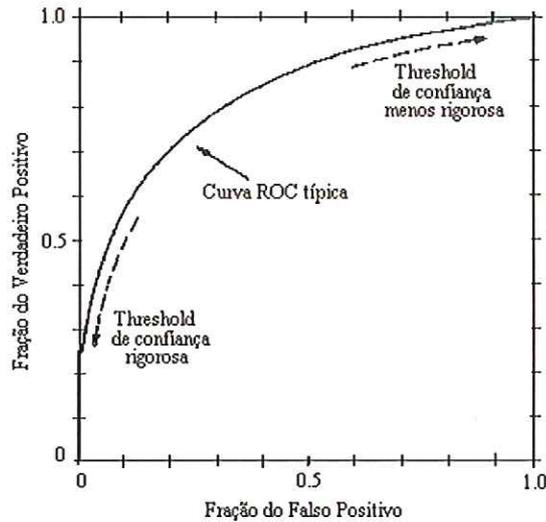


Figura 4 - Curva ROC típica [METZ – 1986]

Quanto maiores forem os valores de VP para valores de FP mais alta será a curva, indicando maior capacidade de discriminação. Ao contrário disso, quanto mais baixa for a curva, menor será a capacidade de discriminação.

### 3.2 GRÁFICO DA CURVA ROC

O critério adotado pelo observador em decidir entre “S” e “N” durante o diagnóstico é o que determina os pontos obtidos no gráfico. Esse critério de seleção é chamado o “threshold” do sistema. O “threshold” pode ser mudado dependendo do critério usado pelo médico.

Estabelecido o critério (“threshold”), relaciona-se à quantidade de verdadeiros positivos em função dos falsos positivos obtidos através dele na análise das imagens. Esses dados representam um ponto na curva ROC [EVANS-1981].

Utilizando um outro critério, faz-se uma nova avaliação da quantidade de acertos e erros, e define-se um novo ponto na curva. Este conjunto de pontos define a **Curva ROC** de um determinado sistema de imagem.

A Figura 5 a seguir mostra como estão relacionados às probabilidades de ocorrência de um verdadeiro positivo com as probabilidades de ocorrência de um falso positivo, para cada um dos critérios adotados pelo observador. O ponto **A** representa a primeira resposta do observador utilizando um determinado critério e o ponto **B** foi gerado após a mudança desse critério.

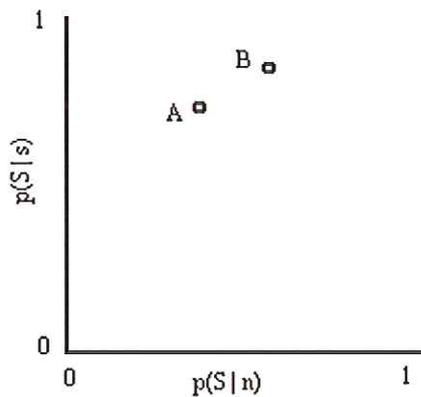


Figura 5 – Informação contida na matriz de estímulos, usando as coordenadas do VP e FP no ponto A. Quando o critério do observador é mudado gera-se o ponto B. [EVANS,1981]

Pela análise da curva ROC, poderemos também avaliar várias etapas do diagnóstico como, por exemplo; o aparelho que foi usado, o receptor (filme, etc), o próprio médico radiologista ou mesmo um sistema de auxílio computacional ao diagnóstico (CAD). Além disto podemos analisar qual classificador apresenta o maior número de acertos e o menor número de erros para o diagnóstico, podendo assim mostrar qual é o melhor classificador para um certo tipo de exame.

Algumas coisas devem ser definidas para essa última análise, como:

- O critério que trará a menor probabilidade de erro.
- O critério que trará a maior quantidade de informação.
- O critério que resultará num menor custo para o paciente.

Alguns critérios são envolvidos na seleção dos pontos de operações nas curvas ROC, como o de máxima informação e o de erro mínimo. Através das curvas de isoinformação, pode-se mostrar os pontos de máxima informação graficamente. Para tanto, o conhecimento da probabilidade de um sinal estar presente  $p(s)$ , é necessário.

Os critérios para a seleção deste ponto de máxima informação não levam em conta os custos dos exames envolvidos no diagnóstico e nem as conseqüências da ocorrência de falsos positivos ou negativos.

Portanto, a eficiência de um diagnóstico médico como um todo pode ser medido considerando simplesmente a detecção ou não da doença através da imagem.

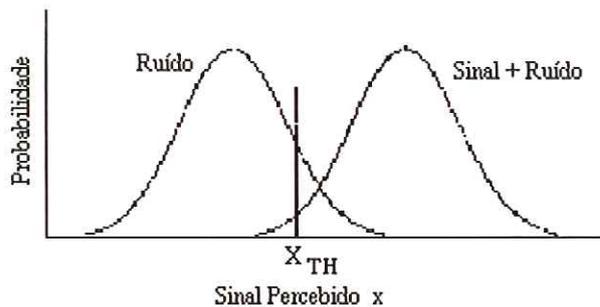
Para um diagnóstico ser preciso temos que levar em conta a sensibilidade e especificidade do sistema. Como já foi mencionada a sensibilidade do sistema é mostrada na taxa de verdadeiro positivo  $p(S/s)$  e a especificidade na taxa do verdadeiro negativo  $p(N/n)$ . Num sistema ideal a sensibilidade e a especificidade devem ser iguais a 1. A sensibilidade do sistema representa a malignidade na imagem, o número de casos verdadeiramente positivos. A especificidade irá representar o número de casos benignos, por isso se faz tão necessário mostrar também este valor, ou seja, as taxas de verdadeiras negativas, são elas que irão representar as pessoas que não possuem a doença e tiveram como diagnóstico uma resposta negativa, ou seja, o paciente não possuía a doença.

Em alguns trabalhos o valor da especificidade é ocultado, mostrando apenas a sensibilidade. Um método considerado bem sucedido pode ter detectado vários casos de positivo verdadeiro, mas para obter esse resultado, muitas pessoas foram encaminhadas para exames desnecessariamente, aumentando a taxa de falso positivo; nesse caso o valor da especificidade seria menor, pois como já foi visto a especificidade é 1 menos a taxa de Falso Positivo, aumentando-se o valor do FP a especificidade diminuirá.

### **3.3 “THRESHOLD” DA CURVA ROC**

Uma curva ROC normalmente pode ser expressa como o resultado da percepção visual representada pela variável aleatória  $X$ , tendo uma distribuição de probabilidade diferente conforme a presença ou não do sinal no sistema de imagem.

O observador decide quando mostrar o sinal de comparação  $X$  com um novo “threshold”  $X_{TH}$ . Como mostrado na Figura 6,  $X$  é o eixo horizontal que representa a certeza do observador de uma imagem que resultou de um observador e  $X_{TH}$  o valor do “threshold”.

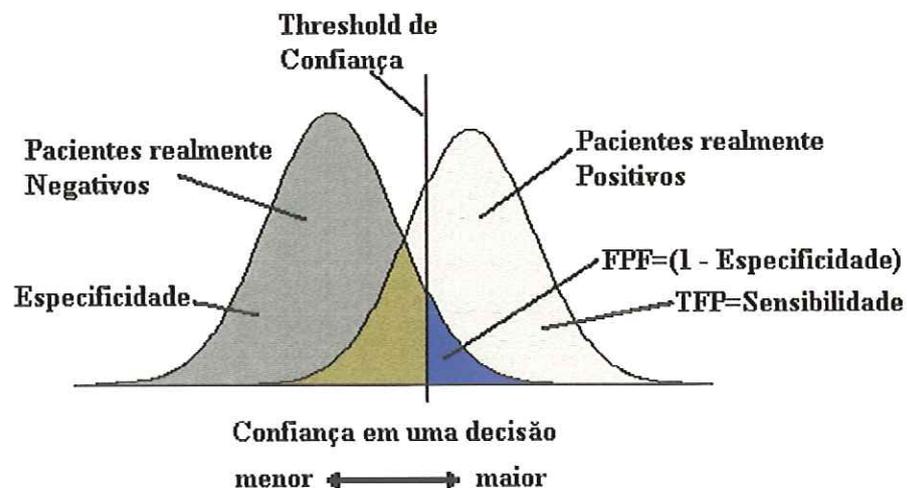


**Figura 6** - Distribuição das probabilidades de um sinal percebido  $X$ , quando introduzido o ruído e o sinal + ruído.  $X_{TH}$  é o valor do “threshold” do sinal percebido sobre qual valor o observador responde Sinal [EVANS, 1981].

Então observando a Figura 6 se  $X \geq X_{TH}$  o observador responde “S” e se  $X < X_{TH}$  ele responde “N”.

A habilidade em fixar os valores do “threshold” é um dos maiores problemas encontrado pelo observador, pois um “threshold” mal escolhido gera erros que são mostrados pela curva ROC. Também altos valores da variável de decisão corresponderão à certeza que estes valores surgiram do sinal de anormalidades. Ao invés disso, valores muito pequenos alertam o observador que estes surgiram de um ruído.

Supostamente as distribuições de probabilidades são Gaussianas (Figura 7) e com a mesma variância. A curva da direita representa pacientes ou imagens realmente positivas, e a da esquerda pacientes ou imagens realmente negativas.



**Figura 7** - Curva representando as probabilidades das distribuições da certeza de um radiologista em um diagnóstico positivo para um particular diagnóstico. O “threshold” é representado pela linha vertical, separando decisões negativas de positivas. [METZ-1986]

O observador pode responder erradamente quando as distribuições se sobrepõem, como mostrado na Figura 7, o que indica um menor nível de certeza, podendo surgir erros para paciente realmente negativo ou positivo.

Conseqüentemente, para o modelo aceito, o observador escolhe entre um diagnóstico positivo e um negativo e compara suas certezas sobre cada imagem com um “threshold” arbitrário. Assim uma imagem é lida como positiva se a certeza do observador em um diagnóstico positivo exceder seu “threshold”.

A Figura 7 mostra este “threshold” como uma linha vertical que divide o eixo horizontal em duas partes. Para esta figura, podemos ver a fração verdadeira positiva, como sendo a fração de pacientes realmente positivos diagnosticada corretamente, representada pela área dentro da distribuição à direita do “threshold”. À esquerda do “threshold” temos a fração de pacientes realmente negativos, ou seja, a taxa de especificidade. Como essas distribuições se sobrepõem, temos também a fração de pacientes realmente negativos, que são incorretamente diagnosticados como positivos, representados pela área azul na Figura 7.

### 3.4 DESVIO PADRÃO

O método de avaliação rende vários pontos no espaço ROC que representam estimativas experimentais dos pontos da curva ROC. O número de casos que podem ser incluídos em algum experimento ROC é limitado por considerações práticas e porque decisões humanas nem sempre são reproduzidas com perfeição. Cada ponto traçado na curva está sujeito a erros estatísticos. Uma curva é então desenhada por meio de aproximação estatística.

O desvio padrão é a variação que pode ser esperada em qualquer ponto traçado da curva sendo que o desvio padrão dos casos verdadeiros positivos ( $DP_{VP}$ ) é dado pela eq. 3.5:

$$DP_{VP} = \sqrt{\frac{VP \times (1 - VP)}{(n. \text{ de casos realmente positivos}) - 1}} \quad 3.5$$

e o desvio padrão dos casos falsos positivos ( $DP_{FP}$ ) é dado pela eq. 3.6:

$$DP_{FP} = \sqrt{\frac{FP \times (1 - FP)}{(n. \text{ de casos realmente negativos}) - 1}} \quad 3.6$$

Note que:

1. O desvio padrão depende da posição de um ponto no espaço ROC, sendo maior quando os valores VP ou FP estão perto de 0,5;
2. O desvio padrão do verdadeiro positivo (VP) é inversamente relacionado ao número de casos realmente positivos usado no experimento;
3. O desvio padrão do falso positivo (FP) é inversamente relacionado ao número de casos realmente negativos usados no experimento.

Embora as precisões de verdadeiro positivo e falso positivo sejam usualmente de igual importância, é habitual tentar usar números aproximadamente iguais de casos realmente positivos e realmente negativos em uma análise da curva ROC.

Esta estimativa da certeza do ponto da curva ROC pode ser usada como um guia na suavização da curva que passa próximo dos pontos tracejados. Frequentemente uma suavização subjetiva da visão humana fornece uma adequada estimativa da curva ROC, quando se aproxima dos pontos traçados.

### 3.5 ÁREA SOB A CURVA ROC

Segundo [METZ - 1986], uma curva ROC representa o compromisso entre sensibilidade e especificidade que podem ser obtidos de um sistema diagnóstico, quando o “threshold” é variado.

A partir do valor da área da curva ROC poderemos comparar os diferentes algoritmos utilizados no diagnóstico, saber quando um é melhor que o outro, ou quando um novo procedimento melhora o desempenho de um algoritmo já estabelecido. Essa área é denotada por  $A_z$ .

A área dentro da curva ROC é uma medida que detecta a presença de um sinal. Se a probabilidade de verdadeiro positivo ( $p(S/s)$ ) se aproxima da unidade para a maioria dos valores de falso negativo ( $p(S/n)$ ), a área dentro da curva ROC aproxima-se também de unidade se o observador estiver correto. Se o sinal não é distinguível do ruído, o observador pode responder “S” ou “N” aleatoriamente e a curva ROC pode ser a diagonal positiva do gráfico. A área dentro da curva é então 0,5 que é a probabilidade do observador responder corretamente.

Em geral, a probabilidade do observador responder corretamente é representada pela equação (3.7) :

$$p_C = p(S/s) p(s) + p(N/n) p(n) \quad (3.7)$$

A área pode ser calculada de três maneiras diferentes:

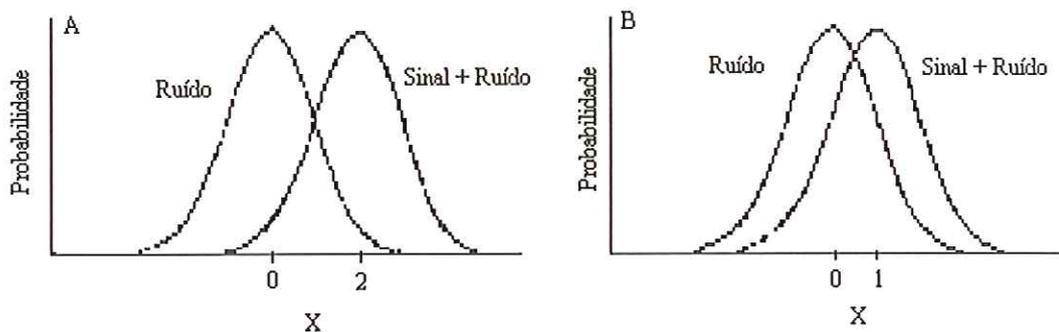
1. Pela regra trapezoidal;
2. Pelo método de avaliação de máxima probabilidade de Dorfman and Alf [DORFMAN - 1969];
3. Pela rampa e interceptação dos dados originais quando traçados num gráfico binomial [HANLEY - 1983].

### 3.6 COMPARAÇÕES ENTRE CURVAS ROC

Em geral curvas ROC altas significam melhor precisão no diagnóstico, por isso elas são muitas vezes usadas como um simples índice “global”, podendo ser interpretadas de várias maneiras intuitivas.

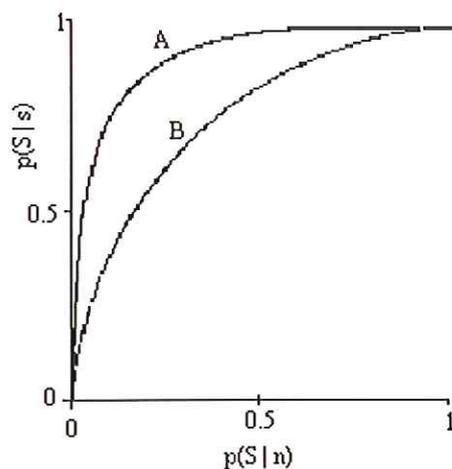
Quanto mais forte for o “sinal” percebido, mais fácil será a decisão, isso quer dizer mais acertos na fração do verdadeiro positivo e menos erro na fração do falso positivo, fazendo com que a curva se aproxime mais do canto esquerdo do gráfico.

Na Figura 8 temos duas curvas, “A” e “B”, fazendo uma análise rápida podemos concluir que o sinal de detectabilidade usado para construir a curva “A” foi melhor que o usado na “B”, pois a curva “A” se aproxima mais do canto esquerdo superior do gráfico.



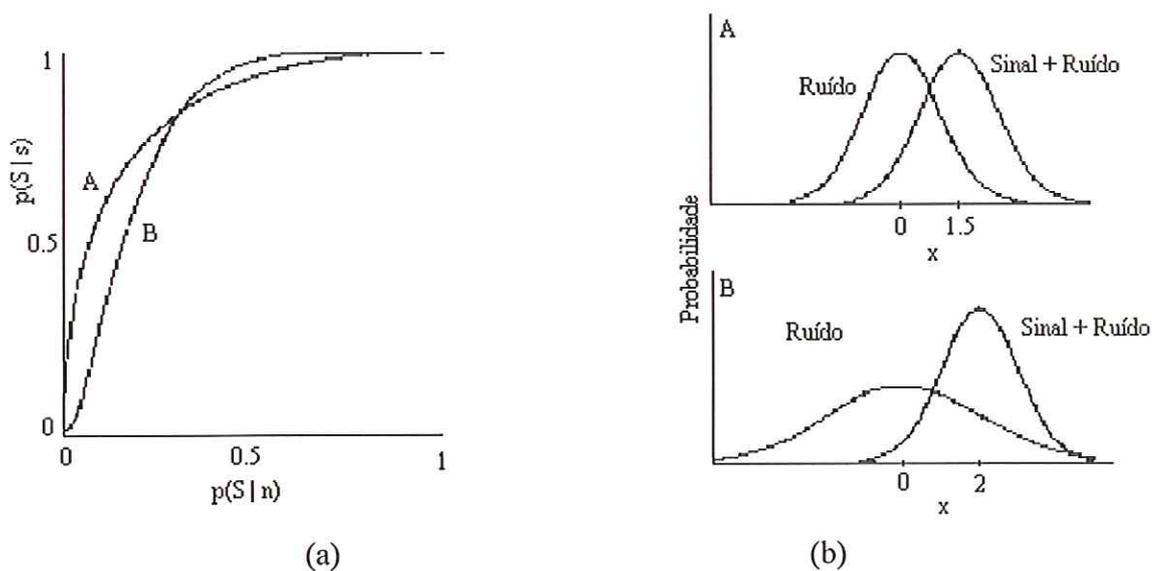
**Figura 8** - Comparação entre duas curvas com diferença de detectabilidade [EVANS, 1981].

A Figura 9 representa a distribuição das probabilidades destas duas curvas “A” e “B”.



**Figura 9** - Distribuição sinal – ruído das duas curvas A e B mencionadas anteriormente [Evans, 1981].

Nem sempre quando se está comparando duas ou mais curvas fica fácil a visualização da qual é a melhor. A Figura 10 a) mostra duas curvas ROC que se interceptam e a 10 b) representa as suas probabilidades.



**Figura 10-** a) duas curvas ROC A e B que se interceptam

b) gráfico das probabilidades destas curvas [EVANS,1981]

Neste momento é preciso saber o que o observador pretende, melhor sensibilidade, ou melhor, especificidade, para determinar qual a melhor curva.

### **3.7 PROCEDIMENTOS DE OBTENÇÃO DA CURVA ROC**

#### **3.7.1 MÉTODO DA TAXAÇÃO**

Vimos até agora que uma curva ROC é então gerada simplesmente pela percepção do observador e este tem que responder sim ou não, a partir do critério por ele selecionado.

Uma outra técnica subjetiva para gerar o chamado de “método de avaliação ou taxação”, foi desenvolvida na psicologia experimental.

Essencialmente, este método representa um compromisso entre aceitar uma resposta sim/não e exigir a seleção de um valor em uma escala contínua representando a certeza de quanto o caso em questão é positivo. O observador ou o sistema de decisão é solicitado para selecionar uma ou várias categorias para representar seu julgamento, baseado na informação vinda do teste de diagnóstico podendo também usar outra informação que ele tenha a respeito do paciente. Estas categorias podem ser definidas como:

1. Definitivamente ou quase definitivamente negativo;
2. Provavelmente negativo;
3. Possivelmente positivo;
4. Provavelmente positivo;
5. Definitivamente ou quase definitivamente positivo.

Estas cinco categorias são utilizadas para representar um razoável compromisso entre a necessidade da análise da curva ROC e a precisão esperada do observador.

A avaliação obtida dessa forma é usada para calcular pontos na curva ROC. Primeiramente, somente as respostas que consideram com certeza que um caso é positivo são contadas como decisões positivas e o resto é considerado como decisões negativas. Assim, para esse tipo de avaliação, somente a quinta resposta das categorias mencionadas acima poderia ser usada como decisão positiva neste estágio da análise de dados. Essas decisões são então comparadas com a presença ou a ausência da doença para cada caso e depois são calculados os valores de VP e FP. Cada combinação de VP e FP é marcada como um ponto da curva ROC podendo ser interpretado como um ponto operante. Neste caso o valor do “threshold” foi bem rigoroso, pois só foram chamados de positivos às decisões que consideram o caso em questão é certamente (ou quase) positivo.

Este procedimento é então repetido interpretando como uma decisão positiva às categorias 5 e 4. Os valores de VP e FP são então calculados e um segundo ponto é marcado no espaço ROC. Este ponto representa um ponto no espaço ROC, o uso de um “threshold” menos rigoroso. Novamente este procedimento é repetido, agora acrescentando a categoria 3, considerando positivas as categorias 5, 4 e a 3.

O critério de decisão é mudado até incluir as cinco categorias como positivo, isto é, até se obter os valores de VP e FP iguais a 1, assim o último ponto operacional sempre está no canto direito superior do gráfico da curva ROC.

### **3.7.2 FROC**

(Free Response Receiver Operating Characteristic)

A metodologia FROC foi introduzida há décadas atrás e mais recentemente estudada por Bunch et al. [BUNCH – 1978] e Chakraborty [CHAKRABORTY – 1990].

O método FROC permite um número arbitrário de anormalidades por imagem e

possibilita que o observador indique o nível de confiança e a localização de toda anormalidade percebida. Este método é o mais próximo de como os radiologistas interpretam mesmo as imagens. Algumas pesquisas mostram que durante a interpretação de um radiologista, dentro das múltiplas decisões algumas são positivas outras negativas, mas muitas são questionáveis. O método convencional ignora esse fato, que é aproveitado no método FROC.

Em uma FROC o observador marca as lesões percebidas numa escala de quatro níveis:

1. Decisão questionável.
2. Possivelmente anormal
3. Provavelmente anormal
4. Definitivamente anormal

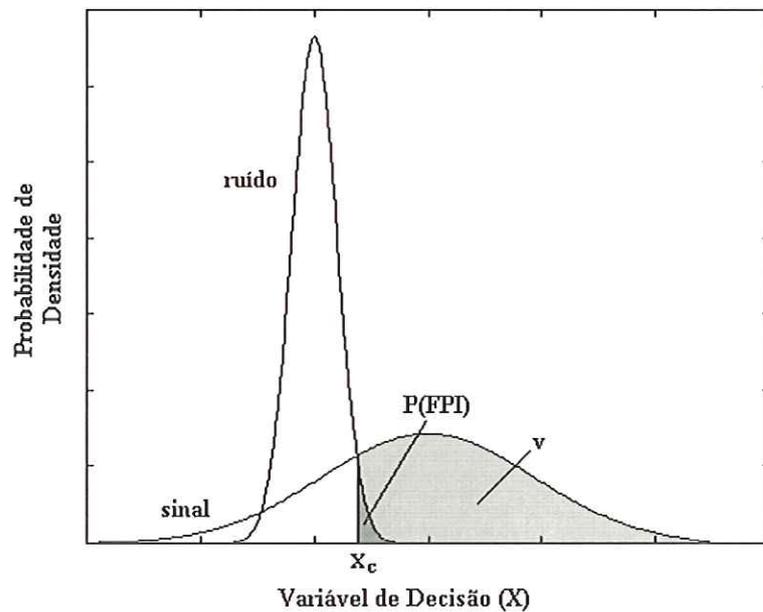
As lesões não marcadas são as definitivamente negativas.

Infelizmente a análise pela FROC não tem um suporte estatístico, por isso sua utilização tem sido limitada.

Em um experimento FROC, as respostas do observador são de dois tipos: respostas verdadeiras positivas (nódulos são localizados com um certo grau de certeza) e respostas falso positiva.

No experimento FROC o gráfico tem na ordenada “ $v$ ” a fração de nódulos com avaliação verdadeira positiva e na abscissa  $\lambda$  a quantidade de respostas falsas positivas por imagem.

Uma imagem é definida falsa positiva (FPI) quando possui uma ou mais respostas falso positivo. O modelo apresentado por [CHAKRABORTY - 1990], consiste de uma variável de decisão “ $x$ ”, representando a distribuição do estímulo de sinal e a distribuição do falso positivo por imagem FPI.



**Figura 11** - FROC - O corte do observador determina a probabilidade de detecção e localização da anormalidade ( $v$ ) e a probabilidade de uma imagem falso positiva  $\{P[FPI]\}$ . [CHAKRABORTY - 1990]

A variável de decisão é uma função de localização de uma imagem. O observador decide marcar a localização, observando se a variável de decisão está excedendo o valor de corte ( $X_c$ ) selecionado por ele.

A área acima de  $X_c$ , abaixo da distribuição de sinal, é denotada por “ $v$ ” (fração de nódulos detectados). Na Figura 11 a parte hachurada corresponde à fração de nódulos corretamente localizados. Para um rigoroso critério de corte  $X_c$ , tendendo a  $+\infty$ , haverá menos sinal percebido, e tomando  $X > X_c$  poderemos obter um valor de “ $v$ ” tendendo a zero.

Para um critério menos rigoroso, quando  $X_c$  tende para  $-\infty$ , a maioria dos eventos observados terá  $X > X_c$ , e “ $v$ ” poderá tender a 1.

A área acima de  $X_c$  (hachurada) dentro da distribuição FPI, denotada por  $P(FPI)$  na Figura 11, dá a probabilidade de gerar um FPI. É visto que o FPI ocorre quando o valor do corte do observador decresce abaixo do mais alto estímulo de ruído gerado na imagem. A distribuição dos FPIs indicam uma idêntica distribuição para estes altos valores.

### 3.7.3 FROCFIT

A falta de uma análise estatística que limita a aplicação do método FROC tem sido motivo para pesquisas nesse assunto. Para isso o grupo do Prof. P. Chakraborty da Universidade da Pensilvânia fez algumas modificações no programa ROCFIT, fazendo com que o programa faça uma análise estatística, após isso o programa passou a se chamar FROCFIT.

### 3.7.4 PORCENTAGEM

Alguns grupos utilizaram para demonstrar seus resultados o número total de casos diagnosticados como verdadeiro, negativo ou indeterminada pelo número total de casos analisados.

## 3.8 TESTE DO QUI-QUADRADO $\chi^2$

Segundo Beiguelman [BEIGUELMAN – 1996], quando é imprescindível comparar a distribuição de certos acontecimentos em diferentes amostras, a fim de avaliar se as proporções observadas desses acontecimentos não mostram diferenças significativas ou se as amostras diferem significativamente, o teste de  $\chi^2$  é indicado.

O teste de  $\chi^2$  é um teste de significância das diferenças entre as proporções comparadas. O  $\chi^2$  é um número com o auxílio do qual se pode testar a hipótese de que os desvios entre as proporções submetidas à comparação podem ser considerado como casuais, contra a de que tais desvios são significativos.

Apesar das proporções observadas serem, às vezes idênticas às esperadas, é mais comum encontrar desvios dos valores teóricos. Essas diferenças são representadas por “*o*” (proporção observada) e “*e*” (proporção esperada). A diferença entre esses valores é o desvio de cada proporção observada e esperada, sendo expressa por “*d*”. Então  $d = o - e$ . Se cada desvio for elevado ao quadrado pela proporção esperada em cada caso o  $\chi^2$  será o somatório de  $d^2/e$ , ou seja:

$$\chi^2 = \frac{(o_1 - e_1)^2}{e_1} + \frac{(o_2 - e_2)^2}{e_2} + \dots + \frac{(o_k - e_k)^2}{e_k} = \sum_{j=1}^k \frac{(o_j - e_j)^2}{e_j} \quad 3.8$$

Se  $\chi^2 = 0$ , as freqüências observadas e esperadas concordam exatamente, mas quando  $\chi^2 > 0$ , isso não se dá. Quanto maior for  $\chi^2$ , maior serão as discrepâncias entre as freqüências observadas e esperadas.

Em estatística, um resultado com probabilidade de ocorrência igual ou inferior a 5% é considerado como tendo pouca probabilidade de acontecer. Com base nisso estabeleceu-se que, a partir de um único experimento, pode-se chegar a uma conclusão a respeito da casualidade ou não dos desvios entre as proporções observadas e esperadas. Em outras palavras, economizando, tempo e trabalho que seriam gastos na repetição de um experimento, pode-se, com base em um único, concluir se os desvios entre as proporções observadas e esperadas devem ter alguma significância ou se eles podem ser considerados como ocorridos casualmente. Na prática, as freqüências esperadas são calculadas com base em uma hipótese “ $H_0$ ” ( hipótese nula ). Se para essa hipótese, o valor calculado de  $\chi^2$  for maior do que alguns valores críticos, deve-se concluir que as freqüências observadas diferem, de modo significativo, das esperadas e rejeitar “ $H_0$ ” no nível de significância correspondente. No caso contrário, deve-se aceitá-la ou, pelo menos, não a rejeitar. Esse processo é denominado teste de qui-quadrado da hipótese ou significância.

Sabendo-se o número total de acontecimentos, e conhecendo a proporção esperada em uma classe, calcula-se a proporção esperada em outra classe. Isso acontece porque os resultados das duas classes não são independentes, não sendo, por isso, independentes cada um dos dois valores de  $(d^2/e)$  que são calculados. A situação de falta de independência entre esses resultados leva a introdução do conceito de grau de liberdade do  $\chi^2$ , que pode ser expresso como o número de classes de resultados menos o número de informações da amostra que são necessários ao cálculo dos valores esperados nessas classes. Esse número é dado por  $v = k - 1$ , quando as freqüências esperadas podem ser calculadas, sem que se façam estimativas dos parâmetros populacionais, a partir de estatísticas amostrais. É utilizado  $v = k - 1 - m$ , quando as freqüências esperadas somente podem ser

calculadas mediante a estimativa de  $m$  parâmetros populacionais, a partir de estatísticas amostrais.

Quando os valores das proporções são tomados em porcentagem, o tamanho amostral é arbitrariamente aumentado (se ele for menor do que 100), ou diminuído (se ele for maior do que 100). Em consequência disso, quando se testa a significância das diferenças entre as proporções observadas e esperadas, elas devem ser tomadas em valores absolutos, a menos que se corrija, posteriormente, o valor do  $\chi^2$  obtido a partir dos percentuais.

Quando os resultados referentes a distribuições contínuas são aplicados a dados discretos, podem ser feitas certas correções relativas à continuidade. Essas correções são denominadas “correções de Yates” [SPIEGEL – 1985]. Essas correções são feitas somente quando o grau de liberdade é igual a 1. Nas grandes amostras essa correção produz o mesmo resultado que o  $\chi^2$  não corrigido. Nas amostras pequenas, nas quais cada frequência esperada está entre 5 e 10, é melhor comparar os valores de  $\chi^2$ , corrigido e não corrigido. Se eles conduzirem a resultados diferentes deverá se recorrer ao aumento do tamanho da amostra.

$$\chi^2_{\text{(corrigido)}} = \frac{(|o_1 - e_1| - 0,5)^2}{e_1} + \frac{(|o_2 - e_2| - 0,5)^2}{e_2} \quad 3.9$$

Na verdade está sendo diminuído 0,5 da diferença absoluta entre valores observados e valor esperado em cada uma das classes, antes de realizar as operações.

O teste do  $\chi^2$  também é aplicado em casos em que não há uma teoria para informar a respeito da probabilidade de ocorrência dos elementos nas diferentes classes esperadas. Quando por exemplo, queremos saber se uma característica qualquer se distribui igualmente em diferentes grupos raciais, sociais, etários, etc, não temos uma teoria que permita calcular os valores esperados nas diferentes classes. Para isso recorreremos a uma tabela de contingência, usualmente construída com o propósito de estudar a relação entre as variáveis de classificação. Na construção de uma tabela quando as frequências observadas ocupam uma única linha, está é denominada tabela de simples entrada. O número de colunas é chamado

de  $k$ , sendo então a tabela de  $l \times k$ . Com a tabela de dupla entrada ou de  $h \times k$ , na qual as frequências observadas ocupam  $h$  linhas e  $k$  colunas. São denominadas tabelas de contingências.

As tabelas de contingências podem ser ampliadas para maiores dimensões, por isso pode haver tabelas de  $h \times k \times l$ , nas quais estão presentes três classificações.

O teste do  $\chi^2$  foi delineado para ser aplicado no estudo da distribuição de caracteres qualitativos, mas existem situações em que ele também pode ser utilizado para análise da distribuição de caracteres quantitativos. Isto ocorre quando existe alguma convenção que permite o agrupamento das medidas desses caracteres em classes.

## **CAPÍTULO 4**

### **DESCRIÇÃO E ANÁLISE DOS RESULTADOS OBTIDOS**

#### **INTRODUÇÃO**

A hipótese que norteou este trabalho, é que a utilização de CAD é limitada principalmente porque não é possível comparar os resultados apresentados pelos pesquisadores e determinar qual é o melhor algoritmo para detectar as microcalcificações. Para comprovar esta hipótese primeiro analisamos todos os trabalhos apresentados na literatura neste assunto e levantamos os métodos utilizados pelos pesquisadores para comprovar seus resultados. A seguir apresentamos três exemplos simplificados para mostrar porque os métodos escolhidos pelos pesquisadores não são adequados. Finalizamos apresentando um método que consideramos mais completo e esclarecedor.

#### **4.1 LEVANTAMENTO DOS MÉTODOS UTILIZADOS PARA APRESENTAR O DESEMPENHO DE ALGORITMOS PARA DETECÇÃO DE MICROCALCIFICAÇÕES**

Os grupos analisados possuem em comum a utilização do CAD para auxiliar o diagnóstico médico a partir da detecção de microcalcificações mamárias e são relacionados a seguir com todos os trabalhos apresentados em revistas ou congressos referentes a esse assunto.

Após feito esse levantamento, construímos as tabelas (Tabelas 3, 4, 5 e 6) relacionando os artigos com a forma utilizada pelos autores para apresentar seus resultados.

Os 107 artigos estão numerados para facilitar a interpretação da tabela, sendo apresentados nas linhas da tabela os 47 grupos selecionados e nas colunas os métodos de análise de resultados por eles utilizados.

Levamos em consideração que alguns grupos utilizam mais de uma maneira para avaliar seus algoritmos.

#### **4.1.1 GRUPOS QUE ESTUDAM A DETECÇÃO DE MICROCALCIFICAÇÃO**

- **Centro Médico Montefiore – Colégio de Medicina Albert Einstein – Departamento de Radiologia - Bronx – NY - USA**

1. [Buchbinder – 1998]

- **Centro Médico e Universidade Georgetown – Departamento de Radiologia – USA**

2. [Hasegawa – 1995]

3. [Loew – 1995]

4. [Shih-Chung – 1996 a]

5. [Shih-Chung – 1996 b]

6. [Shih-Chung – 1998]

- **Centro de Pesquisa Lxi (Leeds Xray Imaging) – Leeds – Yorkshire – UK**

7. [McLeod – 1996]

- **Centro de Pesquisa – David Sarnoff - Princeton - USA**

8. [Sajda – 1996 a]

9. [Sajda – 1996 b]

- **Eastman Kodak Co. - Rochester - NY – USA**

10. [Lure – 1996]

- **Força Aérea Instituto de Tecnologia –Wright Patterson – OH - USA**

11. [McCandless – 1996]

- **Hospital Universidade Graz - Austria**

12. [Sorantin – 1998]

13. [Winkler – 1998]

- **Hospital Universidade Haddasah - Jerusalem - Israel**

14. [Lederman – 1998]

- **Hospital Universidade Nijmegen – Alemanha**

15. [Veldkamp – 1996]

16. [Veldkamp – 1998]

- **INSERM - CHU Pitié-Salpetrière - Paris - França**

17. [Diahi – 1998]

18. [Guillemet – 1996]

- **Instituto Científico - H San Raffaele - Milano - Itália**

19. [Panizza – 1996]

- **Instituto Gustave Roussy - Villejuif - França**

20. [Lefebvre – 1995]

- **Instituto de Pesquisa do Câncer e Hospital Royal Marsden - Londres – Reino Unido**

21. [Davies – 1990]

- **Korea Instituto Avançado de Ciência e Tecnologia – Korea do Sul**

22. [Kim-Jong-Kook – 1997a]

23. [Kim-Jong-Kook – 1997b]

- **Naval Surface Warfare Center Dahigren Division – Advanced Computation Technology Group – VA - USA**

24. [Wallet – 1997]

- **Oak Ridge Natl Lab. – Oak Ridge –TN – USA**

25. [Gleason-Shaun – 1997]

- **Universidade de Agricultura e Tecnologia - Tokyo - Japão**

26. [Kobatake –1998]

- **Universidade da Antwerp – Belgica**

27. [Meersman – 1996]

28. [Meersman – 1998]

- **Universidade do Arizona –Departamento de Radiologia - USA**

29. [Krupinski –1997]

30. [Strickland – 1996]

- **Universidade de Bremen - Alemanha**

31. [Netsch – 1998]

- **Universidade Bilkent - Ankara - Turquia**

32. [Gurcan – 1998]

- **Universidade de British Columbia – Vancouver - Canadá**

33. [Nesbitt –1995]

- **Universidade de Cincinnatti - USA**

34. [Dhawan – 1993]

35. [Chitre – 1994]

36. [Patrick – 1991]

- **Universidade de Chicago – Laboratório Kurt Rossmann para Pesquisas em Imagens Radiológicas - USA**

37. [Anastasio – 1998]
38. [Chan, HP – 1990]
39. [Chan, HP – 1987]
40. [Chan, HP – 1988]
41. [Doi – 1997]
42. [Ema – 1995]
43. [Jiang – 1998]
44. [Jiang – 1996a]
45. [Jiang – 1994]
46. [Jiang – 1996b]
47. [Nagel – 1995]
48. [Nagel – 1998]
49. [Nishikawa – 1994a]
50. [Nishikawa – 1996]
51. [Nishikawa – 1994b]
52. [Nishikawa, - 1994c]
53. [Nishikawa, - 1997]
54. [Nishikawa – 1994d]
55. [Nishikawa, - 1995]
56. [Ping - 1988]
57. [Wu - 1992a]
58. [Wu - 1992b]
59. [Zhang – 1994]
60. [Yoshida –1996a]
61. [Yoshida –1996b]
62. [Yoshida –1996c]
63. [Zhang – 1996]

- **Universidade de Edinburgh – Departamento Ciências da Computação – Escócia**

64. [Hume – 1996]

- **Universidade do Estado do Colorado – USA**

65. [Sadjadi – 1995]

- **Universidade do Estado de Utah – USA**

66. [Heng – 1998]

67. [Heng – 1996]

- **Universidade de Extremadura – Carceres - Espanha**

68. [Cernadas - 1998]

- **Universidade Federal de Pernambuco - Brasil**

69. [Braga - 1995]

- **Universidade GIFU - Japão**

70. [Ibrahim – 1997]

71. [Hara – 1996]

72. [Fukuoka – 1998]

73. [Horino – 1998]

- **Universidade Johns Hopkins – Baltimore - USA**

74. [Bankman – 1993]

75. [Bankman – 1997]

76. [Bankman – 1993]

- **Universidade Kumamoto Colégio de Ciências Médicas – Departamento de Tecnologia Radiológica – Japão**

77. [Higashida – 1992]

- **Universidade de Liverpool - Centro de Pesquisa em Ressonância Magnética e Análise de Imagem – Reino Unido**

78. [Betal – 1997]

- **Universidade de Manchester – Reino Unido**

79. [Bates – 1996]

- **Universidade de Massachusetts Dartmouth – USA**

80. [Chen – 1997]

81. [Oestmann - 1988]

- **Universidade de Michigan – Departamento de Radiologia - USA**

82. [Chan – 1997]

83. [Chan – 1996a]

84. [Chan – 1996b]

85. [Chan – 1995]

86. [Chan – 1998]

87. [Chan – 1987]

88. [Chan – 1994]

- **Universidade Pierre e Marie Curie – Paris - França**

89. [Sehad – 1997]

- **Universidade de Pittsburgh – Departamento de Radiologia – USA**

90. [Chang - 1997]

91. [Chang - 1998]

92. [Shaw – 1998]

93. [Zheng – 1995]

- **Universidade de Salerno – Itália**

94. [Boccignone – 1998]

- **Universidade Santiago de Compostela - Espanha**

95. [Lado – 1997]

- **Universidade de Sydney - Austrália**

96. [Yu – 1996]

- **Universidade de Sint-Andriesziekenhuis – Departamento de Radiologia –  
Bélgica**

97. [De Maeseneer – 1992]

- **Universidade Stavanger - Noruega**

98. [Gulsrud – 1996]

- **Universidade do Sul da Flórida –Departamento de Radiologia- USA**

99. [Clarke – 1994]

100. [Qian – 1995]

101. [Qian – 1994]

102. [Qian – 1998]

103. [Woods – 1993]

104. [Zheng – 1996]

- **Universidade de Surrey – Reino Unido**

105. [Hojjatoleslami – 1996]

- **Universidade Texas A&M – College Station – USA**

106. [Estevez – 1995]

- **Universidade Waltham – MA - USA**

107. [Rosen – 1996]

## **4.1.2 CLASSIFICAÇÃO DOS RESULTADOS**

Com base nesta bibliografia levantamos os métodos utilizados pelos pesquisadores selecionados. Esses métodos são:

- ROC convencional
- Área sob a curva ROC
- FROC convencional - (Free-Response Receiver Operating Characteristic)
- FROC - utilizando o método FROCFIT.
- Sensibilidade ou Verdadeiro Positivo – VP
- Especificidade – Negativo Verdadeiro – VN
- Falso Positivo – FP
- Falso Negativo – FN
- Falso Positivo por Imagem
- Precisão
- Porcentagem

### **4.1.2.1 ROC CONVENCIONAL**

Quando mencionamos o modo ROC convencional estamos relacionando, o gráfico da curva ROC e a área sob a curva. Lembrando que a curva ROC convencional só permite uma resposta do examinador por imagem (Sim/ Não).

### **4.1.2.2 ÁREA SOB A CURVA ROC**

A área sob a curva ROC, comumente denotada por  $A_z$ , é usada como uma medida objetiva da performance do sistema de imagem.

### **4.1.2.3 FROC CONVENCIONAL - (FREE-RESPONSE RECEIVER OPERATING CHARACTERISTIC)**

Os dados com esse método são mostrados em um gráfico, onde o eixo “y” (eixo vertical) representa a fração dos nódulos com avaliação positiva verdadeira, e o eixo

“x” (eixo horizontal), o número de respostas falso negativas por imagem.

#### **4.1.2.4 FROC – utilizando o método FROCFIT**

Programa computacional, desenvolvido pelo Prof. Chakraborty para auxiliar no cálculo da FROC.

#### **4.1.2.5 Sensibilidade ou Verdadeiro Positivo – VP**

Nas Tabelas a seguir os artigos que mostraram seus resultados com a taxa de sensibilidade estão agrupados com os artigos que os demonstraram em forma de Verdadeiro Positivo. Isto significa o quanto o observador acertou nos casos em que o paciente possuía alguma anormalidade, representando a taxa de malignidade.

#### **4.1.2.6 Especificidade – Verdadeiro Negativo – VN**

Uma taxa também muito importante, mas muitas vezes esquecida. É ela que vai indicar a quantidade de pacientes que não possuíam uma anormalidade e foram corretamente observados.

#### **4.1.2.7 Falso Positivo – FP**

O número de casos em que o observador deu um diagnóstico errado, ou seja, o paciente foi diagnosticado como portador de uma anormalidade, quando na verdade ele era um paciente saudável.

#### **4.1.2.8 Falso Negativo – FN**

Número de casos que são diagnosticados como negativo, mas na verdade não são. Esses casos são os piores, muitos pacientes são mandados para casa como saudáveis e depois de um certo tempo voltam com a doença num estágio avançado.

#### **4.1.2.9 Falso positivo (FP) por imagem**

Número de falso positivo encontrado em cada imagem.

#### **4.1.2.10 Precisão**

Números de decisões corretas pelo número total de casos.

#### **4.1.2.11 Porcentagem (%)**

Números de casos diagnosticados malignos e benignos pelo número de caso total.

**Tabela 3** - Grupos de pesquisa analisados pelos tipos de métodos por eles utilizados

	ROC conv.	ÁREA	FROC	Sensibilidade (VP)	Especifidade (VN)	Falso Positivo	Falso Negativo	Precisão	Falso Posit./Imag.	Percentual (%)
C. M. Montefiore – NY- USA		1								
C. M. e U. Georgetown	6	6		2- 3- 4- 5	2-3-	3- 5	3	6	4- 5	
C. M. LXI- Yorkshire				7					7	
C.P. Princeton – USA		8- 9		8- 9		8- 9				
Eastman Kodak – USA								10	10	
Força Aérea – Ohaio – USA				11					11	
Hosp. U. Graz – Austria	13	12- 13		12	12			12		
Hosp. U. Haddasah – Israel		14		14						
Hosp. U. Nijmegen– Alem. França	15	15	16							
INSERM –Paris – França				17- 18	17	18				
Inst. Cient. Milano – Itália	19	19								
Inst Gustave Roussy França										20

Tabela 4 - Grupos de pesquisa analisados pelos tipos de métodos por eles utilizados

	ROC conv.	ÁREA	FROC	Sensibilidade (VP)	Especificidade (VN)	Falso Positivo	Falso Negativo	Precisão	Falso Posit./Imag.	Percentual (%)
Inst. Pesq. Londres - UK				21		21	21			
Korea Inst. Avan.	22	-								
Korea	23									
Naval Surface USA	24									
Oak Ridge Natl Lab USA						25				
U. Agric. T.. Japão									26	
U. Antwerp - Bélgica	27-28									
U. Arizona - USA	29		29	30		30				
U. Bremen - Alem.			31							
U. Bilkent Turquia									32	
U. British Columbia, Canadá				33		33				
U. Cincinnati - USA	35	35		34		34		36		
U. Chicago - USA	38-39-41-46-50-53-58-62-63	41-43-45-46-53-54-59-62-63	40-49-51-56	37-39-40-41-43-44-47-48-49-50-51-52-54-55-57-58-60-61-	43-54-60	38-40-42-48-51-57-58-59-63			37-40-41-42-47-52-55-61	

Tabela 5 - Grupos de pesquisa analisados pelos tipos de métodos por eles utilizados

	ROC conv.	ÁREA	FROC	Sensibilidade (VP)	Especificidade (VN)	Falso Positivo	Falso Negativo	Precisão	Falso Posit./Imag.	Percentual (%)
U. Edinburgh – Escócia	64		64							
U. Colorado USA				65					65	
U. Utah – USA				67		67	67		66	
U. Extremadura – Espanha	68			68					68	
U. Fed. Pernambuco – Brasil										69
U. GIFU – Japão	73			70-71-72	72				70-71	
U. Johns Hopkins – USA	75	75		74-76		74			76	
U. Kumamoto – Japão	77									
U. Liverpool – UK	78	78		78	78					
U. Manchester – UK								79		
U. Massachussets –USA	81	81		80	80					
U. Michigan – USA	82-83-84-85-86-88	83-86		87		87		88	87	
U. P. Marie Curie – França										89

Tabela 6 - Grupos de pesquisa analisados pelos tipos de métodos por eles utilizados

	ROC conv.	ÁREA	FROC	Sensibilidade (VP)	Especificidade (VN)	Falso Positivo	Falso Negativo	Precisão	Falso Positi./ Imag.	Percentual (%)
U..Pitssburgh – USA	92	92		90-91-93		91	90		90-91-93	
U. Salerno – itál.	94			94		94				
U. Santiago de Compostela – Espanha	95									
U. Sydney – Austrália	96	96								
U. Sint-Andriesziekenhuis – Belgica	97			97					97	
U..Stavanger – Noruega				98					98	
U. Sul Florida – USA	99			100-101-102-103-104					100-102-103-104	
U. Surrey – UK				105		105			105	
U. Texas Aadm – USA				106		106				
U. Waltham – USA				107					107	

### 4.1.3 RESUMO DOS MÉTODOS DE AVALIAÇÃO APRESENTADOS NA LITERATURA

A Tabela 7 mostra um resumo das Tabelas 3, 4, 5 e 6 anteriores, fazendo avaliação geral dos 107 artigos que utilizaram cada método.

**Tabela 7** - Números de vezes que os métodos foram utilizados

ROC (Conv.)	35
Área	28
FROCFIT	2
FROC	8
Sensibilidade	59
Especificidade	10
Falso Positivo	26
Falso Negativo	4
Precisão	6
FP/I	33
%	3

A forma mais utilizada pelos pesquisadores para mostrar seus resultados foi à sensibilidade, ou seja, a taxa de verdadeiro positivo.

Trinta e três grupos utilizaram para mostrar seus resultados os valores dos falsos positivos por imagens, não como é mostrado quando se faz o levantamento da FROC, mas como uma quantidade numérica.

Trinta e três grupos utilizaram a curva ROC e através dela demonstraram características como a área. Os demais somente citaram que haviam utilizado a curva ROC obtendo dela uma não significativa diferença entre a análise do radiologista e os resultados do algoritmo.

Dez grupos que apresentaram em forma de especificidade também utilizaram para demonstrar seus resultados os valores da sensibilidade (VP) e área.

Outros métodos bastante utilizados são os números dos falsos positivos (FP),



26 casos. Já a precisão, a taxa de falso negativo foi muito pouco explorada, talvez os pesquisadores não gostem de mostrar o quanto seu classificador errou.

Apenas três grupos utilizaram a porcentagem, mas o grupo de Pernambuco foi o único que mostrou a taxa de acerto das benignas, malignas e também a taxa de duvidosas.

De todos os artigos mencionados, somente um demonstrou seus resultados na forma de sensibilidade, especificidade, falso positivo e falso negativo. Esse artigo era da Universidade de Georgetown e foi o único que mostrou de uma forma completa e clara seus resultados.

## 4.2 EXEMPLOS DE ALTA E BAIXA SENSIBILIDADE

Observamos na tabela 7 que a forma mais utilizada pelos pesquisadores para apresentar os resultados do processamento é a sensibilidade (números de decisões verdadeiras positivas pelo número de casos realmente positivos) com omissão dos dados referentes à especificidade (números de decisões verdadeiras negativas pelo número de casos realmente negativos). Iremos mostrar através de exemplos que essa maneira não é a mais indicada para apresentar os resultados.

Apresentaremos 3 exemplos com os diagnósticos mostrados através da taxa de sensibilidade, especificidade e curva ROC.

**Exemplo 1** – Todos os pacientes são encaminhados para biopsia.

Na tabela 8 consta a descrição de uma população de 1.200 pessoas, sendo que 200 pessoas possuem a doença (“D” uma suposta doença) e 1000 pessoas não a possuem. Supondo que todas elas fossem encaminhadas a uma biopsia, teríamos “T” como resultado, sendo “T+” como o resultado positivo do teste e “T-” o negativo. Dessas 1200 pessoas, as 200 que foram corretamente diagnosticadas como positivas são os verdadeiros positivos (VP). Como todas as pessoas foram encaminhadas para a biopsia não temos nenhum caso de Falso Negativo (FN) e nenhum Verdadeiro Negativo (VN), mas temos 1000 casos de Falso Positivo, ou seja, 1000 pessoas que

não possuíam a doença e desnecessariamente foram encaminhadas para exames.

**Tabela 8 - 100% de Sensibilidade**

Casos reais	Positivo (T+)	Negativo(T-)	Total de casos reais
D+	200(VP)	0(FN)	200
D-	1000 (FP)	0(VN)	1000
<b>Total resultado teste</b>	1200	0	

Com base neste exemplo podemos chamar as probabilidades supostas como sendo:

- ♦ FP como sendo  $P(T + / D -)$  (1 – especificidade)
- ♦ VP como sendo  $P(T + / D +)$  (sensibilidade)
- ♦ FN como sendo  $P(T - / D +)$  (1 – sensibilidade)
- ♦ VN como sendo  $P(T - / D -)$  (especificidade)

Onde  $P(D+)$  é a população sujeita a doença e  $P(D-)$  é a população que não a possui:

$$P(D-) = 1 - P(D+) \quad 4.1$$

Já vimos no Capítulo 3 que  $VP + FN = 1$ , ou seja, substituindo as nomenclaturas pelas do exemplo temos:

$$P(T + / D +) + P(T - / D +) = 1 \quad 4.2$$

$$NV + FP = P(T - / D -) + P(T + / D -) = 1 \quad 4.3$$

Pela Equação 3.3 vista anteriormente temos:

$$P = \text{sensibilidade} \times \left[ \begin{array}{l} \text{Fração da população que} \\ \text{é realmente positiva} \end{array} \right] + \text{especificidade} \times \left[ \begin{array}{l} \text{Fração da população que} \\ \text{é realmente negativa} \end{array} \right] \quad 4.4$$

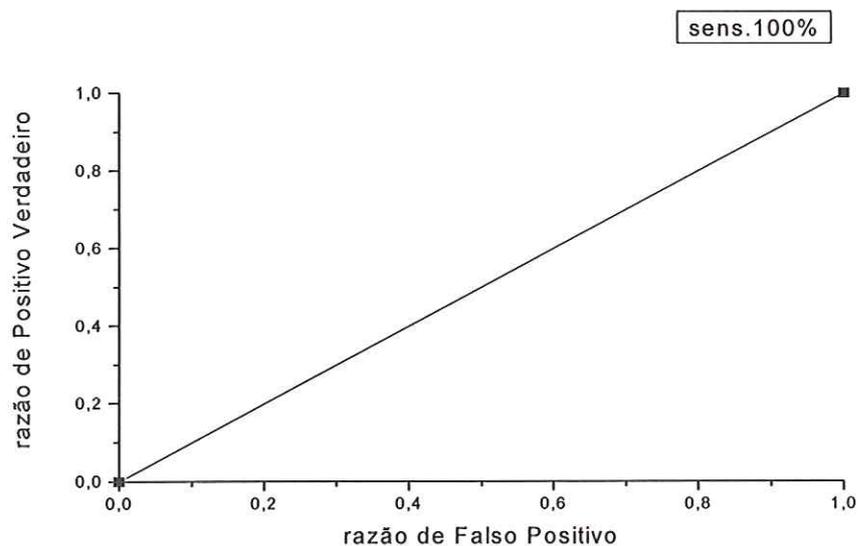
$$P = VP \times P(D+) + VN \times P(D-) \quad 4.5$$

$$P = P(T+/D+) \times P(D+) + P(T-/D-) \times P(D-) \quad 4.6$$

Com base nessas equações os valores das possibilidades são:

- ♦  $(VP) = 200/200 = 1$  sensibilidade
- ♦  $(FN) = 1 - VP = 1 - 1 = 0$
- ♦  $(VN) = 0/1000 = 0$  especificidade
- ♦  $(FP) = 1 - VN = 1$
- ♦  $P(D+) = 200/1.200 = 0,17$
- ♦  $P(D-) = 1 - 0,17 = 0,83$
- ♦  $\text{Precisão} = VP \times P(D+) + VN \times P(D-)$   
 $= 1 \times 0,17 + 0$   
 $= 0,17$

O gráfico da curva ROC utilizando esses dados é dado na figura 12:



**Figura 12** - Curva ROC com 100% de sensibilidade

Neste caso todas as pessoas que possuíam algum tipo de anomalia foram

corretamente diagnosticadas. Isto implica numa alta taxa de sensibilidade, 100%. Poderíamos então concluir que um algoritmo de classificação com esta sensibilidade possui um desempenho excelente, mas como pudemos observar o número de Falso Positivo foi também muito grande, gerando uma precisão muito baixa. Concluimos então que quando não se leva em conta o custo/benefício do sistema, pode-se mostrar seus resultados em forma de sensibilidade, mas quando se quer um equilíbrio entre os dois, esta não é a maneira mais correta. Portanto pesquisadores que mostram seus resultados em forma de sensibilidade, sem associa-la à especificidade, adotam uma forma enganosa para comprovar o desempenho de seus algoritmos.

**Exemplo 2** É o caso inverso ao primeiro exemplo, não se manda ninguém para biopsia.

A população é a mesma que no exemplo 1, só que neste caso ninguém seria diagnosticado como doente. Na tabela (abaixo) consta a descrição de uma população de 1.200 pessoas, sendo que dessa população, 200 pessoas possuem a doença (“D” uma suposta doença) e 1000 pessoas não a possui. Dessas 1200 pessoas, 1000 foram corretamente diagnosticadas como negativas, ou seja, 1000 VN. Como nenhuma foi encaminhada para o teste não temos nenhum caso de Falso Positivo (FP) e Verdadeiro Positivo, mas temos 200 casos de Falso Negativo, ou seja, pessoas que possuíam a doença e não foram encaminhadas para exames.

**Tabela 9** - 100% de Especificidade

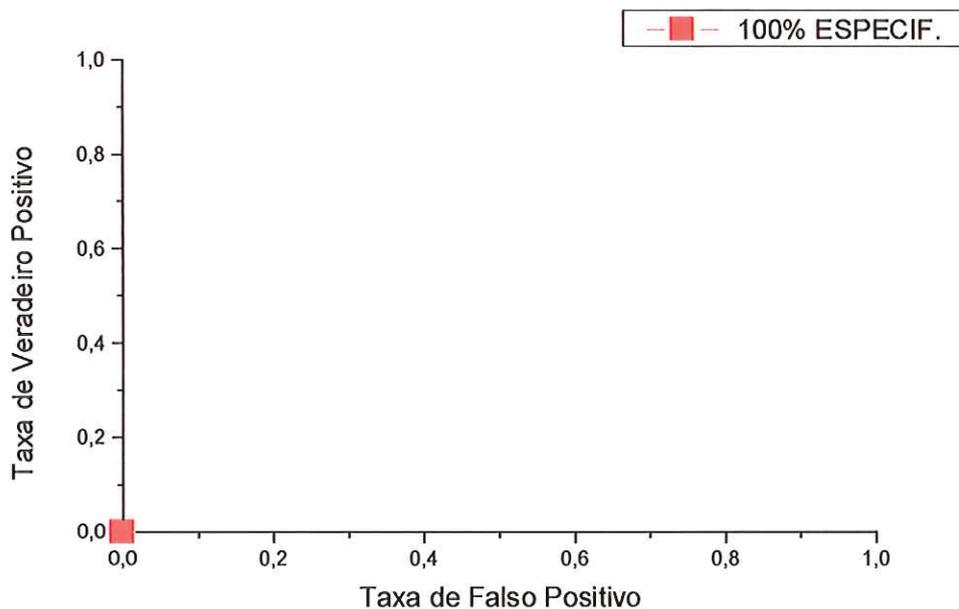
Casos reais	Positivo (T+)	Negativo(T-)	Total de casos reais
D+	0(VP)	200(FN)	200
D-	0(FP)	1000(VN)	1000
Total resultado teste	0	1200	

Calculando os valores das possibilidades temos:

- ♦  $(VP) = \frac{0}{0} = 0$  sensibilidade
- ♦  $(FP) = \frac{0}{0} = 0$

- ♦  $(FN) = 1 - VP = 1 - 0 = 1$
- ♦  $(NV) = 1 - FP = 1 - 0 = 1$  especificidade
- ♦  $P(D+) = 200/1.200 = 0.17$
- ♦  $P(D-) = 1 - 0.17 = 0.83$
- ♦  $Precisão = VP \times P(D+) + VN \times P(D-)$   
 $= 0 + 1 \times 0.83$   
 $= 0.83$

Neste segundo caso, ao contrário do primeiro, a sensibilidade é de 0%, já que nenhum caso positivo foi diagnosticado e a especificidade é de 100%, já que todos os casos negativos foram realmente diagnosticados. A precisão do sistema foi melhor porque o número de pessoas que não possuíam a doença era maior do que as que a possuía. O gráfico da curva ROC que representa esse exemplo é demonstrado a seguir:



**Figura 13** - Curva ROC 100% de especificidade

Um outro exemplo que podemos levar em conta é um caso intermediário o que na maioria das vezes acontece.

**Exemplo 3** – Quando não se usa um critério extremista. Esse exemplo será dividido em duas partes A e B a diferença será a mudança do critério utilizado pelo médico.

- Exemplo A: Tomemos como exemplo a mesma amostra do exemplo 1 e 2, mas mudando o critério de classificação. Na tabela 10 temos a descrição de uma população de 1.200 pessoas, sendo que dessa população, 200 pessoas possuem a doença e 1000 pessoas não a possui. Supondo que em todas fosse aplicado um teste de diagnóstico, teríamos: “T” o resultado do teste de diagnóstico, sendo “T+” como o resultado positivo do teste e “T-” o negativo. Dessas 1200 pessoas, 240 foram diagnosticadas como positivas, sendo 140 Verdadeiros Positivos e 100 Falsos Positivos, e 960 como negativos, sendo 900 Verdadeiros Negativos e 60 Falsos Negativos.

**Tabela 10** – Mudança no critério do observador (exemplo A)

Casos reais	Positivo (T+)	Negativo(T-)	Total de casos reais
<b>D+</b>	140(VP)	60(FN)	200
<b>D-</b>	100(FP)	900(VN)	1000
<b>Total resultado teste</b>	240	960	

Calculando os valores das possibilidades temos:

- ♦  $(VP) = 140/200 = 0.70$
- ♦  $(FN) = 1 - VP = 1 - 0.70 = 0.30$
- ♦  $(FP) = 100/1000 = 0.10$
- ♦  $(VN) = 1 - FP = 1 - 0.10 = 0.90$
- ♦  $P(D+) = 200/1.200 = 0.17$
- ♦  $P(D-) = 1 - 0.17 = 0.83$
- ♦  $Precisão = VP \times P(D+) + VN \times P(D-)$

$$= 0.70 \times 0.17 + 0.90 \times 0.83$$

$$= 0.87$$

Com base nestes valores podemos perceber que os valores de VN são maiores que o do VP, (VN > VP).

- Exemplo B: No mesmo exemplo utilizado anteriormente, mas mudando o critério de decisão 120 pessoas foram diagnosticadas como positivas sendo 80 VP para 40 FP e 1080 pessoas foram diagnosticadas como negativas sendo 120 FN e 960 VN como mostramos na tabela 11.

**Tabela 11** - Mudança no critério do observador (exemplo B)

Casos reais	Positivo (T+)	Negativo(T-)	Total de casos reais
D+	80(VP)	120(FN)	200
D-	40(FP)	960(VN)	1000
Total resultado teste	120	1080	

Calculando os valores das possibilidades temos:

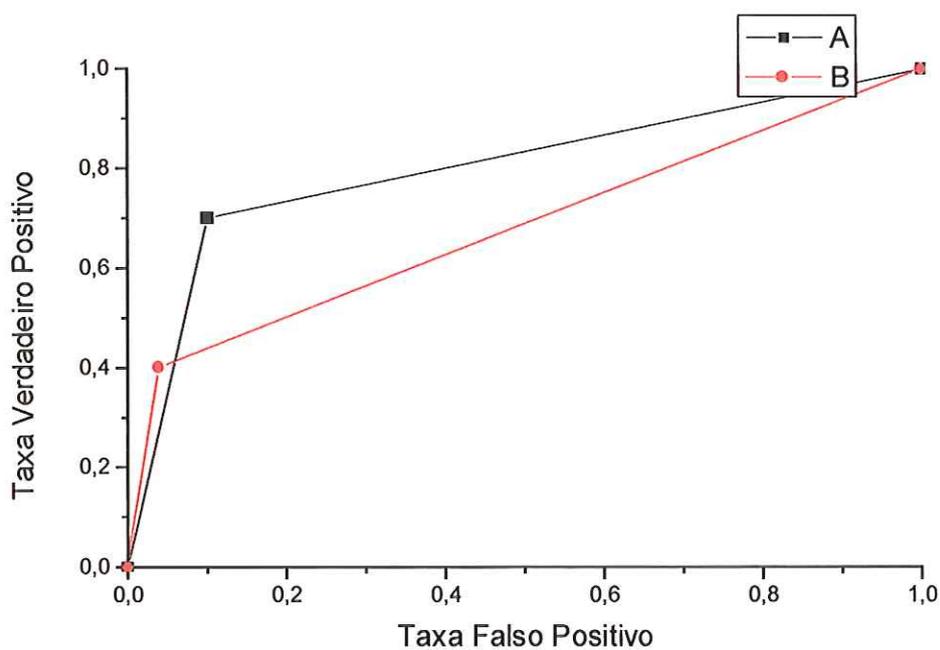
- ♦  $(VP) = 80/200 = 0.40$
- ♦  $(FN) = 1 - VP = 1 - 0.40 = 0.60$
- ♦  $(FP) = 40/1000 = 0.04$
- ♦  $(VN) = 1 - FP = 1 - 0.04 = 0.96$
- ♦  $P(D+) = 200/1.200 = 0.17$
- ♦  $P(D-) = 1 - 0.17 = 0.83$
- ♦  $Precisão = VP \times P(D+) + NV \times P(D-)$   
 $= 0.70 \times 0.17 + 0.90 \times 0.83$   
 $= 0.87$

Na tabela 12 comparamos os valores obtidos nos exemplo 3, casos A e B.

**Tabela 12** - Possibilidades obtidas com os exemplos A e B.

Valores	Exemplo A	Exemplo B
VP( sensibilidade)	0.70	0.40
FP	0.10	0.04
FN	0.30	0.60
VN (especificidade)	0.90	0.96
P(D+)	0.17	0.17
P(D-)	0.83	0.83
Precisão	0.87	0.87

Como podemos notar embora a precisão dos dois testes seja a mesma, as performances dos testes são muito diferentes o que é evidenciada pelas curvas ROC da figura 14.



**Figura 14** - Gráfico da curva ROC dos exemplos do Teste A e B.

Na figura 14, fica claro quando analisamos a área encontrada sob a curva que o teste A é melhor que o teste B.

Para analisar melhor o desempenho dos diversos métodos utilizados para apresentar os resultados dos três exemplos descritos, agrupamos os mesmos na tabela 13.

**Tabela 13** - Comparação dos resultados encontrados nos três exemplos anteriores

Valores	Exemplo 1	Exemplo 2	Exemplo 3A	Exemplo 3B
VP( sensibilidade)	1,00	0	0.70	0.40
FP	1,00	0	0.10	0.04
FN	0	1,00	0.30	0.60
VN (especificidade)	0	1,00	0.90	0.96
P(D+)	0.17	0,17	0.17	0.17
P(D-)	0.83	0,83	0.83	0.83
Precisão	0.17	0,83	0.87	0.87

Analisando a tabela 13, fica claro que a maior sensibilidade foi à encontrada no exemplo 1, mas este exemplo mostrou que nem sempre um alto valor de sensibilidade representa o melhor diagnóstico, o que é mostrado pelos altos valores de Falso Positivo e pela especificidade zero. Isto estará acarretando no custo benefício do exame. No exemplo 2, obtivemos um alto valor de especificidade (Verdadeiro Negativo), mas com um alto índice de casos Falsos Negativos. Lembramos que esse caso é o mais grave, pois pessoas estão sendo diagnosticadas como não portadoras da doença, quando na verdade ocorre o contrário. No exemplo 3 a performance do teste “B” é pior que a do “A” quando se trata dos valores dos positivos verdadeiros, isto é, a sensibilidade. Mas no teste “A” a performance foi pior para a especificidade.

Isto mostra o quanto é difícil apresentar os resultados de forma clara, completa e objetiva, levando em conta o custo-benefício.

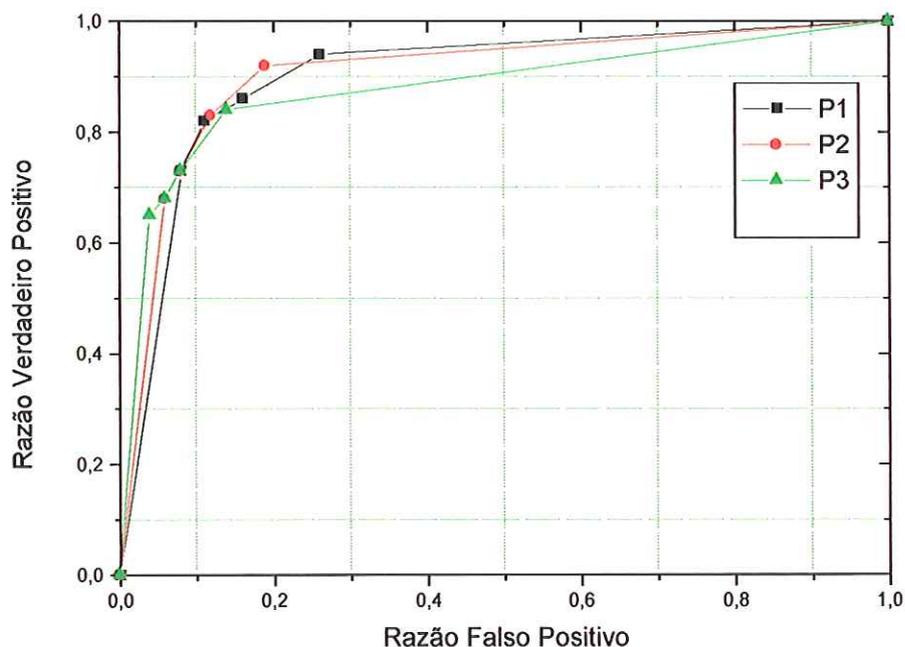
### 4.3 NOVO MÉTODO PARA APRESENTAR OS RESULTADOS DE DIAGNÓSTICOS

Consideramos primeiro o banco de imagens como ele é apresentado tradicionalmente aos pesquisadores que desenvolvem algoritmos para CAD. São 300 imagens, que foram classificadas pela biopsia em malignas e benignas, contendo microcalcificações associadas a tumores mamários. Essas imagens foram analisadas por três pesquisadores ( $P_1$ ,  $P_2$  e  $P_3$ ) que utilizaram algoritmos de processamento diferentes. Neste caso, os classificadores são forçados a decidir entre classes malignas e benignas. Os pesquisadores realizaram uma primeira classificação depois eles adotaram critérios de classificação mais e menos rígidos o que implica mudar os valores do “threshold” dos seus algoritmos, obtendo assim três outros resultados. A tabela 14 apresenta esses dados junto com o resultado da biopsia.

**Tabela 14** - Resultados obtidos pela biopsia e pelos pesquisadores com 4 valores de “threshold” diferentes para os classificadores.

	$P_1$	$P_2$	$P_3$	Biopsia
<b>Benig.</b>	111/126/ 134/138	124/132/ 138/141	129/138/ 141/144	150
<b>FP</b>	39/24/ 17/12	20/18/ 12/9	21/13/ 9/6	0
<b>Malig.</b>	141/129/ 122/109	132/124/ 109/102	126/109/ 102/98	150
<b>FN</b>	9/21/ 27/41	24/26/ 41/48	24/40/ 48/52	0
<b>Total</b>	300	300	300	300

Com esses dados construímos o gráfico da curva ROC, apresentado na figura 15.



**Figura 15** - Curva ROC com relação aos três pesquisadores

Calculando as áreas correspondentes a  $P_1$ ,  $P_2$  e  $P_3$ , mostramos que  $P_2$  possui a maior área. Por este motivo seu algoritmo seria considerado como o melhor.

Entretanto o desempenho do algoritmo pode também ser mostrado pela sensibilidade e especificidade. Para tal finalidade calculamos as médias dos valores da tabela 14, cujos resultados estão sendo mostrados na tabela 15.

**Tabela 15** - Média dos resultados encontrados com diversos “threshold” pelos três pesquisadores.

	$P_1$	$P_2$	$P_3$
<b>Benig.</b>	128	134	138
<b>FP</b>	22	16	12
<b>Malig.</b>	125	116	109
<b>FN</b>	25	34	41
<b>Total</b>	300	300	300

Calculando as taxas de Verdadeiros Positivos (VP), Verdadeiros Negativos (VN), Falsos Positivos (FP) e Falsos Negativos (FN) para cada pesquisador individualmente temos:

- $P_1$  -  $VP = 125/150 = 0,83$   
 $FN = 1 - VP = 0,17$   
 $VN = 128/150 = 0,86$   
 $FP = 1 - 0,86 = 0,14$

A sensibilidade do algoritmo do pesquisador  $P_1$  foi de 0.83 e a especificidade de 0.86.

- $P_2$  -  $VP = 116/150 = 0,77$   
 $FN = 1 - VP = 0,23$   
 $VN = 134/150 = 0,89$   
 $FP = 1 - 0,89 = 0,11$

A sensibilidade do algoritmo do pesquisador  $P_2$  foi de 0.77 e a especificidade de 0.89.

- $P_3$  -  $VP = 109/150 = 0,73$   
 $FN = 1 - VP = 0,27$   
 $VN = 138/150 = 0,92$   
 $FP = 1 - 0,92 = 0,08$

A sensibilidade do algoritmo do pesquisador  $P_3$  foi de 0.73 e a especificidade de 0.92.

Considerando a sensibilidade o algoritmo utilizado por  $P_1$  é o melhor, considerando a especificidade  $P_3$  teve o melhor resultado. Entretanto este pesquisador apresenta muitos falsos positivos o que gera um custo social muito elevado. Nesta análise, podemos ver que os modos tradicionalmente utilizados são conflitantes, mas o fator mais importante é que as imagens que os pesquisadores utilizaram tinham sido classificadas como malignas ou benignas pela biopsia. Portanto são imagens para as quais o médico deu um parecer inicial de malignidade e

encaminhou os pacientes para realizar uma biopsia. Entretanto os sistemas CAD foram desenvolvidos para fornecer uma segunda opinião ao médico, seus resultados devem ser comparados com os resultados deste e não com os da biopsia.

Na biopsia é possível obter informações que não são mostradas pelos aparelhos de raios-X. É, portanto mais apropriado desenvolver algoritmos que processam com confiabilidade os detalhes importantes presentes nas mamografias, mesmo se estes são sutis, do que procurar alcançar resultados parecidos com a biopsia. Consideramos, portanto que uma classificação segura do que é maligno e principalmente do que é benigno, é mais importante do que uma classificação razoável de todas as imagens.

Para satisfazer esse propósito propomos que os algoritmos possam também classificar certas imagens como indeterminadas, mas que, em compensação, classificam as benignas e as malignas com maior rigor. Neste caso pretendemos avaliar os resultados em função dos diagnósticos dados pelo médico que, na primeira análise, também criou uma classe de indeterminadas formada pelas imagens para as quais ele quer mais informações. Sejam estas informações fornecidas pelo prontuário da paciente ou por outro tipo de exame.

No exemplo a seguir os três pesquisadores e o médico analisaram 300 novas imagens e as classificaram em benignas, malignas e indeterminadas. Os resultados obtidos estão na tabela 16 sendo que as colunas  $P_1$ ,  $P_2$  e  $P_3$  apresentam a média dos resultados obtidos por 4 valores de “threshold” diferentes.

**Tabela 16** - Média dos resultados obtidos pelo médico e pelos pesquisadores utilizando algoritmos diferentes.

	$P_1$	$P_2$	$P_3$	M
<b>Malig.</b>	92	94	96	105
<b>Benig.</b>	73	89	65	95
<b>Indeter.</b>	135	117	139	100
<b>Total</b>	300	300	300	300

Com base nestes resultados montamos três tabelas de contingência. A tabela 17 entre os resultados do pesquisador P<sub>1</sub> e do médico, a tabela 18 entre os resultados do pesquisador P<sub>2</sub> e do médico e a tabela 19 entre os resultados do pesquisador P<sub>3</sub> e do médico.

**Tabela 17** - Tabela de contingência entre pesquisador P<sub>1</sub> e médico.

	P <sub>1</sub>		M		Total
	o	e	o	e	
<b>Malig.</b>	92	98.5	105	98.5	197
<b>Benig.</b>	73	84	95	84	168
<b>Indeterm.</b>	135	117.5	100	117.5	235
<b>Total</b>	300	300	300	300	600

**Tabela 18** - Tabela de contingência entre pesquisador P<sub>2</sub> e médico.

	P <sub>2</sub>		M		Total
	o	e	o	e	
<b>Malig.</b>	94	99.5	105	99.5	199
<b>Benig.</b>	89	92	95	92.5	184
<b>Indeterm.</b>	117	108.5	100	108.5	217
<b>Total</b>	300	300	300	300	600

**Tabela 19** - Tabela de contingência entre pesquisador P<sub>3</sub> e médico.

	P <sub>3</sub>		M		Total
	o	e	o	e	
<b>Malig.</b>	96	100.5	105	100.5	201
<b>Benig.</b>	65	80	95	80	160
<b>Indeterm.</b>	139	119.5	100	119.5	239
<b>Total</b>	300	300	300	300	600

Analisando estes resultados pelo teste do  $\chi^2$  visto no capítulo 3 e considerando a hipótese nula (H<sub>0</sub>), ou seja, que a proporção esperada entre as classes; benigna, maligna e indeterminada deve ser independente do observador.

Calculamos portanto os valores dos  $\chi^2$  e observamos se a hipótese nula ( $H_0$ ) é válida para cada pesquisador, considerando o nível de significância de 5%.

Para essas tabelas (17,18 e 19) calculamos que o grau de liberdade é igual a 2, utilizando a equação  $v = (h-1)(l-1)$  e analisamos os valores dos  $\chi^2$  separadamente

$$\text{pela equação } \chi^2 = \sum_{j=1}^k \frac{(o_j - e_j)^2}{e_j} : \quad 3.8$$

$\chi^2_{(2)} P_1 = 8.5$	$0.01 < P < 0.02$
$\chi^2_{(2)} P_2 = 2.10$	$0.30 < P < 0.50$
$\chi^2_{(2)} P_3 = 12.38$	$0.001 < P < 0.01$

Podemos observar que o pesquisador  $P_2$  obteve o valor do  $\chi^2$  menor que o crítico, portanto deve ser aceita a hipótese nula de que a proporção observada não se desvia significativamente da esperada, isto é, a proporção esperada é independente do observador. Já para os pesquisadores  $P_1$  e  $P_3$  a hipótese nula deve ser rejeitada, aceitando a hipótese alternativa de que os desvios entre as proporções observadas e esperadas são significativos.

Nesta etapa estamos propondo que as indeterminadas sejam retiradas da amostra e que os pesquisadores determinem o desempenho dos seus algoritmos pelas curvas ROC, já que o sistema computadorizado não tem acesso às informações suplementares. A análise seria, portanto, realizada em duas partes. A primeira pelo  $\chi^2$  para determinar se o volume das indeterminadas está dentro das expectativas e a segunda pelas curvas ROC sem as indeterminadas.

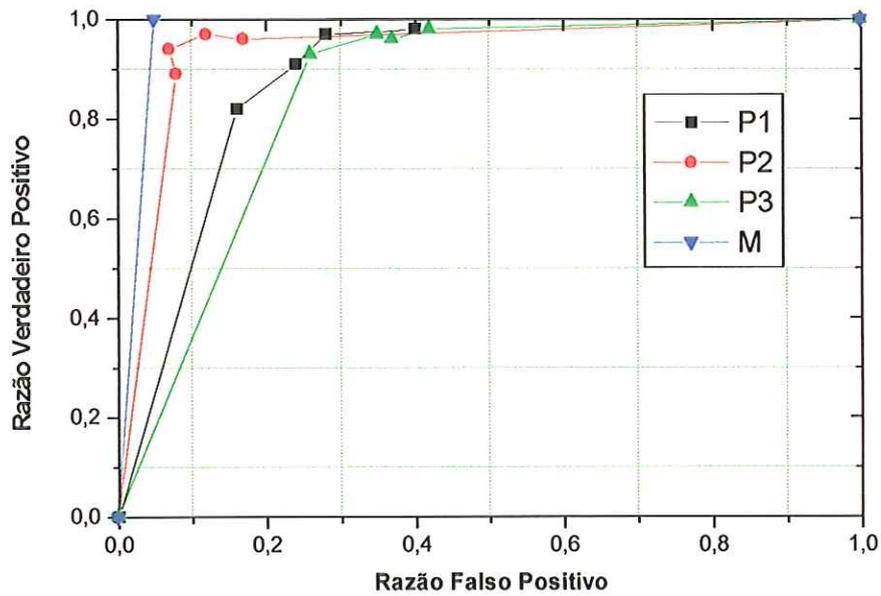
Das 200 imagens restantes, após serem descartadas as 100 imagens classificadas pelo médico como indeterminadas, a biopsia confirmou que 100 eram malignas e o tempo (dois anos) provou que 100 realmente eram benignas.

Mudando os valores do “threshold” utilizados pelos classificadores obtivemos quatro novos valores de benignos e malignos. Comparando esses valores com os resultados da biopsia surgem então os falsos positivos e falsos negativos para cada pesquisador, como mostrado na tabela 20.

**Tabela 20** - Resultados obtidos pela biopsia, pelo médico e pelos pesquisadores, mudando os valores de “threshold” dos classificadores para 200 imagens.

	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	Biop./T	Méd
Malig.	82/91/ 97/98	89/94/ 96/97	93/96/ 98/97	100	100
FN	18/9/ 11/8	11/6 4/3	7/4/ 2/3	0	0
Benig.	84/76/ 72/60	92/93/ 83/88	74/63/ 58/65	100	95
FP	16/24/ 28/40	8/7/ 17/12	26/37/ 42/35	0	5
Total	200	200	200	200	200

Traçando esses pontos no gráfico obtivemos as seguintes curvas ROC:



**Figura 16** - Curvas ROC dos pesquisadores e do médico.

Pelo cálculo das áreas chega-se ao resultado que o pesquisador P<sub>2</sub> foi o que se manteve mais próximo das respostas dadas, tanto pelo médico quanto pela biopsia.

Na tabela 21 apresentamos os valores médios encontrados na tabela 20.

**Tabela 21** - Valor médio encontrado pelo médico, pela biopsia e pelos pesquisadores.

	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	Biop.	M.
<b>Malig.</b>	92	94	96	100	100
<b>FN</b>	8	6	4	0	0
<b>Benig.</b>	73	89	65	100	95
<b>FP</b>	27	11	35	0	5
<b>Tot. Malig.</b>	100	100	100	100	100
<b>Tot. Benig.</b>	100	100	100	100	100

Calculando as taxas de Verdadeiros Positivos (VP), Verdadeiros Negativos (VN), Falsos Positivos (FP) e Falsos Negativos (FN) para cada pesquisador individualmente, obtivemos:

- **P1** -  $VP = 92/100 = 0,92$   
 $FN = 1 - VP = 0,08$   
 $FP = 27/100 = 0,27$   
 $VN = 1 - 0,27 = 0,73$

Portanto, a sensibilidade do algoritmo do pesquisador P<sub>1</sub> foi de 0,92 e a especificidade de 0,73.

- **P2** -  $VP = 94/100 = 0,94$   
 $FN = 1 - VP = 0,06$   
 $FP = 11/100 = 0,11$   
 $VN = 1 - FP = 0,89$

Portanto, a sensibilidade do algoritmo do pesquisador P<sub>2</sub> foi de 0,94 e a especificidade de 0,89.

- **P3** -  $VP = 96/100 = 0,96$   
 $FN = 1 - VP = 0,04$   
 $FP = 35/100 = 0,35$   
 $VN = 1 - FP = 0,65$

Portanto, a sensibilidade do algoritmo do pesquisador P<sub>3</sub> foi de 0,96 e a especificidade de 0,65.

Analisando a sensibilidade vemos que o algoritmo do terceiro pesquisador é o melhor. Entretanto, a geração de uma classe de indeterminadas deve ser compensada por uma precisão aumentada na detecção dos verdadeiros positivos e tolerância baixa para os erros cometidos na detecção dos verdadeiros negativos. Neste caso, com a ênfase dada na especificidade, sem, entretanto desprezar a sensibilidade, o pesquisador P<sub>2</sub> é o melhor.

Precisamos também fazer uma comparação entre as imagens analisadas, observando se todas as imagens diagnosticadas como positivas, negativas ou indeterminadas são as mesmas que o pesquisador classificou como positivas, negativas ou indeterminadas. Suponhamos um médico que analisou 100 imagens, e detectou 40 benignas, 20 malignas e 40 indeterminadas. O classificador pode dar a mesma resposta na média, mas sem que o diagnóstico dado pelo médico para uma imagem corresponde necessariamente a classificação dada pelo pesquisador para esta mesma imagem. Precisamos também comparar os resultados do médico com os da biopsia e o tempo.

Podemos verificar esta condição aplicando um teste de concordância utilizando os dados da tabela 21.

**Tabela 22** - A concordância de P<sub>1</sub> com a Biopsia/Tempo.

<b>Biopsia/Tempo</b>				
<b>P<sub>1</sub></b>		-	+	
	-	73	8	
	+	27	92	
	<b>Total</b>	100	100	

**Tabela 23** - A concordância de P<sub>2</sub> com a Biopsia/Tempo.

<b>Biopsia/Tempo</b>				
<b>P<sub>2</sub></b>		-	+	
	-	89	6	
	+	11	94	
	<b>Total</b>	100	100	

**Tabela 24** - A concordância de P<sub>3</sub> com a Biopsia/Tempo.

<b>Biopsia/Tempo</b>				
<b>P<sub>3</sub></b>		-	+	
	-	65	4	
	+	35	96	
	<b>Total</b>	100	100	

**Tabela 25** - A concordância do médico com a Biopsia/Tempo.

		Biopsia/Tempo	
		-	+
Médico	-	95	0
	+	5	100
	Total	100	100

Construindo essas 4 tabelas de concordância entre os pesquisadores/biopsia e médico/biopsia deveríamos obter os resultados na diagonal negativa mais próximo o possível de 100, em cada células das tabelas correspondentes ao (++) e ao (--), pois foram diagnosticadas pela biopsia/tempo, 100 imagens como benignas e 100 como malignas. Os resultados encontrados nas diagonais positivas são os Falsos Positivos e Falsos Negativos (+ -) e (- +) que deveriam ser de zero.

Obtendo uma coincidência para o  $P_1$  de 73% benignas e 92% malignas, para o  $P_2$  89% benignas e 94% de malignas, para o  $P_3$  65% de benignas e 96% de malignas e para o Médico 95% de benignas e 100% de malignas.

Fica claro que os pesquisadores, embora considerassem a classe das indeterminadas, ainda erraram muito em relação à biopsia. Este erro é mais grave quando gera muitos falsos positivos o que traz um custo social muito elevado.

Mas o que realmente interessa neste trabalho, não é se os pesquisadores conseguirem desenvolver algoritmos eficientes, mas se o método de avaliação é suficiente para deixar claro os benefícios que cada um pode trazer.

Consideramos também a importância de realizar um teste de concordância entre os pesquisadores e o médico com relação à biopsia ou, quando contamos com um banco de imagens que contenha a classe de indeterminadas, entre o classificador e o médico. Para que as respostas dadas pelos pesquisadores tenham a mesma relação perante cada imagem, ou seja, o número de malignas benignas e indeterminadas sejam resultados das mesmas imagens tanto para o médico quanto para os pesquisadores.

## CAPÍTULO 5

### CONCLUSÕES

#### 5.1 CONCLUSÕES GERAIS

A primeira conclusão é que os métodos mais utilizados pelos pesquisadores para apresentar seus resultados, nem sempre são os mais corretos. Pudemos observar, pelo primeiro exemplo, que somente a alta sensibilidade não determina necessariamente o melhor algoritmo. Ela só pode ser considerada suficiente quando não se leva em conta o custo/benefício o que não é possível nas aplicações médicas.

A outra maneira muito utilizada pelos pesquisadores para demonstrar o melhor algoritmo, o cálculo da área da curva ROC, nem sempre está em concordância com os resultados de sensibilidade e especificidade. Nos exemplos apresentados, quando se faz o cálculo desses valores fica claro que o cálculo da área não mostra sempre o melhor algoritmo.

A segunda conclusão é que, para o CAD que está sendo desenvolvido pelo grupo, é mais interessante que o algoritmo não seja forçado a tomar decisões, mas que seja gerada uma terceira classe: as indeterminadas. Esta classe corresponde às imagens de difícil visualização, tanto malignas quanto benignas. Uma classe de indeterminadas também é formada pelos radiologistas quando estes, para fornecer o diagnóstico, recorrem ao prontuário da paciente ou a um exame mamográfico mais específico, ou a outro tipo de exame (por exemplo, o ultra-som).

Entretanto, se forem analisados algoritmos que permitem a classe de indeterminadas é necessário adotar um novo método de avaliação. De fato, se fossemos mostrar os resultados através da curva ROC, a classe das indeterminadas não poderia ser levada em conta, pois esta só considera as classes malignas e benignas, das quais são deduzidos os falsos positivos e os falsos negativos, a sensibilidade e a especificidade.

Como o objetivo do CAD é fornecer um diagnóstico com um maior grau de certeza para a classe Verdadeira Negativa, é possível que o valor da sensibilidade

seja baixo. Entretanto consideramos preferível que o sistema não seja forçado a decidir entre maligno e benigno quando há dúvidas e que, no lugar da sensibilidade, o teste de concordância seja adotado.

A conclusão deste trabalho é que deve ser permitida a classe de indeterminadas e que, neste caso, a avaliação seja feita primeiro sobre o número total de casos, analisando com o teste do  $\chi^2$  se o número de imagens indeterminadas é condizente com o do médico. Depois, sugerimos que as indeterminadas sejam descartadas e que a análise dos resultados seja realizada com o teste de concordância. Este poderá dizer se todas as imagens diagnosticadas como positivas, negativas e indeterminadas pelo pesquisador, são na realidade as mesmas detectadas pelo médico.

Este método, por ser baseado no  $\chi^2$  (Qui-quadrado), é de fácil entendimento pelos radiologistas já que ele é muito utilizado na área médica. Por outro lado o teste de concordância considera os verdadeiros e falsos positivos que são os fatores mais prezados pelos pesquisadores para avaliar seus algoritmos.

## **5.2 CONTINUAÇÃO DESTA PESQUISA**

Mostramos neste trabalho que é necessário permitir que os algoritmos de detecção de microcalcificações possam gerar uma classe de imagens indeterminadas. Entretanto os bancos de imagens à disposição dos pesquisadores não contemplam esta classe, apresentando apenas casos benignos e malignos. Um banco de imagens mais adequado deveria, portanto ser implementado, permitindo que a seleção das imagens mamográficas, realizada num centro de radiodiagnóstico, fosse feita por meio de métodos estatísticos.

## CAPÍTULO 6

### REFERÊNCIAS BIBLIOGRÁFICAS

- [Anastasio – 1998] Anastasio, MA.; Yoshida, H.; Nagel, R.; Nishikawa, RM.; Doi, K. “A genetic algorithm-based method for optimizing the performance of a computer-aided diagnosis scheme for detection of clustered microcalcifications in mammograms”. *Medical Physics*, vol. 25, n . 9, p .1613-20, Sep. 1998
- [Bankman – 1993] Bankman, I. N.; Christens, B.W.A.; King, D.W.; Weinberg I. N.; Gatewwod, O.B.; Brody, W.R. “Algorithm for Detection of Microcalcification Clusters in Mammograms”. *Proceedings of the Annual Conference on Engineering in Medicine and Biology.*, vol. 15 pt 1, 1993. Publ by IEEE, Service Center, NJ., USA., p. 52-53.
- [Bankman – 1993] Bankman, I. N.; Christens-Barry, W. A.; Kim, D. W.; Weinbeg, I. N.; Gatewood, O. B. and Brody, W. R “Automated Recognition of Microcalcification Clusters in Mammograms”, *SPIE*, vol.1905, p. 731-38., 1993.
- [Bankman – 1997] Bankman, I. N.; Nizialek T.; Simon Inpakala; Gatewood, O. B.; Weinbeg, I. N. and Brody, W. R “Segmentation Algorithms for Detecting Microcalcification in Mammograms”. *IEEE – Transactions on Information Technology in Biomedicine.*, vol.1, n.2, p. 141 – 149, Jun. 1997.
- [Bates –1996] Bates, S.P.; Astley, S. M.; Davies, J. D.; Sharp, S. “Three-Dimensional Reconstruction and Shape Classification of Microcalcification Clusters in Breast Lesions”. *3RD International Workshop on Digital Mammography Chicago, Illinois, U.S.A.* June 9-12, p. 239-44, 1996.

- [Beiguelman – 1996] Beiguelman, B. “Curso Prático de Bioestatística”. Revista Brasileira de Genética, 1996
- [Betal – 1997] Betal, D.; Roberts N; Br “Segmentation and numerical analysis of microcalcifications on mammograms using mathematical morphology”. Radiology, vol. 70, n. 837, p. 903-17, Sep. 1997
- [Boccignone – 1998] Boccignone, G.; Chianese, A.; Picariello, A ”Entropy Based Detection of Microcalcifications in Wavelets Space”. ICASSP, IEEE – International Conference on Acoustics, Speech and Signal Processing Proceedings, vol. 5, p. 2713-16, 1998.
- [Boring - 1993] Boring C.C. , Squires T. S. , And Tong T . “Cancer Statistic 1993”, Ca: Cancer J. Clinicians, vol. 43, p. 7 – 26, 1993.
- [Braga - 1995] Braga, Ulisses de Mendonça Neto and Siqueira, Winston Araújo de Neto “A technique for the Automatic Detection and Classification of Mammographic Calcifications Using the Watershed Method”. VIII SIBGRAPI, 1995.
- [Buchbinder - 1998] Buchbinder, S.S.; Leichter, I.S.; Bamberger, P.N.; Novak, B.; Lederman, R.; Fields, S.; Behar, D.J. “ Analysis of clustered microcalcifications by using a single numeric classifier extracted from mammographic digital images”., Academic Radiology, vol. 5, n.11, p. 779 – 784, Nov. 1998.
- [Bunch – 1978] Bunch, PC.; Hamilton, JF.; Sanderson, GK and Simmons AH “A free response approach to the measurement and characterization of radiographic observer performance”. J. Appl. Photogr Eng, vol. 4, p. 166-71, 1978.
- [Cernadas - 1998] Cernadas, E.; Zwiggelaar, R.; Veldkamp, W.; Parr, TC.; Astley,

SM.; Taylor, C.J.; Boggis, CRM. "Detection of mammographic microcalcifications using a statistical model". 4TH International Workshop on Digital Mammography, University of Nijmegen, The Netherlands, p. 205-208, June 7-10, 1998.

[Chakraborty – 1990] Chakraborty, P. D. ; Winter, L.H. "Free-Response Methodology: Alternate Analysis and A New Observer-Performance Experiment". Radiology, vol. 174, p. 873-81, 1990.

[Chan – 1987] Chan, H.P.; Doi, K.; Galhotra, S.; Vyborny, C.J.; MacMahon, H.; Jokich, P.M. "Image Feature Analysis and Computer-Aided Diagnosis in Digital Radiography. I. Automated Detection of Microcalcifications in Mammography". Medical Physics, vol. 14, n. 4, p. 538-48, Jul-Aug. 1987.

[Chan – 1994] Chan, H.P.; Niklason, L.T.; Ikeda, D.M.; Lam, K.L.; Adler, D.D. "Digitization requirements in mammography: effects on computer-aided detection of microcalcifications". Medical Physics, vol. 21, n. 7, p. 1203-11, Jul. 1994.

[Chan – 1995] Chan, H.P.; Lo, Shih-Chung B.; Sahiner, Berkman; Lam, Kwok Leung and Helvie, Mark <sup>a</sup> "Computer-Aided Detection of Mammographic Microcalcifications: Pattern Recognition With an Artificial Neural Network". Medical Physics, vol. 22, n. 10, p. 1555-67, October 1995.

[Chan – 1996] Chan, H. P.; Lo, Shih-Chung B.; Niklason, Loren T.; Ikeda, Debra M. and Lam, Kwok Leung "Image Compression in Digital Mammography: Effects on Computerized Detection of Subtle Microcalcifications". Medical Physics. Vol. 23, n. 8, August 1996.

[Chan – 1996] Chan, Heang-Ping.; Sahiner, Berkman; Petrick, Nicholas; Lam, Kwok L. ; Helvie, Mark A., "Effects of pixel size on classification of microcalcifications on digitized mammograms". SPIE 96, vol. 2710, p.30-

- [Chan – 1997] Chan, H.P.; Sahiner, B.; Petrick, N.; Helvie, Mark A.; Lam, K. L.; Adler, Dorit D. and Goodsitt, Mitchell M. “Computerized Classification of Malignant and Benign Microcalcification on Mammograms: Texture Analysis Using an Artificial Neural Network”, *Medical Physics Biol.*, vol. 42, n. 3, p. 549-67, 1997.
- [Chan – 1998] Chan, HP.; Sahiner, B.; Lam, KL.; Petrick, N.; Helvie, MA.; Goodsitt, MM.; Adler, DD. “Computerized Analysis of Mammographic Microcalcifications in Morphological and Texture Feature Spaces”. *Medical Physics*, vol. 25, n. 10, p.2007-19, October 1998.
- [Chan, HP – 1987] Chan, HP.; Doi, K.; Galhotra, S.; Vyborny, CJ.; MacMahon, H.; Jokich, PM. “Image feature analysis and computer-aided diagnosis in digital radiography. I. Automated detection of microcalcifications in mammography”. *Medical Physics*, vol. 14, n.4, p.538-48, Jul-Aug, 1987.
- [Chan, HP – 1988] Chan HP, Doi K, Vyborny CJ, Lam KL, Schmidt RA “Computer-aided detection of microcalcifications in mammograms. Methodology and preliminary clinical study”. *Invest. Radiol.*, vol. 23, n. 9, p. 664-71, Sep., 1988.
- [Chan, HP – 1990] Chan, HP.; Doi, K.; Vyborny, CJ.; Schmidt, RA.; Metz, CE.; Lam, KL.; Ogura, T.; Wu, YZ.; MacMahon, H. “Improvement in radiologists' detection of clustered microcalcifications on mammograms. The potential of computer-aided diagnosis”. *Investigative Radiology*, vol. 25, n.10, p.1102-10, Oct. 1990
- [Chang - 1997] Chang, YH.; Zheng, B. “Computer-aided detection of clustered microcalcifications on digitized mammograms: a robustness experiment”. *Academic Radiology*, vol. 4 , n. 6, p. 415-8, Jun. 1997.

- [Chang - 1998] Chang, YH.; Zheng, B.; Good, WF. "Identification of clustered microcalcifications on digitized mammograms using morphology and topography-based computer-aided detection schemes. A preliminary experiment". *Investigative Radiology*, vol. 33, n. 10, p. 746-51, Oct. 1998.
- [Chen – 1997] Chen-CH; Lee-GG. "On digital mammogram segmentation and microcalcification detection using multiresolution wavelet analysis". *Graphical Models and Image Processing.*, vol. 59, n. 5, p.349-64, Sep. 1997.
- [Chitre – 1994] Chitre, Y.; Dhawan, Atam P. and Moskowitz, M "Artificial Neural Network Based Classification of Mammographic Microcalcifications Using Image Structure Feature". *Annual-International-Conference of the IEEE Engineering in Medicine and Biology Society Proceedings*, vol. 16, n pt 1, IEEE, Piscataway, NJ, USA, p. 592-593, 1994.
- [Clarke – 1994] Clarke, LP.; Kallergi, M.; Qian, W.; Li, HD.; Clark, RA.; Silbiger, ML. "Tree-structured non-linear filter and wavelet transform for microcalcification segmentation in digital mammography". *Cancer Lett*, vol.1577, n. 2-3, p. 173-81, Mar. 1994.
- [Davies - 1990] Davies, D. H.; Dance, D. R. "Automatic Computer Detection Of Agrupamento Calcification In Digital Mammograms". *Physics In Medicine And Biology*, vol. 35, n. 8, p.1111 – 1118, 1990.
- [Davies – 1990] Davies, D H. and Dance, D R "Automatic Computer Detection of Clustered Calcification in Digital Mammograms". *Medical Physics Biol.*, vol. 35, n.8, p. 1111-18, 1990
- [De Maeseneer – 1992] De Maeseneer, M.; Beeckman, P.; Osteaux, M.; Mattheus, R.; Hoste, M.; Bastaerts, Y.; Jong, B. "Detecting clustered microcalcifications in the female breast: secondary digitized images versus

mammograms". J Belge Radiol., vol. 75, n. 3, p. 173-8, Jun. 1992.

[Dhawan – 1993] Dhawan, Atam P. ; and Moskowitz, Myron “Artificial Neural Network Based Classification of Mammographic Microcalcifications Using Image Structure Feature”. SPIE, Vol.1905, p. 820-31, August, 1993.

[Diahi – 1998] Diahi, JG.; Giron, A.; Brahmi, D.; Frouge, C.; Fertil, B. “Evaluation of a neural network classifier for detection of microcalcifications and opacities in digital mammograms”. 4TH International Workshop on Digital Mammography, University of Nijmegen, The Netherlands, p. 151-56, June 7-10, 1998.

[Doi – 1991] Doi, K.; Giger, M.L.; Macmahon, H.; Hoffman, K.R.; Katsuragawa, S.; Nishikaw, R.M.; Yoshimura, H.; Sanada, S.; Chen, X.; Metz, C.E.; Vyborny, C.J.; Schmidt, R. A.; Montener, S.M.; Matsumoto, T. And Chua, K.G. . “Computer-Aided Diagnosis: Present And Future”.

[Doi – 1995] Doi, K.; Giger, M.L.; Nishikawa, R.M.; Hoffman, K.R.; Macmahon, H.; Schmidt, R. A.; Metz, C.E., “Recent Progress in Development of Computer-Aided Diagnosis (CAD) Schemes In Radiology”. Medical Imaging Technology, vol. 13, n.. 6, p.. 822-35, Nov. 1995.

[Doi – 1997] Doi, K.; Giger, M.L.; Nishikawa, R.M. And Schmidt, R. A “Computer-Aided Diagnosis of Breast Cancer on Mammograms”. The Journal of the Japanese Breast Cancer Society, vol. 4, n. 4, p. 228-32, Dec. 1997.

[Doi – 1997] Doi, K.; Giger, M.L.; Nishikawa, R.M.; Schimidt, R. <sup>a</sup> .. “Computer – Aided Diagnosis of Breast Cancer on Mammograms”. Breast Cancer , vol. 4, n. 4, p. 228-33. December 1997.

[Dorfman – 1969] Dorfaman, D.D.; Alf, E. Jr “Maximum-Likelihood Estimation of Parameters of Signal-Detection Theory and Determination of Confidence

Intervals – Rating-Method Data”. *Journal of Mathematical Psychology*, vol. 6, p. 487-96, 1969.

[Ema – 1995] Ema T, Doi, K.; Nishikawa, Robert M.; Jiang, Y.; Papaioannou, J .  
“Image Feature Analysis and Computer-Aided Diagnosis in Mammography: Reduction of False-Positive Clustered Microcalcifications Using Local Edge-Gradient Analysis”, *Medical Physics*, vol.22, n.2, p.161-69, Feb. 1995.

[Estevez – 1995] Estevez-Leonardo-W; Kehtarnavaz-Nasser-D. “Computer assisted enhancement of mammograms for detection of microcalcifications”. *IEEE-Symposium-on-Computer-Based-Medical-Systems.*, p. 16-23, Los Alamitos, CA, USA 1995.

[Evans – 1981] Evans, A LI “The Evaluation of Medical Images”. *Medical Physics Handbooks* 10, 1981.

[Fukuoka – 1998] Fukuoka, D.; Kasai, S.; Fujita, H.; Hara, T.; Kato, M.; Endo, T.; Yoshimura, H “Automated detection of clustered microcalcifications on digitized mammograms”. *4TH International Workshop on Digital Mammography*, University of Nijmegen, The Netherlands, p. 197-200, June 7-10, 1998.

[Gleason-Shaun – 1997] Gleason-Shaun-S; Sari-Sarraf-Hamed; Hudson-KT; Hubner-K “Higher accuracy and throughput in computer-aided screening of mammographic microcalcifications”. *IEEE - Nuclear Science Symposium and Medical Imaging Conference*. Vol. 2, IEEE, Piscataway, p. 1464-1468, 1997.

[Guillemet – 1996] Guillemet, H.; Benali, H.; Kahn, E.; Paola, Robert Di. “Detection and Characterization of Microcalcifications in Digital Mammography”. *3RD International Workshop on Digital Mammography* Chicago, Illinois,

U.S.A. June 9-12, p. 225-30, 1996.

[Gulsrud – 1996] Gulsrud, Thor ° “ Automated detection of clusters of microcalcifications using a multichannel filtering approach”. In International Symposium on Digital Signal Processing 1996, London, UK, 22-25 July 1996, p.16-21.

[Gurcan – 1998] Gurcan, MN.; Yardimci, Y.; Cetin, AE. .. “Microcalcifications detection using adaptive filtering and gaussianity tests”. 4TH International Workshop on Digital Mammography, University of Nijmegen, The Netherlands, p. 157-64, June 7-10, 1998

[Hanley – 1983] Hanley, J. A .And Macneil, B “A Method of Comparing the Areas Under Receiver Operating Characteristic Derived From the Same Cases”. Radiology, vol. 148. p. 839-43, Sep. 1983.

[Hara – 1996] Hara, T.; Hirako, K.; Fujita, H.; Endo, T.; Horita, K.; Ikeda, M.; Kido, C.; Ishigaki, T. “Automated Detection Algorithm for Clustered Microcalcifications Based on Density Gradient and Triple-ring Filter Analysis”. 3RD International Workshop on Digital Mammography Chicago, Illinois, U.S.A. June 9-12, 1996.

[Hasegawa -1995] Akira Hasegawa, Yuzheng C. Wu, Matthew T. Freedman, and Seong K. Mun.. “Adaptive-sized neural networks-based computer-aided diagnosis of microcalcifications”. Proceedings SPIE 1995, vol. 2434, p.557-562.

[Heng – 1996] Heng, C.; Lui, Yui M.; Freimamis, R.I “A New Approach to Microcalcification Detection in Digital Mammograms”. IEEE – Nuclear Science Symposium and Medical Imaging Conference, vol. 2, p. 1094-98, 1996.

- [Heng – 1998] Heng, C.; Lui, Yui M.; Freimam, R.I “Novel approach to microcalcification detection using fuzzy logic technique”. IEEE – Transactions on Medical Imaging, vol. 17, n. 3, p. 442-450, Jun. 1998
- [Higashida – 1992] Higashida, Y.; Moribe, N.; Morita, K.; Katsuda, N.; Hatemura, M.; Takada, T.; Takahashi, M.; Yamashita, J. “Detection of subtle microcalcifications: comparison of computed radiography and screen-film mammography”. Radiology, vol.183, n. 2, p. 483-6, May. 1992.
- [Hoffmann - 1993] Kenneth R. Hoffmann, Heber Macmahon Kunio Doi Charles Metz Lian Yao And Katsumi Abe “Evaluation Of An Enhanced Digital Film – Duplication System By Receiver Operating Characteristic Analysis”., Investigative Radiology, vol. 18, n. 12, p. 1134-1138. 1993.
- [Hojjatoleslami – 1996] Hojjatoleslami, S. A. and Kittler, J. “Detection of clusters of microcalcification using a K-nearest neighbour rule with locally optimum distance metric”. 3RD International Workshop on Digital Mammography Chicago, Illinois, U.S.A., June 9-12, p. 267-72, 1996.
- [Horino – 1998] Horino, K “Computer-aided diagnosis in the detection of simulated clustered microcalcifications on mammography”. Nippon Igaku Hoshasen Gakkai Zasshi , vol. 58, n. 7, p. 343 – 348, Jun. 1998.
- [Hume – 1996] Hume, A.; Thanisch, P.; Hartswood, M. “On The Evaluation of Microcalcification Detection Algorithms”. 3RD International Workshop on Digital Mammography Chicago, Illinois, U.S.A. June 9-12, p. 273-76, 1996.
- [Ibrahim – 1997] Ibrahim, N.; Fujita, H.; Hara, T.; Endo, T. “Automated Detection of Clustered Microcalcifications on Mammograms: CAD System Application to MIAS Database”. Medical Physical Biol., vol. 42, n. 12, December 1997.

- [INCA – 97-98] Ministério Da Saúde. Instituto Nacional De Câncer. Coordenação De Programas De Controle Do Câncer " O Problema Do Câncer No Brasil", Quarta Edição Revisada e Atualizada. Rio De Janeiro, 1997.
- [Jiang – 1994] Jiang, Yulei; Nishikawa, Robert M.; Wolverton, Dulcy E.; Giger, Maryellen L.; Doi, Kunio; Schmidt, Robert A. and Vyborny, Carl J. "Mammographic Feature Analysis of Clustered Microcalcifications for Classification of Breast Cancer and Benign Breast Diseases". Annual International Conference of the IEEE – Engineering in Medicine and Biology Society Proceedings, vol. 16, n pt 1, p. 594-595, 1994.
- [Jiang – 1996] Jiang, Y.; Nishikawa, Robert M.; Metz, Charles E.; Wolverton, Dulcy E.; Schmidt, Robert A.; Papaioannou, John and Doi Kuni "A Computer - Aided Diagnostic Scheme for Classification of Malignant and Benign Clustered Microcalcifications on Mammograms". 3RD International Workshop on Digital Mammography Chicago, Illinois, U.S.A. June 9-12. P. 219-24, 1996.
- [Jiang – 1996] Jiang, Yulei; Nishikawa, Robert M.; Wolverton, Dulcy E.; Metz, Charles E.; Giger, Maryellen L.; Schmidt, Robert A.; Vyborny, Carl J. and Doi K "Malignant and Benign Clustered Microcalcifications: Automated Feature Analysis and Classification", Radiology, vol.198, n. 3, p. 671-78. March 1996.
- [Jiang – 1998] Jiang, Y.; Nishikawa, RM.; Schmidt, RA.; Metz, CE.; Giger, ML.; Doi, K "Benefits of computer-aided diagnosis in mammographic diagnosis of malignant and benign clustered microcalcifications". 4TH International Workshop on Digital Mammography, University of Nijmegen, The Netherlands, p. 215-20, June 7-10, 1998.
- [Kim-Jong-Kook – 1997] Kim-Jong-Kook; Park-Hyun-Wook. "Surrounding region dependence method for detection of clustered microcalcifications on

mammograms”. IEEE-International-Conference-on-Image-Processing. vol. 3, p.535-38, IEEE Comp Soc, Los Alamitos, CA, USA, 1997.

[Kim-Jong-Kook – 1997] Kim-Jong-Kook; Park-Jeong-Mi; Song-Kung-Sik; Park-Hyun-Wook. “Texture analysis and artificial neural network for detection of clustered microcalcifications on mammograms”. Neural for Signal Processing Proceedings of the IEEE – Workshop., p. 199-206. IEEE, Piscataway, NJ, USA. 1997.

[Kobatake –1998] Kobatake, H.; Takeo, H.; Nawano, S. “Microcalcification detection system for full-digital mammography”. 4TH International Workshop on Digital Mammography, University of Nijmegen, The Netherlands, p. 201-204, June 7-10, 1998.

[Krupinski –1997] Krupinski E.A.; Nishikawa RM “Comparison of eye position versus computer identified microcalcification clusters on mammograms”. Medical Physics. vol. 24, n.1, p.17-23, Jan 1997.

[Lado – 1997] Lado, M. J.; Tahoces, PG.; Souto, M.; Mendez, AJ. and Vidal, JJ. “Real and Simulated Clustered Microcalcifications in Digital Mammograms. ROC Study of Observer Performance”., Medical Physics, vol. 26, n. 9, p. 1385-94, September 1997

[Laine – 1995] Laine, Andrew; Fan, J. And Yang Wuha “Wavelets For Contrast Enhancement Of Digital Mammography, IEEE Engineering In Medicine And Biolofy, p. 536-50, Sep/Out., 1995.

[Lederman – 1998] Lederman, R.; Leichter, I.; Bamberger, P.; Novak,B.; Fields, S.; Buchbinder, B. “Interactive quantitative characterization of microcalcifications on digitized film-screen mammograms”. 4TH International Workshop on Digital Mammography, University of Nijmegen, The Netherlands, p. 221-24, June 7-10, 1998.

- [Lefebvre – 1994] Lefebvre, F.; Benali, H.; Gilles, R. and Paola, R. D “A simulation model of clustered breast microcalcifications”. *Medical Physics*, vol. 21, p. 1865-74, Dez. 1994.
- [Lefebvre – 1995] Lefebvre, F.; Benali, H.; Gilles, R.; Kahn, E.; Di Paola, R. “A fractal approach to the segmentation of microcalcifications in digital mammograms”. *Medical Physics*, vol. 22, n. 4, p.381-90, Apr. 1995.
- [Loew-1995] Loew, Murray H. and Mia, Rashidus S. “Detection of Microcalcifications in Mammograms Using Eyetrack Data”. *IEEE International Conference on Image Processing*, vol. 3, p. 145-48, 1995.
- [Lure - 1996] Lure, Fleming Y.; Gaborski, Roger S.; Pawlicki, Thaddeus F.. “Application of neural-network-based multistage system for detection of microcalcification clusters in mammogram images”, *SPIE vol.2710*, p.16-23, 1996.
- [McCandless-Donald-1996] McCandless-Donald - A.; Rogers- Steven-K.; Hoffmeister-Jeffrey-W-MD; Ruck-Dennis-W; Raines-Richard-A; Suter-Bruce-W. “Wavelet detection of Microcalcifications”. *Proceeding-of- SPIE-The International Society for Optical Engineering.*, vol. 2762 , p. 388-99, Society of Photo Optical Instrumentation Engineers, Bellingham, WA, USA, 1996.
- [McLeod-1996] G. McLeod, G.; GJS. Parkin, GJS.; Cowen, AR. “Automatic Detection of Clustered Microcalcifications using Wavelets”. *3RD International Workshop on Digital Mammography Chicago, Illinois, U.S.A. June 9-12*, p. 311-16, 1996.
- [Meersman – 1996] Meersman, D.; Scheunders, P.; Van, Dyck. D “Detection of Microcalcifications using Neural Networks”. *3RD International Workshop*

on Digital Mammography Chicago, Illinois, U.S.A. June 9-12, p. 287-90, 1996.

[Meersman – 1998] Meersman, D.; Scheunders, P.; Van, Dyck. D “Classification of microcalcifications using texture-based features”. 4TH International Workshop on Digital Mammography, University of Nijmegen, The Netherlands, p. 233-36, June 7-10, 1998.

[Metz – 1978] Metz, C.E “Basic Principles of ROC Analysis”. Semin. Nucl. Med. vol. 8, p. 283-98, 1978.

[Metz – 1986] Metz, C.E “ROC Methodology In Radiologic Imaging”. Investigative Radiology, vol. 21, n. 9 , p. 720-33. Sep. 1986.

[Metz – 1989] Metz, C.E “Some Practical Of Experimental Design And Data Analysis in Radiological ROC Studies”. Investigative Radiology, vol. 24, n. 3, p. 234 – 245, 1989.

[Metz – 1996] Metz, C. E. “Evaluation of Digital Mammography by ROC analysis”. 3RD International Workshop on Digital Mammography Chicago, Illinois, U.S.A. June 9-12, p. 61-68, 1996.

[Metz – 1998] Metz, C.E.; Herman, B. A. And Shen, J.H “Maximum Likelihood Estimation of Receiver Operating Characteristic (ROC) Curves From Continuously-Distributed Data”. Statistics In Medicine, vol. 17, p. 1033-53, 1998.

[Nagel – 1995] Nagel, R.H.; Nishikawa, J.P.; Giger, M.L.; Doi, K “Comparison of rule-based and artificial neural network approaches for improving the automated detection of clustered microcalcifications in mammograms’. SPIE 2622, p.775-779, 1995.

- [Nagel – 1998] Nagel, RH; Nishikawa, RM; Papaioannou, J.; Doi, K “Analysis of methods for reducing false positives in the automated detection of clustered microcalcifications in mammograms”, *Medical Physics*, vol.25, n.8, p.1502-6, Aug. 1998.
- [Nesbitt –1995] Nesbitt, D.; Aghdasi, F.; Ward,R.; Morgan, P. J. “Detection of Microcalcifications in digitized mammogram film images using wavelet enhancement and local adaptative false positive suppression”. *IEEE – Pacific – RIM- - Conference on Communications, Computers and Signal Processing Proceedings*. 1995, IEEE, Piscataway, NJ, USA, p. 594-97.
- [Netsch – 1998] Netsch, T.; Lang, M.; Peitgen, HO. “Automatic detection of microcalcification clusters after lossy compression of digitized mammograms”. *4TH International Workshop on Digital Mammography*, University of Nijmegen, The Netherlands, p. 465-72, June 7-10, 1998.
- [Nishikawa – 1994] Nishikawa, Robert M. ; Giger, Maryellen L. ; Doi, Kunio; Vyborny, Carl J.; Schmidt, Robert. “Computer-Aided Detection and Diagnosis of Masses and Clustered Microcalcifications from Digital mammograms”; “State of the Art in Digital Mammographic Image Analysis” K. Bowyer and S. Astley, Editors. *World Scientific Publishing Co.*, p. 82-102, 1994.
- [Nishikawa – 1994] Nishikawa, Robert M.; Jiang, Yulei; Giger, Maryellen L.; Schmidt, Robert A.; Vyborny, Carl J.; Zhang, W.; Papaioannou, J. ; Bick, U.; Nagel, R.; and Doi K “Performance of Automated CAD Schemes for the Detection and Classification of Clustered Microcalcifications”. *Digital Mammography, Proc. Of 2 nd International Workshop on Digital Mammography*, York, UK, Elsevier Science, p. 13-20, 1994.
- [Nishikawa – 1994] Nishikawa, Robert M.; Giger, Maryellen L ; Doi, Kunio; Metz, Charles E., Yin, Fang- Fang; Vyborny, Carl J. and Schmidt Robert A

“Effect of Case Selection on the Performance of Computer - Aided Detection Schemes”, *Medical Physics*. Vol.21, p. 265-269, February 1994.

[Nishikawa – 1996] Nishikawa, Robert M.; Papaioannou, J.; Collins, Stephen A., “Reproducibility of an automated scheme for the detection of clustered microcalcifications on digital mammograms”. *SPIE*, vol. 2710, p.24-29, 1996.

[Nishikawa, - 1994] Nishikawa, Robert M.; Vyborny, Carl J.; Giger, Maryellen L.; and Doi, Kunio “Analysis of False-Positive Microcalcification Clusters Identified by a Mammographic Computer-Aided Detection Scheme”. *Image Processing SPIE*, vol. 2167, p. 773-77, Febr. 1994.

[Nishikawa, - 1995] Nishikawa, Robert M.; Giger, Maryellen L.; Doi, K.; Vyborny, Carl J. and. Schmidt, Robert <sup>a</sup> “Computer-Aided Detection of Clustered Microcalcifications on Digital Mammograms”. *Medical Biol. Eng. Comput.*; vol. 33, n. 2, p.174-8, Mar. 1995.

[Nishikawa, - 1997] Nishikawa, Robert M., Wolverton, Dulcy E., Schmid, Robert A., Papaioannou, John “Radiologists' ability to discriminate computer-detected true and false positives from an automated scheme for the detection of clustered microcalcifications on digital mammograms”, *SPIE* 97, vol.3036, p.198-204.

[Oestmann - 1988] Oestmann, JW.; Kopans, D.; Hall, DA.; McCarthy, KA.; Rubens, JR.; Greene, R. “A comparison of digitized storage phosphors and conventional mammography in the detection of malignant microcalcifications”. *Investigative Radiology*, vol. 23, n . 10, p. 725-8, Oct. 1988.

[Panizza – 1996] Panizza, P.; Rodighiero, MG.; De Gaspari, A.; Tacchini, S.; Camalori, M.; Del Maschio “The diagnostic accuracy of a digital

mammography system with photostimulable storage phosphors used with automatic reading”. *Radiol. Med. (Torino)*, vol. 91, n. (1-2), p. 46-51, Jan-Feb. 1996.

[Patrick – 1991] Patrick, Edward A.; Moskowitz, M.; Mansukhani, Vineet T. and Gruenstein, Eric I “Expert Learning System Network for Diagnosis of Breast Calcifications’, *Investigative Radiology*, Vol. 26, p. 534-39, June 1991.

[Ping - 1988] Ping, Heang; Doi, Kunio; Vyborny, Carl J.; Lam, Kwok-Leung and Schmidt, Robert A “Computer-Aided Detection of Microcalcifications in Mammograms Methodology and Preliminary Clinical Study”, *Investigative Radiology*, Vol. 23, n. 9, p.664 – 671, March 1988

[Qian – 1994] Qian, W.; Clarke, LP.; Li, HD.; Clark, R.; Silbiger, ML. “Digital mammography: M-channel quadrature mirror filters (QMFs) for microcalcification extraction”. *Comput Med Imaging Graph*, vol. 18, n. 5 p. 301-14, Sep-Oct. 1994.

[Qian – 1995] Qian, W.; Kallergi, M.; Clarke, Laurence P.; Huai-Dong; Venugopal, P.; Song, D. and Clark, Robert <sup>a</sup> “Tree Structured Wavelet Transform Segmentation of Microcalcifications in Digital Mammography”. *Medical Physics*, vol. 22, n. 8, August 1995.

[Qian – 1998] Qian, W.; Clarke, LP.; Song, D.; Clark, RA. “Digital mammography: hybrid four-channel wavelet transform for microcalcification segmentation”. *Acad. Radiol.*, vol. 5, n. 5, p. 354-64, May 1998.

[Rosen – 1996] Rosen, D.; Martin, B.; Monheit, Mark.; Wolff, G.; and Stanton, M. “A Bayesian Neural Network to Detect Microcalcifications in Digitized Mammograms”. *3RD International Workshop on Digital Mammography Chicago, Illinois, U.S.A. June 9-12, p. 139-144, 1996.*

- [Sadjadi – 1995] Sadjadi, MRA.; Suarez, O. Y. ;Newman, F.; Zeligman, B.  
 “Detection of Clusters of Microcalcifications Using Neural Network Based Schemes”. Annual International Conference of the IEEE Engineering in Medicine and Biology Proceedings, vol. 17, n. 1, p. 533-34, 1995.
- [Sajda - 1996] Sajda, P.; Clay D. Spence, C.D.; Pearson, J.C.; Nishikawa, R..  
 “Exploiting context in mammograms: a hierarchical neural network for detecting microcalcifications”. SPIE 2710, p.733-742, 1996.
- [Sajda -1996] Sajda, P.; Spence, Clay D.; Pearson, John C. and Nishikawa, Robert M..  
 “Integrating multi-resolution and contextual information for improved microcalcification detection in CAD”. 3RD International Workshop on Digital Mammography Chicago, Illinois, U.S.A. June 9-12, p. 291-96, 1996.
- [Sehad – 1997] Sehad, Samira; Desarnaud, Serge and Strauss, Anne “Artificial Neural Classification of Clustered Microcalcifications on Digitized Mammograms”. Proceedings of the IEEE international Conference on Systems mMan and Cybernetics, vol. 5, p. 4217-22, 1997.
- [Shaw – 1998] Shaw, CC.; Wang, T.; King, JL.; Breitenstein, DS.; Chang, TS.; Harris, KM.; Baratz, AB.; Ganott, MA.; Reginella, R.; Sumkin, JH.; Gur, D. “Computed radiography versus screen-film mammography in detection of simulated microcalcifications: a receiver operating characteristic study based on phantom images”. Academic Radiology, vol. 5, n. 3, p. 173-80, Mar. 1998.
- [Shih-Chung –1996] Shih-Chung B. Lo.; Freedman, Matthew T.; Lin, Jyh-Shan J.; Li, H.; Hasegawa, A. and Mun, Seong K.. “Preliminary Study of Computer-Aided Search for Clustered Microcalcifications on Mammograms”. 3RD International Workshop on Digital Mammography Chicago, Illinois, U.S.A.

June 9-12, 1996.

- [Shih-Chung -1996] Shih-Chung B. Lo, Huai Li, Jyh-Shyan Lin, Akira Hasegawa, Osamu Tsujii, Matthew T. Freedman, and Seong K. Mun.. “Detection of Clustered Microcalcifications Using Fuzzy Modeling and Convolution Neural Network”. Proceedings SPIE 1996, vol.2710, p. 8-15.
- [Shih-Chung -1998] Shih-Chung B. Lo, Jyh-Shyan J. Lin, Matthew T. Freedman and Seong K. Mun;. “Application of Artificial Neural Networks to Medical Image Pattern Recognition: Detection of Clustered Microcalcifications on Mammograms and Lung Cancer on Chest Radiographs”. Journal of VLSI Signal Processing for Signal Image and Video Technology, vol. 18, n. 3, p. 263-274, Apr. 1998.
- [Sorantin – 1998] Sorantin, E.; Schmidt, F.; Mayer, H.; Winkler, H.; Szepesvari, C.; Graif, E.; Schuetz, E. “Automated detection and classification of microcalcifications in mammograms using artificial neural nets”. 4TH International Workshop on Digital Mammography, University of Nijmegen, The Netherlands, p. 225-32, June 7-10, 1998.
- [Spiegel – 1985] Spiegel, M. R “Estatística”, tradução, revisão, e adapatação de Carlos Augusto Crusius – 2. Ed. – São Paulo: McGraw-Hill do Brasil, 1985.
- [Strickland – 1996] Strickland, N. R.; Hlm, H. H “Wavelet Transforms for Detecting Microcalcifications in Mammograms”. IEEE Transactions on Medical Imaging, vol. 15, n.2 , p. 218 – 229, April 1996.
- [Veldkamp – 1996] Veldkamp, W.; Karssemeijer, N “Automatic Classification of Clustered Microcalcifications in Digital Mammography”. 3RD International Workshop on Digital Mammography Chicago, Illinois, U.S.A., June 9-12, p. 231-34, 1996.



- [Veldkamp – 1998] Veldkamp, W.; Karssemeijer, N. “Improved correction for signal dependent noise applied to automatic detection of microcalcifications”, 4TH International Workshop on Digital Mammography, University of Nijmegen, The Netherlands, p. 169-76, June 7-10, 1998.
- [Wallet – 1997] Wallet, BC.; Solka, JL. “A method for detecting microcalcifications in digital mammograms”. J. Digit. Imaging, vol. 10, n. 3 Suppl. 1 , p. 136-9, Aug. 1997.
- [Winkler – 1998] Winkler, P.; Sorantin, E.; Tanacs, A.; Schmidt, F.; Mayer, H.; Szepesvari, C. “Performance evaluation for automated detection of microcalcifications in mammo using three different film-digitizers”. 4TH International Workshop on Digital Mammography, University of Nijmegen, The Netherlands, p. 485-86, June 7-10, 1998
- [Woods – 1993] Woods, Kevin S.; Solka, Jeffrey L.; Priebe, Carey E.; Kegelmeyer, W. Philip Jr.; Doss, Christopher C.; Bowyer, Kevin W. “Comparative Evaluation of Pattern Recognition Techniques for Detection of Microcalcifications in Mammography”. Int. J. Pattern Recognition Artificial Intell., vol. 7, p. 1417-36, Dec. 1993.
- [Wu - 1992] Wu, Y.; Doi, K.; Giger, ML.; Nishikawa, RM. “Computerized detection of clustered microcalcifications in digital mammograms: applications of artificial neural networks”, Medical Physics, vol. 19, n. 3, p. 555-60, May-Jun. 1992.
- [Yoshida –1996] Yoshida, H.; Anastasio, Mark A.; Nishikawa, Robert M.; Giger, Maryellen L.; Doi, K “Optimally Weighted Wavelet Packet Transform for Detection of Clustered Microcalcifications in Digital Mammograms”, 3RD International Workshop on Digital Mammography Chicago, Illinois, June 9-12, p. 317-22, 1996.

- [Yoshida –1996] Yoshida, H.; Doi, K.; Nishikawa, RM.; Giger, ML.; Schmidt, RA.  
“An improved computer-assisted diagnostic scheme using wavelet transform for detecting clustered microcalcifications in digital mammograms”, Acad Radiol, vol. 3, n. 8, p.621-7, Aug. 1996.
- [Yoshida –1996] Yoshida, H.; Nishikawa, R.; Giger, M.L.; Doi, K.  
“Signal/background separation by wavelet packets for detection of microcalcifications in mammograms”. SPIE 2825, p.805-811, 1996.
- [Yu – 1996] Yu, S.; Brown, S.; Xue, Yu; Guan, L “Enhancement and Identification of Microcalcifications in Mammograms Image Using Wavelets”. Proceedings of the IEEE international Conference on Systems Man and Cybernetics, vol. 2, p. 1166-71, 1996
- [Zhang – 1994] Zhang, W.; Doi, K.; Giger, Maryellen L.; Wu, Y.; Nishikawa, Robert M and Schmidt, Robert A “Computerized Detection of Clustered Microcalcifications in Digital Mammograms Using a Shift-Invariant Artificial Neural Network”, Medical Physics, vol. 21, n.4, p. 517-24, Apr.,1994.
- [Zhang – 1996] Zhang, W.; Doi, K.; Giger, ML.; Nishikawa, RM.; Schmidt, RA.  
“An improved shift-invariant artificial neural network for computerized detection of clustered microcalcifications in digital mammograms”. Medical Physics, vol. 23, n .4, p. 595-601, Apr. 1996.
- [Zheng – 1995] Zheng, B.; Chang, YH.; Staiger, M.; Good, W.; Gur, D. “Computer-Aided Detection of Clusters Microcalcifications in Digital Mammograms”. Academic Radiology, vol. 2, n. 8, p.655-62, August 1995.
- [Zheng – 1996] Zheng, B.; Qian, Ww.; Clarke, LP “Digital Mammography: Mixed Feature Neural Network With Spectral Entropy Decision for Detection of

Microcalcifications". IEEE Transactions on Medical Imaging. , vol. 15, n.  
5, p. 589-97, Oct. 1996.