



UNIVERSIDADE DE SÃO PAULO

Escola de Artes, Ciências e Humanidades

Relatório Técnico PPgSI-004/2013
*Aplicação e Avaliação de Sistemas de Auxílio ao
Diagnóstico: uma Revisão Sistemática*

Vagner M. Gonçalves
Márcio E. Delamaro
Fátima L. S. Nunes

Outubro - 2013

O conteúdo do presente relatório é de única responsabilidade dos autores.

Série de Relatórios Técnicos

PPgSI-EACH-USP. Rua Arlindo Béttio, 1000 - Ermelino Matarazzo -
03828-000

São Paulo, SP.

TEL: (11) 3091-8197

<http://www.each.usp.br/ppgsi>

Aplicação e Avaliação de Sistemas de Auxílio ao Diagnóstico: uma Revisão Sistemática

Vagner M. Gonçalves^{1,2}, Márcio E. Delamaro^{1,3}, Fátima L. S. Nunes^{1,2}

¹Laboratório de Aplicações de Informática em Saúde

²Escola de Artes, Ciências e Humanidades – Universidade de São Paulo
São Paulo – SP, Brasil

{vagner.goncalves, fatima.nunes}@usp.br

³Instituto de Ciências Matemáticas e de Computação
Universidade de São Paulo – Campus de São Carlos
Caixa Postal 668, 13560-970, São Carlos, SP, Brasil

delamaro@icmc.usp.br

Resumo. *Sistemas de Diagnóstico Auxiliado por Computador (CAD – Computer-Aided Diagnosis) têm sido desenvolvidos por diversos grupos de pesquisa a fim de auxiliar na detecção e classificação de estruturas de interesse em imagens médicas que podem indicar a presença de doenças. Atualmente, os objetos de pesquisa nesse contexto compreendem muitas aplicações e consideram diferentes modalidades de imagens médicas para o diagnóstico de diversas doenças. Um dos desafios no desenvolvimento de tais sistemas é a sua avaliação. Este Relatório Técnico apresenta uma Revisão Sistemática sobre sistemas de CAD desenvolvidos recentemente, com foco principal em métodos de avaliação empregados. O objetivo é proporcionar uma visão geral dos trabalhos da área considerando-se aspectos tais como técnicas computacionais e bases de imagens utilizadas, destacando-se uma compilação de técnicas de avaliação empregadas nas pesquisas. Os resultados da revisão indicam que os cânceres de mama e de pulmão são, ainda, as anomalias mais exploradas no desenvolvimento de sistemas de CAD. Sensibilidade, Especificidade e curvas ROC são os principais métodos de avaliação aplicados a CAD, segundo os trabalhos incluídos.*

1. Considerações Iniciais

1.1. Diagnóstico Auxiliado por Computador

Sistemas de Diagnóstico Auxiliado por Computador (CAD – *Computer-Aided Diagnosis*) ou, simplesmente, Sistemas de Auxílio ao Diagnóstico consistem em sistemas computacionais com a finalidade de fornecer uma segunda opinião aos especialistas, contribuindo na tomada de decisão a respeito de um diagnóstico [Doi 2007]. Esses sistemas computam a saída a partir de informações provenientes de diversas fontes, sendo que as imagens médicas de diversas modalidades são as principais delas. Segundo van Ginneken et al. [2010], CAD se tornou a área de pesquisa mais ativa dentro do contexto de análise de imagens médicas. Doi [2006] enfatiza que os sistemas CADs podem melhorar a precisão do diagnóstico e a consistência da interpretação de tais imagens.

O início da pesquisa desta categoria de sistemas ocorreu nas décadas de 1960 e 1970, quando ainda imaginava-se construir um diagnóstico sem a presença de radiologistas. Na década de 1980, surgiram vários grupos de pesquisa no mundo com o objetivo de

desenvolver sistemas de CAD para finalidades diversas, tendo principalmente as imagens provenientes de Raios-X como entrada de dados. No final da década de 90, alguns grupos avançaram no desenvolvimento, incluindo imagens de outras modalidades, como Ressonância Magnética Nuclear e Ultrassom. Nessa década, os primeiros sistemas de CAD comerciais começaram a ser aprovados nos Estados Unidos para emprego na rotina médica, como o *ImageChecker*, comercializado pela empresa *R2 Technology*, de Los Altos, Califórnia [Brown 2001, Giger 1999].

Segundo Doi [2007], no desenvolvimento de esquemas de CAD, é necessário empregar o processamento de imagens e técnicas para o processamento de informação para uma análise computacional quantitativa e, além disso, compreender a relevância médica das imagens com base nas suas características. No entanto, para se obter sistemas de CAD com métodos eficientes, no seu desenvolvimento é necessário empregar técnicas mais recentes, como redes neurais artificiais, transformadas *Wavelet* e análise fractal, por exemplo.

De forma geral, um sistema de CAD tem uma imagem médica digital como fonte de dados e possui algumas fases bem definidas até chegar a uma saída final, conforme ilustrado na Figura 1.



Figura 1. Fases de processamento em um sistema de CAD.

De acordo com Brown [2001], os sistemas de CAD deixaram de ser somente objetos de pesquisa acadêmica, passando por uma transição de sucesso para implementações comerciais. Ainda segundo o autor, antes de chegar à comunidade médica, tais sistemas devem oferecer provas de segurança e eficácia, determinadas, por exemplo, pelo nível intrínseco de risco associado ao sistema e as condições médicas em sua aplicação, pela precisão das alegações do sistema e pelo nível de supervisão exercido durante a sua utilização.

Um das maiores dificuldades encontradas durante o desenvolvimento de sistemas de CAD diz respeito ao processo de avaliação. Não é fácil saber se uma determinada técnica é eficaz ou não, pois os resultados podem variar de acordo com o conjunto de imagens utilizado nos testes. Para se atestar a viabilidade do uso de uma técnica, são necessários testes com um conjunto de imagens que tenham, preferencialmente, características de aquisição variadas e que atendam aos requisitos da finalidade da técnica, isto é, conter as estruturas procuradas pelo sistema computacional. Isso envolve uma pesquisa intensa junto a hospitais e clínicas para obtenção das imagens médicas e dos respectivos laudos médicos e uma exaustiva tarefa de catalogar essas imagens a fim de permitir uma recuperação rápida e precisa de acordo com as suas características.

1.2. Revisão Sistemática

Este Relatório Técnico apresenta e discute os resultados de uma Revisão Sistemática (RS) que teve como objetivo levantar o estado da arte com relação à avaliação de sistemas de CAD. Simultaneamente, foram analisados dados adicionais dos trabalhos incluídos, tais como técnicas computacionais empregadas no desenvolvimento de esquemas

de CAD, anomalias investigadas e modalidades de imagens médicas mais exploradas. Tal levantamento possibilitou uma análise geral do panorama atual no que diz respeito a projeto, desenvolvimento e avaliação de tais sistemas.

A RS é uma metodologia rigorosa de pesquisa bibliográfica que visa a identificar estudos primários e secundários relacionados a um determinado tema de pesquisa. Ela permite avaliar e interpretar toda a pesquisa relevante desenvolvida sobre uma questão particular ou sobre um tópico de interesse [Kitchenham 2004].

Segundo Kitchenham [2004] e Biolchini et al. [2007], uma RS é conduzida em três fases bem definidas: Planejamento de Revisão, Execução de Revisão e Análise de Resultados. Na fase de *Planejamento* se define um protocolo no qual são especificadas a questão de pesquisa e a metodologia que será empregada na execução da revisão. Integrarão esse protocolo os objetivos da RS, as fontes de consulta, os critérios de inclusão e de exclusão de estudos primários, as palavras-chave e outros tópicos de interesse para a pesquisa bibliográfica.

Na fase de *Execução*, a pesquisa bibliográfica propriamente dita é realizada. É nessa etapa que ocorrem a seleção dos estudos primários, de acordo com os critérios de inclusão e de exclusão definidos, e a extração e síntese de dados dos trabalhos incluídos. Por fim, na fase de *Análise de Resultados*, são realizadas a análise e a documentação dos resultados e das conclusões obtidas por meio deles.

Uma importante diferença entre a RS e a revisão de literatura comum é o fato de que o estabelecimento de um protocolo permite que a revisão seja passível de reprodução. Isso significa que outros pesquisadores podem reproduzir o mesmo protocolo, sendo também capazes de avaliar os métodos empregados para o caso em questão [Biolchini et al. 2007].

1.3. Revisões de Literatura Correlatas

Visões gerais sobre o desenvolvimento e tendências de CAD foram apresentadas por Doi [2007] e por Shiraishi et al. [2009]. Doi [2007] realizou uma revisão histórica com relação ao desenvolvimento de esquemas de CAD. Foram apresentados, também, exemplos de trabalhos que exploram diferentes modalidades de imagens médicas, buscando o auxílio ao diagnóstico para diversas anomalias, tais como nódulos de pulmão, fraturas vertebrais e aneurismas intracraniais. O pesquisador discorreu, ainda, sobre o potencial de tais esquemas para aplicação na rotina clínica. Shiraishi et al. [2009] apresentaram uma revisão a respeito da aplicação de análise utilizando a curva Característica de Operação do Receptor (ROC – *Receiver Operating Characteristic*) para a avaliação de sistemas de CAD. A revisão incluiu trabalhos publicados no periódico *Radiology*¹ entre os anos de 1997 e 2006. Foram analisadas, ainda, a participação de observadores humanos nos processos de avaliação reportados em tais trabalhos e as modalidades de imagens médicas mais exploradas.

A revisão aqui apresentada, diferentemente das anteriormente mencionadas, foi conduzida de forma a atender determinados requisitos. Tais requisitos, que caracterizam o seu diferencial e a relevância dos resultados apresentados, são listados a seguir:

- seleção mais abrangente de bases científicas como fontes de consulta, permitindo o acesso a publicações diversificadas da área; tais bases foram selecionadas após a

¹<http://radiology.rsna.org/>

condução de um estudo exploratório que norteou a definição das fontes relevantes para a pesquisa;

- definição e divulgação de um protocolo de RS que foi seguido rigorosamente durante a condução da revisão; tal protocolo permite a reprodução da mesma e a auditoria dos critérios utilizados;
- análise de resultados levando-se em consideração cada tarefa automatizada para o auxílio ao diagnóstico individualmente (i.e., segmentação, detecção e classificação de anomalias e regiões de interesse).

1.4. Organização do Relatório

Além da presente seção introdutória, na qual foram apresentados os principais conceitos relacionados ao Diagnóstico Auxiliado por Computador e à Revisão Sistemática, bem como os objetivos e diferenciais do trabalho conduzido e revisões de literatura correlatas, o relatório está organizado da seguinte forma. Na Seção 2 é apresentado o protocolo de Revisão Sistemática de forma descritiva. Na Seção 3 são descritos os procedimentos e materiais de condução de revisão empregados. Na Seção 4 são apresentados e discutidos os resultados obtidos. Por fim, na Seção 5 são apresentadas as considerações finais obtidas por meio dos resultados da revisão apresentada.

2. Planejamento: Protocolo da Revisão Sistemática

O ponto de partida no planejamento de uma RS é a definição das questões de pesquisa. As respostas para tais questões devem possibilitar o alcance dos objetivos propostos. A revisão apresentada neste relatório teve como objetivo inicial levantar o estado da arte com relação à avaliação de sistemas de CAD. Nesse contexto, foram definidas as seguintes questões de pesquisa:

- Quais são as metodologias aplicadas atualmente para a avaliação de sistemas de CAD baseados em imagens médicas?
- Quais são as modalidades de imagens médicas, anomalias e técnicas computacionais envolvidas no desenvolvimento de sistemas CAD?
- Quais são as vantagens e limitações apresentadas pelas metodologias de avaliação empregadas em sistemas CAD?

Uma análise exploratória sobre CAD foi previamente conduzida em diversas bases científicas. Tal pesquisa norteou a seleção das fontes de consulta e a definição das palavras-chave utilizadas na RS. Foram consultadas as bases que tradicionalmente publicam artigos sobre o tema, de acordo com a experiência dos autores e os dados colhidos na análise exploratória preliminar. Diante desse cenário, as seguintes bases foram selecionadas:

- *ACM Digital Library (ACM)*²
- *IEEE Xplore Digital Library (IEEE)*³
- *ISI Web of Knowledge (ISIWK)*⁴
- *Scientific Electronic Library Online (SciELO)*⁵

²<http://dl.acm.org/>

³<http://ieeexplore.ieee.org/Xplore/dynhome.jsp>

⁴<http://apps.webofknowledge.com/>

⁵<http://www.scielo.org/php/index.php>

- *ScienceDirect* (SD)⁶

A partir das palavras-chave definidas, buscas por artigos de periódicos ou de anais de eventos científicos foram realizadas nas bases selecionadas. Para o levantamento do estado da arte com relação ao tema abordado foram considerados trabalhos recentes (publicados a partir de 2006).

A seleção de fontes de consulta reconhecidas na área tornou possível a recuperação de uma quantidade significativa de trabalhos que abordam CAD, a qual será detalhada adiante. Além disso, contribuíram para essa tarefa a definição de critérios adequados de inclusão e de exclusão de trabalhos.

Foi empregada uma composição de termos aplicável em todas as bases selecionadas. Os termos deveriam estar presentes no título, no resumo ou nas palavras-chave de um trabalho para que o mesmo fosse recuperado. Essas opções foram definidas por meio de ferramentas de busca avançada disponíveis nas bases. A única base selecionada que não possui uma ferramenta de busca avançada é a SciELO. Em tal base foi realizada, então, uma busca simples, utilizando-se os termos *computer-aided diagnosis* (com opções de pesquisa integrada e regional selecionadas). Para as demais bases consultadas a composição de termos definida foi a seguinte:

```
(evaluation OR testing OR assessment)
AND
("computer-aided diagnosis system" OR "computer-aided diagnosis scheme")
```

Na Tabela 1 são apresentadas as composições de termos utilizadas para cada uma das bases consultadas.

Tabela 1. Composições de termos utilizadas nas buscas.

BASE	FERRAMENTA DE BUSCA	REFINAMENTO POR ANO JÁ INCLuíDO?	STRING DE BUSCA
ACM	Advanced Search	Sim	(evaluation or assessment or testing or Keywords:evaluation or Keywords:assessment or Keywords:testing) and ("computer-aided diagnosis system" or "computer-aided diagnosis scheme" or Keywords:"computer-aided diagnosis system" or Keywords:"computer-aided diagnosis scheme")
IEEE	Command Search	Sim	("Document Title":evaluation OR "Abstract":evaluation OR "Author Keywords":evaluation OR "Document Title":assessment OR "Abstract":assessment OR "Author Keywords":assessment OR "Document Title":testing OR "Abstract":testing OR "Author Keywords":testing) AND ("Document Title":"computer-aided diagnosis system" OR "Abstract":"computer-aided diagnosis system" OR "Author Keywords":"computer-aided diagnosis system" OR "Document Title":"computer-aided diagnosis scheme" OR "Abstract":"computer-aided diagnosis scheme" OR "Author Keywords":"computer-aided diagnosis scheme")
ISIWK	Search / Lemmatization=Off	Sim	Topic=(evaluation or testing or assessment) AND Topic=("computer-aided diagnosis system" or "computer-aided diagnosis scheme")
SciELO	Busca Simples	Não	computer-aided diagnosis
SD	Advanced Search	Sim	pub-date > 2005 and TITLE-ABSTR-KEY(evaluation OR testing OR assessment) and TITLE-ABSTR-KEY("computer-aided diagnosis system" OR "computer-aided diagnosis scheme")

A fim de selecionar somente trabalhos relevantes para o tema de estudo foram definidos critérios de inclusão e exclusão. Foram incluídos somente trabalhos que atendessem a, pelo menos, um dos critérios de inclusão e *nenhum* dos critérios de exclusão. Os *critérios de inclusão* definidos foram:

⁶<http://www.sciencedirect.com/>

- (a) apresentar ou abordar conceitos, critérios e metodologias para a avaliação de sistemas de CAD cujas saídas se apresentam em forma de imagens ou utilizam imagens processadas de alguma maneira para realizar diagnósticos;
- (b) aplicar alguma metodologia específica na avaliação de algum sistema de CAD com as características mencionadas no critério de inclusão (a).
- (c) apresentar conceitos ou revisões históricas sobre CAD.

Já os *critérios de exclusão* definidos foram os seguintes:

- (d) ser semelhante, em conteúdo e resultados, a outro(s) trabalho(s) dos mesmos autores que, porventura, venha(m) também a ser recuperado(s) em alguma das bases consultadas;
- (e) o ano de publicação está fora do limite estabelecido (2006 até a data de término da condução da RS);
- (f) não estar disponível integralmente nas bases consultadas ou em alguma outra base acessível por meio, por exemplo, da ferramenta Google Acadêmico.

3. Condução da Revisão

As buscas foram realizadas durante o segundo semestre de 2011. Foram recuperados, no total, 57 trabalhos distintos. Ao todo, foram incluídos 36 trabalhos (63,16%) dos quais 34 são trabalhos que apresentam esquemas de CAD, bem como as metodologias de avaliação empregadas, e dois são artigos de conteúdo mais conceitual. O primeiro discute a participação de profissionais de saúde e técnicos no ciclo de vida (projeto, avaliação e implantação) de sistemas de CAD [Pietka et al. 2011]. O segundo, por sua vez, retrata um estudo a respeito do uso de análise ROC para a avaliação de sistemas de CAD em trabalhos publicados no periódico *Radiology* [Shiraishi et al. 2009].

Toda a etapa de condução da RS foi devidamente documentada tomando-se como base os modelos propostos em Biolchini et al. [2007] e Kitchenham [2004]. Os documentos produzidos, bem como ferramentas utilizadas são descritas a seguir.

Formulário de condução Um formulário de condução foi produzido para cada base científica consultada. Nele são documentadas todas as informações de interesse da busca: datas de acesso, composição de termos utilizada, lista de trabalhos recuperados, critérios de inclusão e de exclusão atendidos por cada um e observações que se fizeram necessárias.

Formulário de extração de dados Para cada trabalho incluído foi produzido um formulário de extração de dados. Em tal formulário, além das informações de publicação e referência, um resumo do trabalho foi incluído e os tópicos de interesse documentados. Os principais tópicos de interesse extraídos foram: finalidade do sistema utilizado/proposto, método(s) de avaliação de CAD empregado(s), modalidade(s) e quantidade(s) de imagens/casos utilizadas(os) nos testes, principais resultados e testes e outras informações que se mostraram relevantes para a pesquisa.

Na Figura 2 é apresentado um diagrama de fluxo, baseado em Liberati et al. [2009], que sumariza a seleção de estudos realizada. É possível observar por meio desse diagrama que houve exclusão de trabalhos somente na fase de avaliação por meio do título e do

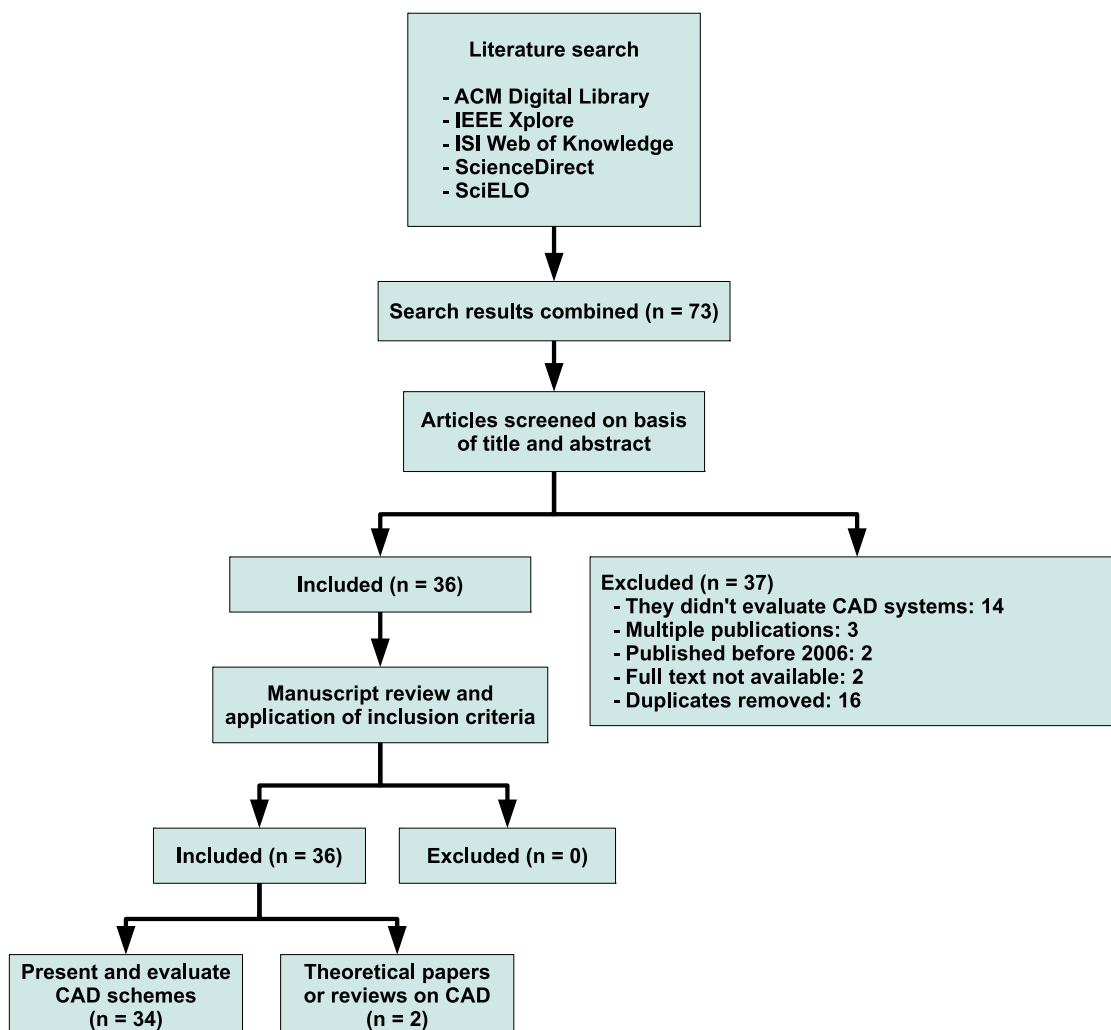


Figura 2. Diagrama de fluxo resumindo a etapa de seleção de estudos.

resumo. Vale ressaltar que, nessa fase, houve a necessidade de uma leitura superficial de cada artigo avaliado, a fim de se ratificar se o mesmo poderia ou não ser excluído.

Na Tabela 2 são apresentados os trabalhos incluídos, bem como os critérios de inclusão atendidos por cada um deles.

Foram recuperados trabalhos publicados em periódicos científicos, anais de eventos e coletâneas de artigos. Em especial, uma variedade significativa de publicações da área médica, de inteligência computacional, de processamento e análise de imagens e de reconhecimento de padrões foram observadas. Na Tabela 3 são apresentados os veículos dos quais foram incluídos trabalhos na presente RS.

Na próxima seção são apresentados e discutidos os resultados obtidos por meio dessa RS.

Tabela 2. Critérios atendidos por cada um dos trabalhos incluídos.

ID	REFERÊNCIA	CRIT.	ID	REFERÊNCIA	CRIT.
1	Giannakopoulou et al. [2010]	(b)	19	Gomathi e Thangaraj [2010]	(b)
2	Jasmine et al. [2009]	(b)	20	Itai et al. [2009]	(b)
3	Osman et al. [2009]	(a) (b)	21	Pietka et al. [2010]	(a) (b)
4	Verma [2009]	(b)	22	Wada et al. [2006]	(b)
5	He et al. [2011]	(b)	23	Korfiatis et al. [2007]	(b)
6	Song et al. [2010]	(a) (b)	24	Kuang e Ye [2008]	(b)
7	Haindl et al. [2007]	(b)	25	Huang et al. [2009b]	(a) (b)
8	Zheng et al. [2008]	(a) (b)	26	Mumcuoglu et al. [2011]	(a) (b)
9	Grusauskas et al. [2008]	(a) (b)	27	Raja et al. [2010]	(b)
10	Grusauskas et al. [2009]	(b)	28	Tan et al. [2010]	(a) (b)
11	Lee et al. [2009]	(a) (b)	29	Chan [2010]	(b)
12	Shen et al. [2007]	(a) (b)	30	Huang et al. [2007]	(b)
13	Barhoumi et al. [2007]	(b)	31	Huang et al. [2009a]	(b)
14	Garnavi et al. [2011]	(a) (b)	32	López et al. [2011]	(a) (b)
15	Odeh et al. [2006]	(a) (b)	33	Sanchez et al. [2011]	(a) (b)
16	Ampeliotis et al. [2007]	(b)	34	Jiménez et al. [2010]	(a) (b)
17	Schilham et al. [2006]	(a) (b)	35	Shiraishi et al. [2009]	(a) (c)
18	Capobianco et al. [2008]	(b)	36	Pietka et al. [2011]	(a) (c)

4. Resultados e Discussões

4.1. Trabalhos Incluídos

Na Tabela 4 são apresentados os 36 sistemas de CAD relatados em 34 dos trabalhos incluídos. Essa quantidade de sistemas justifica-se, pois Pietka et al. [2010] relataram em seu artigo o desenvolvimento de três sistemas distintos. Nas subseções seguintes, uma análise geral do estado da arte com relação ao desenvolvimento e avaliação de sistemas de CAD é apresentada com base nos trabalhos incluídos por meio da RS.

4.2. Anomalias Estudadas

Como se pode observar na Tabela 4, existe uma variedade significativa de anomalias que são atualmente objetos de estudo para o desenvolvimento de sistemas de CAD. Como apresentado no gráfico da Figura 3, câncer de mama e câncer de pulmão são as doenças mais comumente estudadas de acordo com os trabalhos incluídos. Tal evidência é significativa, considerando-se que já é bem conhecida a importância do diagnóstico precoce dos mais diferentes tipos de neoplasias.

Além das doenças mencionadas, outras anomalias foram também relatadas nos trabalhos incluídos, envolvendo o processamento de imagens médicas de diferentes estruturas e órgãos do corpo humano. São eles: pele, retina, ossos, próstata, dentes, rins e cérebro (ver Tabela 4). Sistemas de CAD são tema de estudo e pesquisa com aplicabilidade nas mais distintas áreas médicas. Apesar de muitas abordagens ainda estarem distantes da aplicação clínica, a variedade de idéias, técnicas e áreas de aplicação evidenciam que se pode esperar bastante das aplicações computacionais de auxílio ao diagnóstico para muitas anomalias já bem conhecidas.

4.3. Modalidades de Imagens Médicas Exploradas

Diferentes tipos de imagens médicas têm sido objetos de estudo para o desenvolvimento de sistemas de CAD. Tal afirmação é ratificada por meio da observação dos resul-

Tabela 3. Veículos científicos dos quais foram incluídos trabalhos recuperados.

PERIÓDICOS	<p>Academic Radiology (ISSN: 1076-6332) Archivos de la Sociedad Española de Oftalmología (ISSN: 0365-6691) Biomedical Signal Processing and Control (ISSN: 1746-8094) Computers in Biology and Medicine (ISSN: 0010-4825) European Journal of Radiology (ISSN: 0720-048X) Expert Systems (ISSN: 1468-0394) IEEE Journal of Selected Topics in Signal Processing (ISSN: 1932-4553) IEEE Transactions on Nuclear Science (ISSN: 0018-9499) Investigative Ophthalmology & Visual Science (ISSN: 0146-0404) Jornal Brasileiro de Pneumologia (ISSN: 1806-3713) Machine Vision and Applications (ISSN: 0932-8092) Medical Image Analysis (ISSN: 1361-8415) Neurocomputing (ISSN: 0925-2312) Nuclear Medicine Communications (ISSN: 0143-3636) Radiology (ISSN: 0033-8419) Skin Research and Technology (ISSN: 1600-0846) Ultrasound in Medicine & Biology (ISSN: 0301-5629)</p>
COLETÂNEAS	<p>Lecture Notes in Artificial Intelligence / Knowledge-based and Intelligent Information and Engineering Systems (ISBN: 978-3-540-74828-1) Lecture Notes in Computer Science / Advances in Neuro-Information Processing (ISBN: 978-3-642-02489-4) Lecture Notes in Computer Science / Computer Analysis of Images and Patterns (ISBN: 978-3-540-74271-5) Lecture Notes in Computer Science / Image Analysis and Recognition (ISBN: 978-3-540-44894-5) Lecture Notes in Computer Science / Medical Biometrics (ISBN: 978-3-642-13922-2) Lecture Notes in Computer Science / Medical Image Computing and Computer-Assisted Intervention (ISBN: 978-3-642-04270-6)</p>
ANAIS	<p>Annual ACM Bangalore Conference (ISBN: 978-1-4503-0001-8) Annual International Conference of the IEEE Engineering in Medicine and Biology Society (ISBN: 978-1-4244-4123-5) IEEE International Conference on Signal Processing and Communications (ISBN: 1-4244-1236-6) IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBN: 978-1-4244-3932-4) IEEE International Workshop on Imaging Systems and Techniques (ISBN: 1-4244-0965-9) International Conference on Control, Automation, Communication and Energy Conservation (ISBN: 978-1-4244-4789-3) International Joint Conference on Neural Networks (ISBN: 978-1-4244-3553-1) International Symposium on Intelligent Information Technology Application (ISBN: 978-0-7695-3497-8) National Radio Science Conference (ISBN: 978-1-4244-4214-0) SPIE Medical Imaging (ISBNs: 978-0-8194-6425-5 e 978-0-8194-7099-7)</p>

tados da presente RS. Dos doze sistemas relatados voltados para o câncer de mama, oito (66,67%) utilizam mamogramas. A mamografia é a técnica mais efetiva no diagnóstico precoce do câncer de mama [Giger 1999]. Sendo assim, é fato que a exploração de imagens de tal domínio continua sendo de grande importância, visto que ainda existe a preocupante possibilidade de perda de sinais por parte do médico que analisa o exame [Giannakopoulou et al. 2010, Jasmine et al. 2009, Osman et al. 2009, Verma 2009]. Os quatro outros sistemas relatados (33,33%) trabalham com imagens de Ultrasonografia, técnica que tem sido muito importante para, em conjunto com a Mamografia, aumentar a sensibilidade no diagnóstico do câncer de mama [Giger 1999, Lee et al. 2009, Shen et al. 2007].

Dos sistemas de CAD relatados voltados para o câncer de pulmão, seis (85,71%) trabalham com imagens provenientes de exames de Tomografia Computadorizada. A única exceção é o sistema apresentado em Schilham et al. [2006] que utiliza imagens radiográficas do tórax.

Dois dos sistemas relatados voltados para o câncer de pele (IDs 13 e 14) utilizam imagens provenientes de exames de Dermatoscopia, enquanto o terceiro (ID 15) trabalha

Tabela 4. Sistemas de CAD apresentados nos trabalhos incluídos na RS.

ID	REFERÊNCIA	ANOMALIAS	MODALIDADE DE IMAGEM	INTERESSES	MÉTODOS E MÉTRICAS DE AVALIAÇÃO
1	Giannakopoulou et al. [2010]	Câncer de mama	Mamografia	Classificação de microcalcificações	Especificidade, ROC (e área sob a curva), Sensibilidade
2	Jasmine et al. [2009]	Câncer de mama	Mamografia	Classificação de microcalcificações	Sensibilidade
3	Osman et al. [2009]	Câncer de mama	Mamografia	Classificação de microcalcificações	Especificidade, Sensibilidade
4	Verma [2009]	Câncer de mama	Mamografia	Classificação de ROIs	Taxa de classificação correta
5	He et al. [2011]	Câncer de mama	Mamografia	Segmentação da estrutura do tecido e Classificação de risco	Segmentação: Comparação entre segmentação automatizada e segmentação manual; Classificação de risco: Grau de satisfação de profissionais da saúde, Sensibilidade
6	Song et al. [2010]	Câncer de mama	Mamografia	Segmentação de lesões	Diferença entre bordas baseada em distâncias, <i>Overlap</i> , <i>Over-segmentation</i> , <i>Under-segmentation</i>
7	Haindl et al. [2007]	Câncer de mama	Mamografia	Segmentação de ROIs	Índices de um <i>Benchmark</i> *
8	Zheng et al. [2008]	Câncer de mama	Mamografia	Segmentação de ROIs e Classificação de ROIs	Segmentação: Relative area difference; Classificação: ROC (e área sob a curva)
9	Grusauskas et al. [2008]	Câncer de mama	Ultrasonografia	Classificação de lesões	Segmentação: <i>Overlap</i> ; Classificação: ROC (e área sob a curva)
10	Grusauskas et al. [2009]	Câncer de mama	Ultrasonografia	Classificação de lesões	ROC (e área sob a curva)
11	Lee et al. [2009]	Câncer de mama	Ultrasonografia	Classificação de lesões	Especificidade, Estimador de Fisher, ROC (e área sob a curva), Precisão, Sensibilidade, Taxa de classificação correta, Valor preditivo negativo
12	Shen et al. [2007]	Câncer de mama	Ultrasonografia	Classificação de lesões	Nível de concordância entre os resultados dos radiologistas e os do sistema: Estatística de Kappa; Classificação: Especificidade, ROC (e área sob a curva), Precisão, Sensibilidade, Taxa de classificação correta, Valor preditivo negativo
13	Barhoumi et al. [2007]	Câncer de pele	Dermatoscopia	Classificação de lesões	Precisão <i>versus</i> Revocação, ROC, Taxa de classificação correta
14	Garnavi et al. [2011]	Câncer de pele	Dermatoscopia	Deteção das bordas de Melanomas	Erro de borda, Especificidade, Precisão, Sensibilidade, Similaridade, Taxa de deteção correta
15	Odeh et al. [2006]	Câncer de pele	Espectroscopia de Fluorescência	Classificação de lesões	Especificidade, Sensibilidade, Taxa de classificação correta
16	Ampeliotis et al. [2007]	Câncer de próstata	Ressonância Magnética	Classificação de tecidos prostáticos	Taxa de classificação correta
17	Schilham et al. [2006]	Câncer de pulmão	Radiografia	Deteção de nódulos	Segmentação: <i>Overlap</i> ; Deteção: FROC
18	Capobianco et al. [2008]	Câncer de pulmão	Tomografia Computadorizada	Deteção de nódulos	Nível de concordância entre os radiologistas: Estatística de Kappa; Deteção: Qtde. de FP por caso e por imagem, Sensibilidade
19	Gomathi e Thangaraj [2010]	Câncer de pulmão	Tomografia Computadorizada	Deteção de nódulos	Qtde. de FP por caso, Qtde. de VP detectados pelo CAD <i>versus</i> Qtde. de VP detectados por especialista, Sensibilidade
20	Itai et al. [2009]	Câncer de pulmão	Tomografia Computadorizada	Deteção de nódulos	Qtde. de FP por caso, Sensibilidade
21	Pietka et al. [2010] (a)	Câncer de pulmão	Tomografia Computadorizada	Deteção de nódulos	Sensibilidade, Taxa (%) de FP
22	Wada et al. [2006]	Câncer de pulmão	Tomografia Computadorizada	Deteção de nódulos	Qtde. de FP por slice, Precisão
23	Korfatis et al. [2007]	Câncer de pulmão	Tomografia Computadorizada	Segmentação dos pulmões	Segmentação: <i>Overlap</i> ; Corretitude da borda segmentada: Diferença na forma da borda segmentada, Máxima distância entre os volumes segmentados, Média; Valor quadrático médio
24	Kuang e Ye [2008]	Cáries dentárias	Radiografia	Deteção de cáries	Sensibilidade
25	Huang et al. [2009b]	Catarata nuclear	Lâmpada de Fenda	Classificação	Ranking dos casos: Ganho Cumulativo Descontado Normalizado; Classificação: Acurácia na predição de notas às imagens, Erro médio entre notas preditas e notas definidas pelos especialistas
26	Mumcuoglu et al. [2011]	Cicatrizes corticais renais	Cintilografia Cortical Renal	Deteção de cicatrizes corticais	FROC
27	Raja et al. [2010]	Cistos renais	Ultrasonografia	Classificação de casos	Medida F (<i>F-score</i>), ROC, Validação cruzada com resultados de análises de especialistas
28	Pietka et al. [2010] (b)	Esclerose múltipla	Ressonância magnética	Deteção de lesões cerebrais	Especificidade, Sensibilidade, Similaridade
29	Tan et al. [2010]	Glaucoma	Imagens de Fundo de Retina	Segmentação do cálice ótico	<i>Area overlap error</i> ($1 - \text{Overlap}$), <i>Relative area difference</i>
30	Chan [2010]	Hemorragia intracraniana aguda	Tomografia Computadorizada	Deteção	ROC (e área sob a curva), Sensibilidade, Taxa de FP por caso e por imagem
31	Huang et al. [2007]	Lesões ósseas	Cintilografia Óssea	Deteção de lesões	Qtde. de FP por caso, Sensibilidade
32	Huang et al. [2009a]	Lesões ósseas	Tomografia Computadorizada	Segmentação do esqueleto e Deteção de lesões	Qtde. de FP por caso, Sensibilidade
33	López et al. [2011]	Mal de Alzheimer	Tomografia Computadorizada	Deteção da doença	Especificidade, Razões de verossimilhança positiva e negativa, ROC (e área sob a curva), Sensibilidade, Taxa de deteção correta
34	Pietka et al. [2010] (c)	Pneumotórax	Tomografia Computadorizada	Segmentação	Especificidade, Sensibilidade
35	Sanchez et al. [2011]	Retinopatia diabética	Imagens de Fundo de Retina	Deteção de lesões	Especificidade, ROC (e área sob a curva), Sensibilidade
36	Jiménez et al. [2010]	Retinopatia diabética	Imagens de Fundo de Retina	Segmentação dos vasos da retina	Especificidade, Sensibilidade, Taxa de classificação correta (de pixels)

* <http://mosaic.utia.cas.cz>

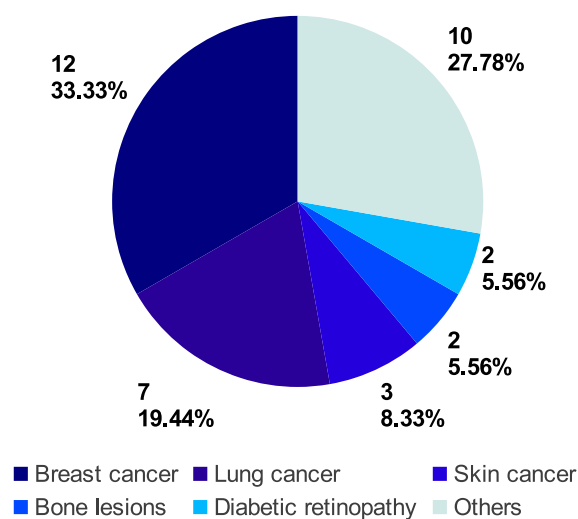


Figura 3. Quantidade de sistemas de CAD relatados nos trabalhos incluídos para cada anomalia.

com imagens provenientes da técnica de Espectroscopia de Fluorescência.

Os três sistemas relatados para doenças oculares que afetam a retina (Retinopatia Diabética e Glaucoma; IDs 29, 35 e 36) trabalham com o processamento de fotografias de retina. Já o sistema relatado para Catarata Nuclear (ID 25) trabalha com imagens obtidas por meio de Lâmpada de Fenda.

No que se refere aos sistemas relatados para a detecção de lesões ósseas, o primeiro (ID 31) trabalha com imagens provenientes de Cintilografia Óssea, enquanto que o segundo (ID 32) trabalha com imagens de Tomografia Computadorizada.

No gráfico apresentado na Figura 4 é apresentada uma visão geral da quantidade de sistemas relatados nos trabalhos incluídos para cada modalidade de imagens médicas. Nesse gráfico são considerados, para cada modalidade, todos os sistemas relatados independentemente da anomalia estudada. De acordo com esse resultado, imagens provenientes de Tomografia Computadorizada e Mamografia são as mais exploradas para o desenvolvimento de sistemas de CAD.

4.4. Tarefas para o Auxílio ao Diagnóstico

Cada técnica desenvolvida tem como objetivo uma tarefa específica para auxílio ao diagnóstico. Um sistema de CAD completo envolve, no geral, etapas de segmentação de estruturas, detecção de anomalias e extração de características das mesmas para uma posterior classificação do problema (normal, benigno(a), maligno(a), entre outras classes possíveis dependendo do caso). Como se pode deduzir, para se chegar, por exemplo, à classificação de estruturas, passa-se, no geral, por etapas prévias de segmentação e detecção. Trabalhos que apresentam esquemas de CAD procuram contribuir para a automatização de tais tarefas, quer seja por meio de técnicas desenvolvidas, quer seja por meio de alguma técnica já existente que foi melhorada. Os gráficos apresentados na Figura 5 apresentam, para cada tarefa de interesse, a quantidade de sistemas relatados cujo objetivo resumiu-se em contribuir principalmente com alguma inovação para a realização da tarefa.

A maior parte dos sistemas de CAD relatados para câncer de mama (gráfico da Figura 5(a))

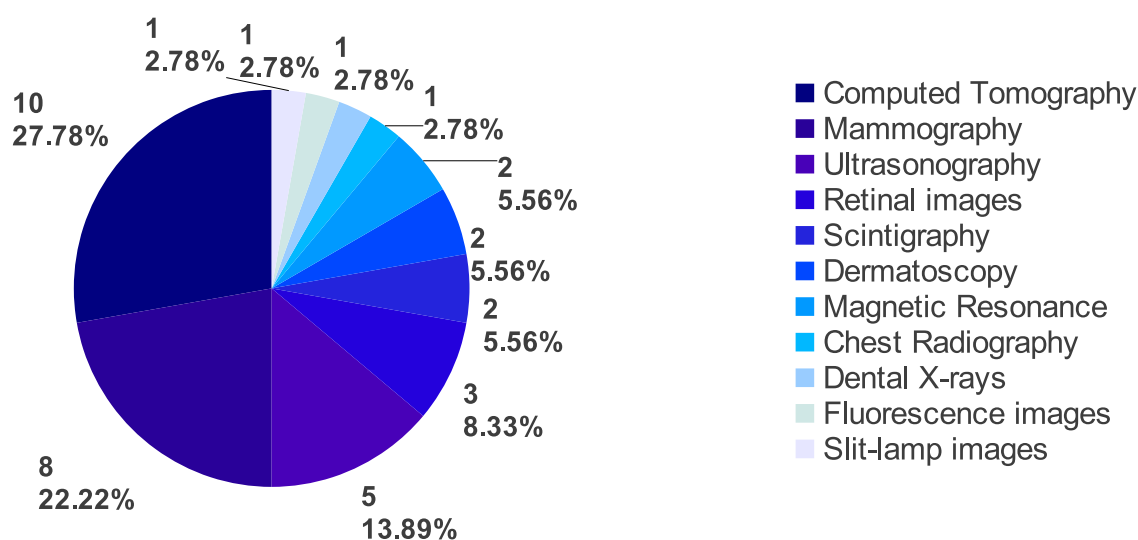


Figura 4. Quantidade de sistemas de CAD relatados para cada modalidade de imagens médicas.

objetivaram principalmente a classificação de microcalcificações, lesões ou regiões de interesse (ROIs – *Regions of Interest*). Isto evidencia que o maior interesse atualmente, quando se trata de CAD para câncer de mama, é a descoberta de estruturas suspeitas na imagem médica e a indicação de quais dessas estruturas podem ser consideradas normais, quais podem constituir achados cancerígenos benignos e quais podem constituir achados cancerígenos malignos. Tal tarefa é de extrema utilidade para auxiliar o médico no diagnóstico da doença e do nível de gravidade da mesma, diminuindo as chances de erros de interpretação por parte do profissional e aumentando a capacidade do mesmo de indicar tratamentos adequados para cada caso.

O trabalho de He et al. [2011] (ID 5) apresentou uma abordagem de classificação de risco para câncer de mama, isto é, calcular o risco de determinado paciente vir a desenvolver a doença. Tal risco é estimado por meio da análise de padrões do parênquima mamário utilizando-se imagens mamográficas.

No caso de sistemas de CAD para câncer de pulmão (gráfico da Figura 5(b)) observa-se que a maior parte dos sistemas relatados tiveram como objetivo principal a detecção de nódulos. Considerando-se todos os sistemas relatados (gráfico da Figura 5(c)), independentemente da doença estudada, observa-se que as principais tarefas de interesse foram detecção e classificação de anomalias. Tal resultado era de se esperar visto que são essas as tarefas que mais refletem a contribuição dos sistemas de CAD.

Vale a pena ressaltar também que diferentes técnicas têm sido propostas para a realização das diferentes tarefas mencionadas. Dentre os sistemas relatados nos trabalhos incluídos nesta revisão, aqueles que visaram a segmentação abordaram principalmente técnicas baseadas em Limiarização ou *Thresholding* (IDs 23 e 34), Modelos de Mistura Gaussiana (IDs 7 e 29), Correspondência por Padrão ou *Template Matching* (ID 6) e Crescimento de Regiões (IDs 8 e 36).

No caso dos sistemas que focaram na detecção de anomalias, observam-se abordagens diretamente ligadas aos resultados de técnicas de segmentação de estruturas nas imagens.

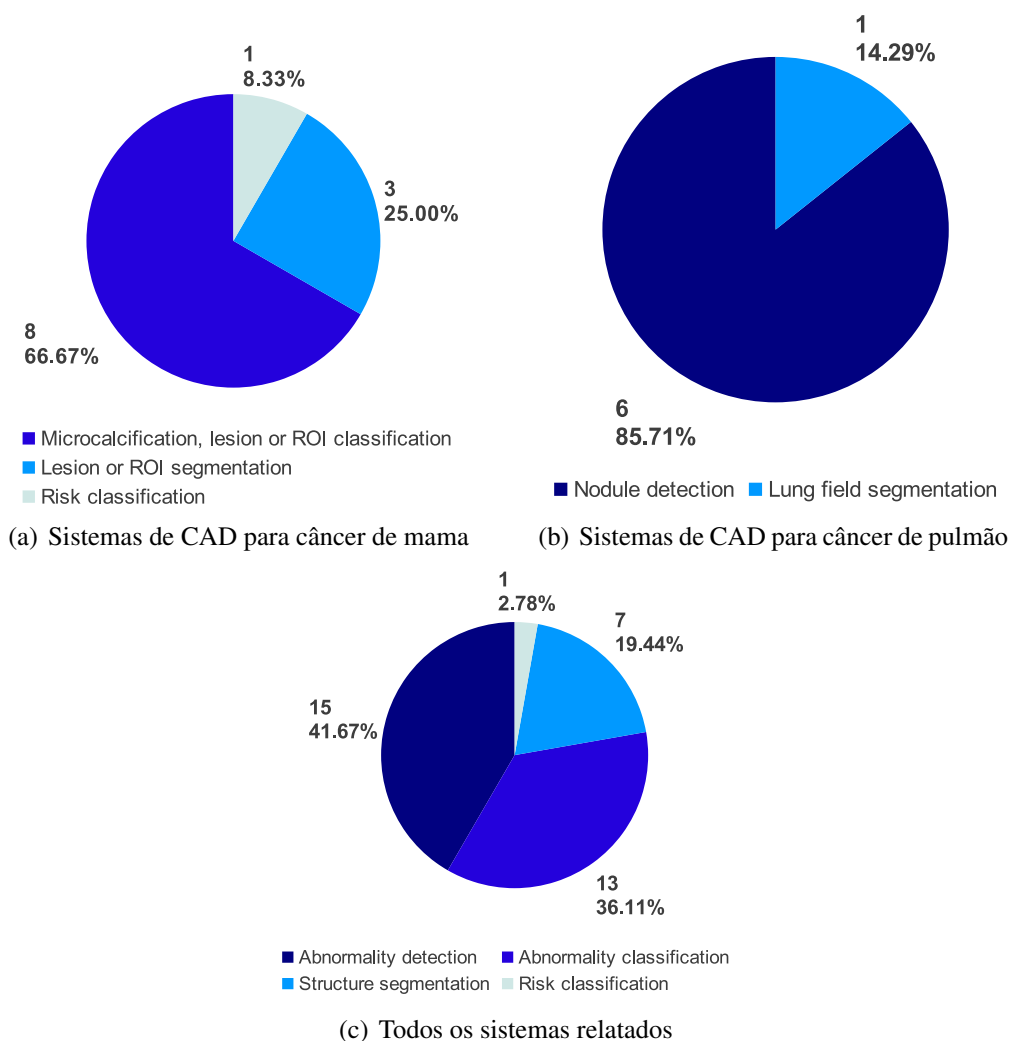


Figura 5. Quantidade de sistemas relacionados que enfatizaram principalmente cada tarefa de auxílio ao diagnóstico.

Dentre elas, destacam-se técnicas de segmentação baseadas na Teoria dos Conjuntos Nebulosos ou *Fuzzy* (IDs 21, 28 e 31), Limiarização (IDs 14 e 19) e baseadas em modelos (IDs 17 e 26). Observam-se também abordagens de detecção baseadas no emprego de classificadores, tais como Redes Neurais Artificiais (RNA) treinadas com o algoritmo Backpropagation (IDs 20 e 33), k-NN (*k-Nearest Neighbor*) (ID 17 e 35) e Máquina de Vetores Suporte (ID 24). Schilham et al. [2006] (ID 17) apresentaram uma abordagem combinada entre segmentação baseada em modelos e classificação utilizando o k-NN. Outras abordagens específicas foram apresentadas por Chan [2010] (ID 30) e Huang et al. [2009a] (ID 32). Capobianco et al. [2008] e Wada et al. [2006] (IDs 18 e 22) não mencionaram a abordagem implementada para a tarefa de detecção de nódulos pulmonares.

Por fim, para as tarefas de classificação observa-se o emprego de classificadores como k-NN (IDs 3, 5, 15 e 27), RNA treinadas com Backpropagation (IDs 2, 4 e 13), Redes Bayesianas (IDs 9, 10 e 16) e técnicas baseadas em Análise Discriminante Linear (ID 11) e Modelos de Regressão Logística (ID 12). Alguns trabalhos apresentaram abordagens combinadas de algoritmos de clusterização com algoritmos de classificação, como é o

caso em He et al. [2011] e Raja et al. [2010] (IDs 5 e 27), que utilizaram os algoritmos k-Means e k-NN, e Verma [2009] (ID 4) que utilizou um algoritmo de clusterização em conjunto com uma RNA treinada com Backpropagation. A abordagem de classificação de lesões de pele proposta por Barhoumi et al. [2007] combinou os resultados da RNA com resultados de um esquema de Recuperação de Imagens Baseada em Conteúdo (CBIR) por meio da Teoria de *Dempster-Shafer*. Huang et al. [2009b] (ID 25) apresentaram uma abordagem específica para a classificação de casos de Catarata Nuclear e Giannakopoulou et al. [2010] não especificaram o algoritmo ou técnica utilizada para a classificação de calcificações em mamogramas.

4.5. Bases Públicas de Imagens Médicas

Uma das grandes dificuldades enfrentadas no processo de desenvolvimento de sistemas de CAD é a obtenção de casos de teste. Obter uma ou mais bases de imagens médicas do domínio de interesse, contendo características de aquisição variadas, estruturas ou anomalias que a técnica propõem detectar e/ou analisar e um parecer correto de diagnóstico para cada caso presente na base, nem sempre é uma tarefa viável. Fatores como parcerias com clínicas e hospitais, questões éticas, permissões de acesso às imagens, entre outros, tendem a dificultar tal tarefa.

Por esse motivo, alguns projetos têm sido desenvolvidos e mantidos ao redor do mundo a fim de disponibilizar imagens médicas para grupos de pesquisa que queiram trabalhar com pesquisa e desenvolvimento na área de CAD. Tais projetos resultam em bases públicas de imagens médicas que contêm uma documentação referente aos laudos realizados por médicos experientes incluindo, algumas vezes, informações relativas ao posicionamento de estruturas de interesse nas imagens.

Dos sistemas relatados nos trabalhos incluídos, 13 foram avaliados utilizando-se bases públicas. As bases relatadas a seguir foram catalogadas dos artigos incluídos. Após a referência ao nome de cada base, seguem entre parênteses os IDs (de acordo com a Tabela 4) dos sistemas relatados que foram testados com imagens e dados provenientes da base.

DDSM – Digital Database for Screening Mammography⁷ (IDs 4, 6, 7 e 8). Mantida pela *University of South Florida*, essa base serve como recurso para a pesquisa em análise de imagens mamográficas [Heath et al. 2001]. A base contém 2620 casos divididos em 43 volumes, cada um composto somente por casos normais, casos contendo estruturas suspeitas comprovadamente benignas ou casos comprovados de câncer. São fornecidos também a demarcação de bordas das lesões, quando existentes, e um conjunto de programas para decodificação e manipulação dos mamogramas.

The mini-MIAS database of mammograms⁸ (IDs 2, 3 e 5). Essa base é mantida pela *Mammographic Image Analysis Society* (MIAS) e oferece 322 mamogramas para utilização em trabalhos de pesquisa. Além dos mamogramas, a base contém informações referentes ao tipo, gravidade e coordenadas do pixel central da anomalia, quando existente, de cada imagem [Suckling et al. 1994].

LIDC – Lung Image Database Consortium⁹ (IDs 21 e 23). Essa base pública é

⁷<http://figment.csee.usf.edu/Mammography/Database.html>

⁸<http://peipa.essex.ac.uk/info/mias.html>

⁹<http://imaging.cancer.gov/programsandresources/informationssystem/>

fruto de uma iniciativa que teve como objetivo fornecer imagens de Tomografia Computadorizada torácica para suportar o desenvolvimento, treinamento e avaliação de esquemas de CAD para detecção de nódulos pulmonares [Armato III et al. 2011]. A base fornece exames de 1010 pacientes, bem como a demarcação de bordas das lesões existentes e um software para a manipulação das imagens.

JSRT – Japanese Society of Radiological Technology database¹⁰ (ID 17). A base de imagens radiológicas de tórax da JSRT foi criada em cooperação com a Sociedade Japonesa de Radiologia em 1998. São 247 imagens, sendo 154 casos de único nódulo pulmonar (agrupados por sutileza) e 93 casos livres de nódulos. São fornecidas ainda informações a respeito dos pacientes (gênero e idade), diagnóstico (benigno ou maligno) e coordenadas centrais de cada nódulo [Shiraishi et al. 2000].

MESSIDOR Digital Retinal Images¹¹ (ID 35). A base MESSIDOR foi desenvolvida com o objetivo de facilitar estudos de CAD para Retinopatia Diabética. Ela contém 1200 imagens de fundo de olho. Além disso, fornece também, para cada imagem, notas atribuídas por especialistas indicando o nível da Retinopatia, quando existente, e o risco de existência de edema macular¹².

DRIVE – Digital Retinal Images for Vessel Extraction¹³ (ID 36). Essa base foi estabelecida com o intuito de possibilitar o estudo comparativo em segmentação dos vasos sanguíneos em imagens de retina. Para tanto, são disponibilizadas 40 fotografias de retina (fundo de olho). Para cada fotografia, existem duas imagens com segmentações manuais dos vasos sanguíneos na imagem [Staal et al. 2004]. Os pesquisadores que utilizam a base para o teste de métodos de segmentação dos vasos sanguíneos podem submeter seus resultados por meio da *home-page* do projeto, a fim de disponibilizá-los para a comparação com outros trabalhos da área.

STARE – Structured Analysis of the Retina¹⁴ (ID 36). O projeto STARE, iniciado em 1975 na Universidade da Califórnia, consiste no desenvolvimento de um sistema para auxílio na detecção de doenças do olho humano [McCormick e Goldbaum 1975]. O projeto possui uma base que contém 402 fotografias de fundo de olho e uma lista com os diagnósticos verdadeiros para cada imagem.

ADNI – Alzheimer’s Disease Neuroimaging Initiative¹⁵ (ID 33). A ADNI [Mueller et al. 2005] tem como missão definir a progressão da Doença de Alzheimer (ou Mal de Alzheimer). Para tanto, objetiva coletar e validar dados como imagens provenientes de Tomografia por Emissão de Pósitrons (PET) ou Ressonância Magnética (MRI) e outras informações como preditores da doença. A iniciativa disponibiliza os dados coletados para fins de pesquisa. Ao todo, a base contém dados e imagens de 895 pacientes.

l1dc

¹⁰http://www.jsrt.or.jp/web_data/english03.php

¹¹<http://messidor.crihan.fr/download-en.php>

¹²Informações extraídas da *home-page* do projeto.

¹³<http://www.isi.uu.nl/Research/Databases/DRIVE/>

¹⁴<http://www.ces.clemson.edu/~ahoover/stare/>

¹⁵<http://adni.loni.ucla.edu/>

4.6. Avaliação de Sistemas de CAD

Como mencionado anteriormente, o principal objetivo da presente RS consistiu no levantamento e análise do estado da arte com relação à avaliação de sistemas de CAD. Já foram apresentados as anomalias estudadas, modalidades de imagens médicas, tarefas de interesse e bases públicas utilizadas no processo de desenvolvimento e teste dos sistemas relatados nos trabalhos incluídos.

Da Tabela 4 pode-se deduzir que os processos de avaliação dos sistemas de CAD envolvem, no geral, executar as rotinas de segmentação, detecção ou classificação sobre um conjunto de entradas e, com base em resultados corretos previamente conhecidos (fornecidos na maioria das vezes por especialistas experientes), obter métricas de desempenho para o sistema.

4.6.1. Métricas de Avaliação

Nos gráficos apresentados na Figura 6 são relatados, para cada tarefa de auxílio ao diagnóstico, as métricas e os métodos de avaliação aplicados aos sistemas relatados nos trabalhos incluídos. Foram considerados somente métricas e métodos mencionados pelos autores nos artigos incluídos. Vale ressaltar que, para a análise de tais resultados, não foram consideradas somente as tarefas principais almejadas por cada sistema. Isto é, se um trabalho cuja tarefa de interesse principal do sistema relatado é a classificação de lesões apresenta também resultados de avaliação para uma tarefa prévia de segmentação (mencionando as métricas e/ou métodos utilizados), tais métricas e/ou métodos foram incluídos no gráfico relativo à segmentação (Figura 6(a)). Para cada gráfico, é mencionada a quantidade total de sistemas relatados considerados.

No gráfico da Figura 6(a) observa-se que metade dos sistemas que avaliaram algum método de segmentação aplicou a medida de *Overlap*. Essa medida consiste na área relativa da intersecção entre duas regiões consideradas [Grusauskas et al. 2008, Korfiatis et al. 2007]. Considerando-se A_{seg} , a área de uma região segmentada automaticamente, e A_{man} , a área considerada correta para o processo de segmentação (gerada, por exemplo, manualmente), a medida de *Overlap* é definida na Equação 1. O valor 0 indica o pior desempenho, ou seja, não há intersecção entre a área correta e a área obtida automaticamente. O valor 1, por sua vez, indica uma segmentação perfeita.

$$Overlap = \frac{A_{seg} \cap A_{man}}{A_{seg} \cup A_{man}} \quad (1)$$

A métrica *Relative Area Difference*, aplicada na avaliação de dois dos sistemas relatados (IDs 8 e 29), mede a extensão da região segmentada automaticamente que não coincide com a região correta esperada [Tan et al. 2010]. Tal medida pode ser obtida por meio da Equação 2. Observa-se que, se $A_{seg} = A_{man}$, então a *Relative Area Difference* é 0.

$$Relative\ Area\ Difference = \frac{A_{seg} - A_{man}}{A_{man}} \quad (2)$$

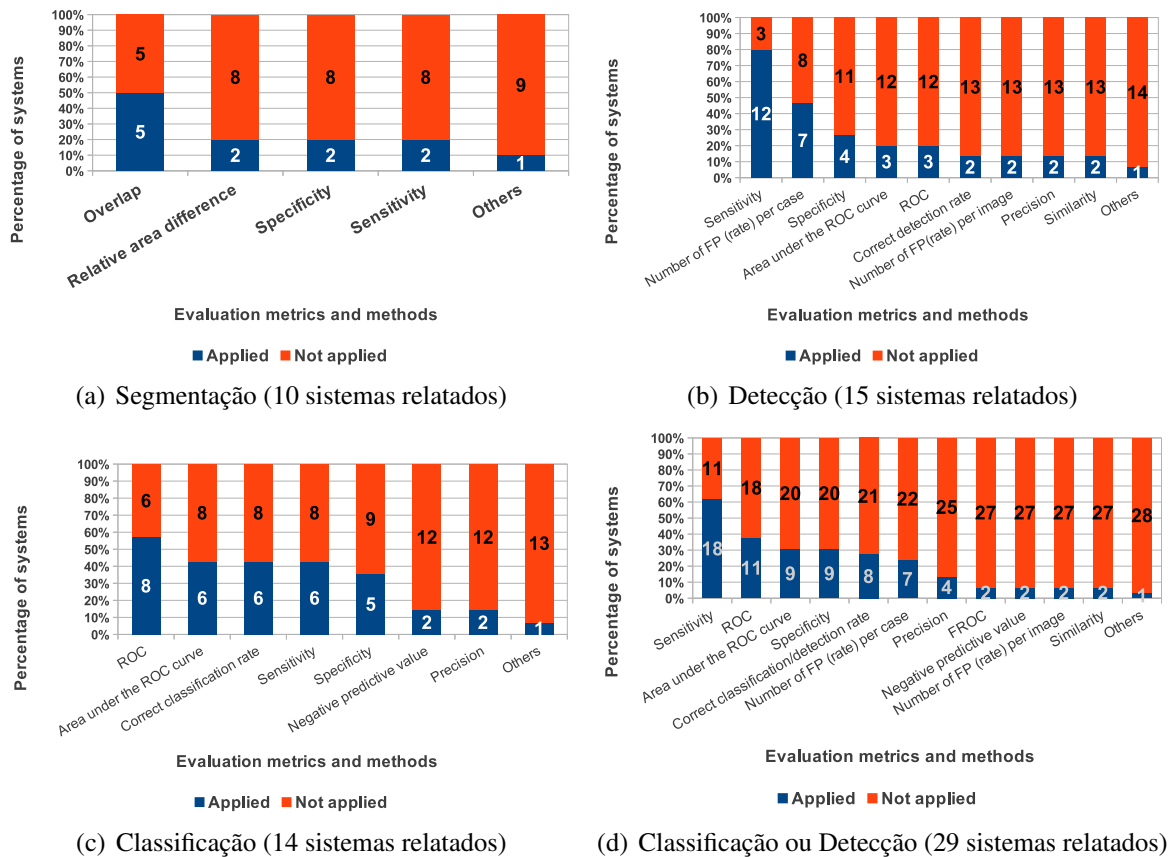


Figura 6. Métricas e métodos de avaliação empregados nas diferentes tarefas de auxílio ao diagnóstico.

Outras métricas também observadas para a avaliação de métodos de segmentação são especificidade e sensibilidade. Tais métricas fazem parte de um conjunto de métricas estatísticas muito tradicionais na avaliação de sistemas de CAD. Elas se baseiam em VP, VN, FP e FN [Garnavi et al. 2011].

No caso da segmentação, uma abordagem para a utilização de tais métricas é definir pixels VP (pertencem à região de interesse e foram segmentados), VN (não pertencem à região de interesse e não foram segmentados), FP (não pertencem à região de interesse, mas foram segmentados) e FN (pertencem à região de interesse, mas não foram segmentados). Adiante, nesta subseção, tais métricas são apresentadas. Os demais métodos e métricas de avaliação, aplicáveis em rotinas de segmentação, observados nos sistemas relatados podem ser conferidos na Tabela 4 e referências.

As métricas e os métodos de avaliação aplicados para as tarefas de classificação e detecção são, em sua maioria, as mesmas. Por esse motivo, os gráficos das Figuras 6(c) e 6(b) apresentam as métricas e os métodos aplicados para cada tarefa em particular e, no gráfico da Figura 6(d), os resultados para as duas tarefas são combinados.

Como se pode observar pelos gráficos, houve predominância do uso de métricas baseadas em VP, VN, FP e FN para a avaliação de tarefas de classificação e detecção nos sistemas relatados. Na combinação dos casos, a sensibilidade demonstra ser a métrica mais considerada para se avaliar um sistema de CAD. Na Tabela 5 são definidas as prin-

cipais métricas relatadas nos trabalhos incluídos com base em Garnavi et al. [2011].

Tabela 5. Principais métricas de avaliação relatadas nos trabalhos incluídos.

MÉTRICA	FÓRMULA	DESCRIÇÃO
Sensibilidade (ou Revocação)	$\frac{TP}{TP + FN} (\times 100\%)$	Porcentagem de anormalidades que foram detectadas / classificadas corretamente.
Especificidade	$\frac{TN}{TN + FP} (\times 100\%)$	Porcentagem de estruturas normais que não foram detectadas / classificadas incorretamente como possíveis anormalidades.
Taxa de classificação / detecção correta	$\frac{TP + TN}{TP + TN + FP + FN} (\times 100\%)$	Porcentagem de anormalidades e estruturas normais classificadas / detectadas corretamente.
Similaridade	$\frac{2TP}{2TP + FN + FP} (\times 100\%)$	Representação do nível de concordância entre os resultados obtidos e os resultados esperados (tomados como verdade).
Precisão (or Valor preditivo positivo)	$\frac{TP}{TP + FP} (\times 100\%)$	Porcentagem de estruturas detectadas que realmente são anormalidades.
Valor preditivo negativo	$\frac{TN}{TN + FN} (\times 100\%)$	Porcentagem de estruturas normais detectadas / classificadas que realmente não representam anormalidades.

Um método muito conhecido e empregado na avaliação de sistemas de CAD é a curva ROC. Como se pode observar nos gráficos da Figura 6, tal método foi empregado na avaliação de mais da metade (57,14%) dos sistemas de classificação relatados. Na gráfico que combina os métodos e métricas aplicados à classificação e à detecção (Figura 6(d)) a curva ROC também se destaca.

Uma métrica de avaliação extraída da curva ROC e também frequentemente empregada no processo de avaliação dos sistemas é a *área sob a curva ROC*. Quanto maior a área sob a curva, maior o desempenho do sistema de CAD avaliado. Mais detalhes a respeito de tal métrica podem ser obtidos em Metz [1999] e nos trabalhos incluídos que a mencionam.

4.6.2. Limitações

Métricas de avaliação baseadas em VP, VN, FP e FN dependem de um conhecimento prévio das características dos casos utilizados para se testar o sistema de CAD. Ou seja, é necessário saber, por exemplo, quais estruturas correspondem verdadeiramente a lesões em uma imagem mamográfica e quais são estruturas normais que podem vir a ser consideradas lesões por engano.

Todo o trabalho de análise “visual” de cada caso e a comparação dos resultados tomados como verdade com os resultados obtidos por meio do sistema de CAD deve ser realizado por, ou em conjunto com, um profissional da área médica experiente (médico, radiologista). O trabalho de teste e avaliação de sistemas de CAD utilizando-se tais métricas é, portanto, grande e desgastante. Isso porque sempre que se modifica alguma parte do sistema, toda a comparação entre os resultados obtidos e aqueles considerados corretos deve ser refeita.

5. Considerações Finais

Neste Relatório Técnico foram apresentados os resultados de uma Revisão Sistemática que possibilitou o levantamento e a análise do estado da arte com relação a projeto, desenvolvimento e avaliação de sistemas de diagnóstico auxiliado por computador. Foram

catalogados 36 sistemas de CAD projetados para a automatização de diferentes tarefas de auxílio ao diagnóstico. Tais sistemas são descritos nos trabalhos incluídos na RS e recuperados em cinco bases de trabalhos científicos publicados.

Assim como acontece em diversas áreas de pesquisa, o desenvolvimento de CAD é contemplado por diversos grupos ao redor do mundo. Assim, há uma vasta quantidade de artigos publicados, considerando-se doenças e modalidades de imagens médicas diversas, como foi apresentado no presente trabalho.

A execução de revisões não sistemáticas pode explorar o estado da arte de forma incompleta ou, ainda, levar a retrabalho devido à ausência de registro dos passos conduzidos durante a pesquisa. Nesse contexto, a execução de uma revisão sistemática especificamente em CAD contribui tanto para proporcionar uma visão geral a pesquisadores que estão iniciando seus trabalhos na área, assim como com grupos de pesquisa que desejam encontrar assuntos ainda pouco explorados relacionados ao tema. Além disso, a possibilidade de atualizar continuamente a revisão sem a necessidade de refazer o trabalho já divulgado e a facilidade de auditar os resultados apresentados constituem diferenciais que proporcionam uma maior produtividade na pesquisa bibliográfica.

Os resultados confirmam as métricas tradicionais baseadas em verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos como os principais meios de se avaliar e comparar o desempenho de tais sistemas. Curvas ROC e FROC consistem em métodos derivados de tais métricas que auxiliam no processo de avaliação do comportamento dos sistemas, dadas variações em seus parâmetros específicos.

Como limitação de tais métricas e métodos, observa-se a repetitiva e desgastante participação de médicos e radiologistas no processo de indicar, para cada versão do sistema, quais resultados são corretos e quais não são.

Agradecimentos

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) – Processo 559931/2010-7, à Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp) – Processos 2010/01496-1 e 2010/15691-0 e ao Instituto Nacional de Ciência e Tecnologia – Medicina Assistida por Computação Científica (INCT-MACC), pelo apoio financeiro.

Referências

- Ampeliotis, D., Antonakoudi, A., Berberidis, K., e Psarakis, E. Z. (2007). Computer aided detection of prostate cancer using fused information from dynamic contrast enhanced and morphological magnetic resonance images. Em *IEEE International Conference on Signal Processing and Communications*, páginas 888–891.
- Armato III, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Zhao, B., Aberle, D. R., Henschke, C. I., Hoffman, E. A., Kazerooni, E. A., MacMahon, H., van Beek, E. J. R., Yankelevitz, D., Biancardi, A. M., Bland, P. H., Brown, M. S., Engelmann, R. M., Laderach, G. E., Max, D., Pais, R. C., Qing, D. P.-Y., Roberts, R. Y., Smith, A. R., Starkey, A., Batra, P., Caligiuri, P., Farooqi, A., Gladish, G. W., Jude, C. M., Munden, R. F., Petkovska, I., Quint, L. E., Schwartz, L. H., Sundaram, B., Dodd, L. E., Fenimore, C., Gur, D., Petrick, N., Freymann, J.,

- Kirby, J., Hughes, B., Castele, A. V., Gupte, S., Sallam, M., Heath, M. D., Kuhn, M. H., Dharaiya, E., Burns, R., Fryd, D. S., Salganicoff, M., Anand, V., Shreter, U., Vastagh, S., Croft, B. Y., e Clarke, L. P. (2011). The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans. *Med. Phys.*, 38(2):915–931.
- Barhoumi, W., Dhahbi, S., e Zagrouba, E. (2007). A collaborative system for pigmented skin lesions malignancy tracking. Em *IEEE International Workshop on Imaging Systems and Techniques*, páginas 1–6.
- Biolchini, J. C. d. A., Mian, P. G., Natali, A. C. C., Conte, T. U., e Travassos, G. H. (2007). Scientific research ontology to support systematic review in software engineering. *Adv. Eng. Inform.*, 21(2):133–151.
- Brown, D. G. (2001). The evaluation of computer-aided diagnosis systems: an FDA perspective. Em *Proceedings of the 30th Applied Imagery Pattern Recognition Workshop*, páginas 17–20.
- Capobianco, J., Jasinowodolinski, D., e Szarf, G. (2008). Diagnóstico auxiliado por computador na detecção de nódulos pulmonares pela tomografia computadorizada com múltiplos detectores: estudo preliminar de 24 casos. *J. Bras. Pneumol.*, 34:27–33.
- Chan, T. (2010). Clinical usage considerations in the development and evaluation of a computer aided diagnosis system for acute intracranial hemorrhage on brain CT. Em Zhang, D. e Sonka, M., editores, *Medical Biometrics*, volume 6165 of *Lecture Notes in Computer Science*, páginas 268–275. Springer Berlin / Heidelberg.
- Doi, K. (2006). Diagnostic imaging over the last 50 years: research and development in medical imaging science and technology. *Phys. Med. Biol.*, 51(13):R5–R27.
- Doi, K. (2007). Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Comput. Med. Imaging Graph.*, 31(4-5):198–211.
- Garnavi, R., Aldeen, M., e Celebi, M. E. (2011). Weighted performance index for objective evaluation of border detection methods in dermoscopy images. *Skin Res. Technol.*, 17(1):35–44.
- Giannakopoulou, G., Spyrou, G. M., Antarakis, A., Andreadis, I., Koulocheri, D., Zagouri, F., Nonni, A., Filippakis, G. M., Nikita, K. S., Ligomenides, P. A., e Zografos, G. C. (2010). Downgrading BIRADS 3 to BIRADS 2 category using a computer-aided microcalcification analysis and risk assessment system for early breast cancer. *Comput. Biol. Med.*, 40(11-12):853–859.
- Giger, M. L. (1999). Overview of computer-aided diagnosis in breast imaging. Em Doi, K., MacMahon, H., Giger, M. L., e Hoffmann, K. R., editores, *Computer-Aided Diagnosis in Medical Imaging*, volume 1182 of *International Congress Series*, páginas 167–176. Elsevier Science BV, 1 edição.
- Gomathi, M. e Thangaraj, P. (2010). Automated CAD for detection of lung nodule using CT scans. Em *Proceedings of the Third Annual ACM Bangalore Conference*, páginas 25:1–25:4, New York, NY, USA. ACM.
- Grusauskas, N. P., Drukker, K., Giger, M. L., Chang, R.-F., Sennett, C. A., Moon, W. K., e Pesce, L. L. (2009). Breast US computer-aided diagnosis system: robustness across urban populations in south korea and the united states. *Radiol.*, 253(3):661–671.
- Grusauskas, N. P., Drukker, K., Giger, M. L., Sennett, C. A., e Pesce, L. L. (2008). Performance of breast ultrasound computer-aided diagnosis: dependence on image selection. *Acad. Radiol.*, 15(10):1234–1245.

- Haindl, M., Mikeš, S., e Scarpa, G. (2007). Unsupervised detection of mammogram regions of interest. Em *Proceedings of the 11th International Conference on Knowledge-based and Intelligent Information and Engineering Systems and 17th Italian Workshop on Neural Networks*, páginas 33–40, Berlin, Heidelberg. Springer-Verlag.
- He, W., Denton, E. R. E., Stafford, K., e Zwiggelaar, R. (2011). Mammographic image segmentation and risk classification based on mammographic parenchymal patterns and geometric moments. *Biomed. Signal Process. Control.*, 6(3):321–329.
- Heath, M., Bowyer, K., Kopans, D., Moore, R., e Kegelmeyer, W. P. (2001). The Digital Database for Screening Mammography. Em Yaffe, M. J., editor, *Proceedings of the Fifth International Workshop on Digital Mammography*, páginas 212–218. Medical Physics Publishing.
- Huang, J.-Y., Kao, P.-F., e Chen, Y.-S. (2007). A set of image processing algorithms for computer-aided diagnosis in nuclear medicine whole body bone scan images. *IEEE Trans. Nucl. Sci.*, 54(3):514–522.
- Huang, S.-F., Chaoa, H.-Y., Hsu, C.-C., Yang, S.-F., e Kao, P.-F. (2009a). A computer-aided diagnosis system for whole body bone scan using single photon emission computed tomography. Em *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, páginas 542–545.
- Huang, W., Li, H., Chan, K., Lim, J., Liu, J., e Wong, T. (2009b). A computer-aided diagnosis system of nuclear cataract via ranking. Em Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., e Taylor, C., editores, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2009*, volume 5762 of *Lecture Notes in Computer Science*, páginas 803–810. Springer Berlin / Heidelberg.
- Itai, Y., Kim, H., Ishikawa, S., Katsuragawa, S., e Doi, K. (2009). Reduction of FPs for lung nodules in MDCT by use of temporal subtraction with voxel-matching technique. Em Koppen, M., Kasabov, N., e Coghill, G., editores, *Advances in Neuro-Information Processing*, volume 5506 of *Lecture Notes in Computer Science*, páginas 504–512. Springer Berlin / Heidelberg.
- Jasmine, J., Govardhan, A., e Baskaran, S. (2009). Microcalcification detection in digital mammograms based on wavelet analysis and neural networks. Em *2009 International Conference on Control, Automation, Communication and Energy Conservation*, páginas 1–6.
- Jiménez, S., Alemany, P., Fondón, I., Foncubierta, A., Acha, B., e Serrano, C. (2010). Detección automática de vasos en retinografías. *Arch. Soc. Esp. Oftalmol.*, 85:103–109.
- Kitchenham, B. A. (2004). Procedures for performing systematic reviews. Technical report, Computer Science Department, Keele University.
- Korfiatis, P., Skiadopoulou, S., Sakellaropoulos, P., Kalogeropoulou, C., e Costaridou, L. (2007). Automated 3D segmentation of lung fields in thin slice CT exploiting wavelet preprocessing. Em *Proceedings of the 12th International Conference on Computer Analysis of Images and Patterns*, páginas 237–244, Berlin, Heidelberg. Springer-Verlag.
- Kuang, W. e Ye, W. (2008). A kernel-modified SVM based computer-aided diagnosis system in initial caries. volume 3, páginas 207–211.
- Lee, H.-W., Liu, B.-D., Hung, K.-C., Lei, S.-F., Wang, P.-C., e Yang, T.-L. (2009). Breast tumor classification of ultrasound images using wavelet-based channel energy and ImageJ. *IEEE J. Sel. Top. Signal Process.*, 3(1):81–93.

- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., e Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med.*, 6(7):e1000100–1–e1000100–28.
- López, M., Ramírez, J., Górriz, J., Álvarez, I., Salas-Gonzalez, D., Segovia, F., Chaves, R., Padilla, P., e Gómez-Río, M. (2011). Principal component analysis-based techniques and supervised classification schemes for the early detection of alzheimer's disease. *Neurocomputing*, 74(8):1260–1271.
- McCormick, B. e Goldbaum, M. (1975). STARE - Structured Analysis of the Retina: image processing of TV fundus image. Em *USA-Japan Workshop on Image Processing*, Pasadena, CA, USA. Jet Propulsion Laboratory.
- Metz, C. E. (1999). Evaluation of cad methods. Em Doi, K., MacMahon, H., Giger, M. L., e Hoffmann, K. R., editores , *Computer-Aided Diagnosis in Medical Imaging*, volume 1182 of *International Congress Series*, páginas 543–554, Amsterdam, Netherlands. Elsevier Science BV.
- Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C. R., Jagust, W., Trojanowski, J. Q., Toga, A. W., e Beckett, L. (2005). Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimer's Dement.*, 1(1):55–66.
- Mumcuoglu, E. U., Bozkurt, F. M., Aslan, M., Sener, E., e Ugur, O. (2011). Computerized scar detection on renal cortical scintigraphy images. *Nucl. Med. Commun.*, 32(11):1070–1078.
- Odeh, S., Ros, E., Rojas, I., e Palomares, J. (2006). Skin lesion diagnosis using fluorescence images. Em Campilho, A. e Kamel, M., editores , *Image Analysis and Recognition*, volume 4142 of *Lecture Notes in Computer Science*, páginas 648–659. Springer Berlin / Heidelberg.
- Osman, M., Wahed, M., Mohamed, A., e Kadah, Y. (2009). Computer aided diagnosis system for classification of microcalcifications in digital mammograms. Em *National Radio Science Conference*, páginas 1–6.
- Pietka, E., Kawa, J., Badura, P., e Spinczyk, D. (2010). Open architecture computer-aided diagnosis system. *Expert Syst.*, 27(1):17–39.
- Pietka, E., Kawa, J., Spinczyk, D., Badura, P., Wieclawek, W., Czajkowska, J., e Rudzki, M. (2011). Role of radiologists in CAD life-cycle. *Eur. J. Radiol.*, 78(2):225–233.
- Raja, K. B., Madheswaran, M., e Thyagarajah, K. (2010). Texture pattern analysis of kidney tissues for disorder identification and classification using dominant gabor wavelet. *Mach. Vis. Appl.*, 21:287–300.
- Sanchez, C. I., Niemeijer, M., Dumitrescu, A. V., Suttorp-Schulten, M. S. A., Abramoff, M. D., e van Ginneken, B. (2011). Evaluation of a computer-aided diagnosis system for diabetic retinopathy screening on public data. *Invest. Ophthalmol. Vis. Sci.*, 52(7):4866–4871.
- Schilham, A. M., van Ginneken, B., e Loog, M. (2006). A computer-aided diagnosis system for detection of lung nodules in chest radiographs with an evaluation on a public database. *Med. Image Anal.*, 10(2):247–258.
- Shen, W.-C., Chang, R.-F., e Moon, W. K. (2007). Computer aided classification system for breast ultrasound based on breast imaging reporting and data system (BI-RADS). *Ultrasound Med. Biol.*, 33(11):1688–1698.

- Shiraishi, J., Katsuragawa, S., Ikezoe, J., Matsumoto, T., Kobayashi, T., Komatsu, K.-i., Matsui, M., Fujita, H., Kodera, Y., e Doi, K. (2000). Development of a digital image database for chest radiographs with and without a lung nodule. *AJR Am. J. Roentgenol.*, 174(1):71–74.
- Shiraishi, J., Pesce, L. L., Metz, C. E., e Doi, K. (2009). Experimental design and data analysis in Receiver Operating Characteristic studies: lessons learned from reports in Radiology from 1997 to 2006. *Radiol.*, 253(3):822–830.
- Song, E., Xu, S., Xu, X., Zeng, J., Lan, Y., Zhang, S., e Hung, C.-C. (2010). Hybrid segmentation of mass in mammograms using template matching and dynamic programming. *Acad. Radiol.*, 17(11):1414–1424.
- Staal, J., Abramoff, M. D., Niemeijer, M., Viergever, M. A., e van Ginneken, B. (2004). Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging*, 23(4):501–509.
- Suckling, J., Parker, J., Dance, D. R., Astley, S., Hutt, I., Boggis, C., Ricketts, I., Stamatakis, E., Cerneaz, N., Kok, S. L., Taylor, P., Betal, D., e Savage, J. (1994). The Mammographic Image Analysis Society digital mammogram database. Em Gale, A. G., Astley, S. M., Dance, D. R., e Cairns, A. Y., editores , *Proceedings of the Second International Workshop on Digital Mammography*, Amsterdam. Excerta Medica.
- Tan, N. M., Liu, J., Wong, D. W. K., Yin, F., Lim, J. H., e Wong, T. Y. (2010). Mixture model-based approach for optic cup segmentation. Em *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, páginas 4817–4820.
- van Ginneken, B., Armato III, S. G., de Hoop, B., van Amelsvoort-van de Vorst, S., Duindam, T., Niemeijer, M., Murphy, K., Schilham, A., Retico, A., Fantacci, M. E., Camarlinghi, N., Bagagli, F., Gori, I., Hara, T., Fujita, H., Gargano, G., Bellotti, R., Tangaro, S., Bolaños, L., Carlo, F. D., Cerello, P., Cheran, S. C., Torres, E. L., e Prokop, M. (2010). Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the ANODE09 study. *Med. Image Anal.*, 14(6):707–722.
- Verma, B. (2009). Impact of multiple clusters on neural classification of ROIs in digital mammograms. Em *Proceedings of the 2009 International Joint Conference on Neural Networks*, páginas 2532–2535, Piscataway, NJ, USA. IEEE Press.
- Wada, S., Matsumoto, T., Murao, K., e Sone, S. (2006). A study on the performance evaluation of computer-aided diagnosis for detecting pulmonary nodules for the various CT reconstruction. Em Jiang, Y. e Eckstein, M. P., editores , *Medical Imaging 2006: Image Perception, Observer Performance and Technology Assessment*, volume 6146 of *Proceedings of the Society of photo-optical instrumentation engineers (SPIE)*, páginas 61461C–1–61461C–8. SPIE.
- Zheng, B., Pu, J., Park, S. C., Zuley, M., e Gur, D. (2008). Assessment of the relationship between lesion segmentation accuracy and computer-aided diagnosis scheme performance. Em Giger, M. L. e Karssemeijer, N., editores , *Medical Imaging 2008: Computer-Aided Diagnosis*, volume 6915 of *Proceedings of the Society of photo-optical instrumentation engineers (SPIE)*, páginas 691530–1–691530–11. SPIE.