

Reconhecimento de câncer de pele do tipo melanoma

Rômulo Lopes Frutuoso¹, Jonas Rodrigues Vieira Santos¹, Robson S. Siqueira¹

¹Instituto Federal de Educação, Ciência e Tecnologia – IFCE, Campus Maracanaú
Ciência da Computação

{frutuoso.romulo, jonascomputacao, siqueira.robson.dasilva}@gmail.com

***Abstract.** This work seeks recognize the skin cancer melanoma type with digital image processing. The work was divided into two stages: the first takes care of the image feature extraction, based on the ABCD rule, while the second, aims to classify the result of the previous step. We used three different algorithms for pattern recognition, DMC, KNN and Naive Bayes. The three algorithms achieved accuracy rates of over 70% and the Naive Bayes had the best result with over 90% accuracy.*

***Resumo.** O presente trabalho visa reconhecer o câncer de pele do tipo melanoma com processamento digital de imagens. O trabalho foi dividido em duas etapas: extração de características da imagem, com base na regra ABCD e classificação das mesmas. Foram utilizados três algoritmos diferentes de reconhecimento de padrões: DMC, KNN e Naive Bayes. Os três algoritmos conseguiram taxas de acerto de mais de 70% sendo o Naive Bayes o de melhor resultado com mais de 90% de acerto.*

1. Introdução

A área de Processamento Digital de Imagens (PDI) teve grande impulso com o advento dos primeiros computadores digitais de grande porte e o início do programa espacial norte-americano [Filho e Neto 1999]. Nesse contexto, com o aumento da capacidade de processamento dos computadores pessoais, temos grande parte da computação antes efetuada apenas por supercomputadores, sendo efetivamente desenvolvida em máquinas comuns. Assim, o PDI que algumas décadas atrás era apenas realizado em supercomputadores, pode ser realizado em computadores pessoais.

Conseqüentemente, muitos problemas reais usam PDI, por exemplo: reconhecimento de impressão de polegar (biometria), leitura automática de placa de veículos e inspeção de trilhas de circuito impresso [Gonzales e Woods 2010]. Uma das áreas médicas onde o estudo destaca-se é na dermatologia. (Ganzeli *et al.* 2011, Rosado 2009 e Sobieranski *et al.* 2007), inclusive, com o mesmo propósito do presente trabalho o reconhecimento do câncer de pele do tipo melanoma.

O Melanoma Maligno, como informado por Pereira, 2012, tem como alguns fatores de risco: exposição em excesso ao sol, pele clara e histórico de câncer de pele na própria pessoa ou membro da família.

Segundo Ganzeli *et al.*, 2011, o primeiro sinal de melanoma é uma alteração no tamanho, forma ou coloração de uma mancha existente. A maioria apresenta uma área escura ou azulada, podendo também surgir como um nevo novo, escuro, anormal ou de

“aspecto repugnante”.

Notamos que, uma das áreas de aplicação do reconhecimento de padrões é a análise de imagens, diagnóstico médico e visão computacional [Castro e Prado 2002]. Neste trabalho o reconhecimento de padrões possui grande participação na classificação das imagens analisadas. Os algoritmos de classificação utilizados foram: Naive Bayes, Distância Mínima ao Centróide (DMC) e k-Nearest Neighbor (KNN).

Portanto, o presente trabalho, a partir do reconhecimento de imagens utilizando algoritmos de reconhecimento de padrões e técnicas de processamento digital de imagens, tem como objetivo diferenciar imagens que possuem câncer de pele do tipo melanoma de outras que apresentem manchas de pele normais.

1.1. Melanoma

Melanoma é o mais perigoso tipo de câncer de pele e também é a principal causa de morte por doenças de pele. As taxas estão aumentando progressivamente e embora o risco aumente com a idade, ela agora é vista frequentemente em pessoas jovens. [Montemayor-Quellenberg 2012]

Nos Estados Unidos, como informado por [Montemayor-Quellenberg 2012], o melanoma representa o quinto tipo mais comum de novos diagnósticos de câncer em homens e o sétimo tipo mais comum em mulheres.

Segundo [INCA s.d], o melanoma tem predominância em adultos brancos. Embora o câncer de pele seja o mais frequente no Brasil e corresponda a 25% de todos os tumores malignos registrados no país, o melanoma representa apenas 4% das neoplasias malignas do órgão, apesar de ser o mais grave devido à sua alta possibilidade de metástase.

Como informado por [INCA s.d], a cirurgia é o tratamento mais indicado para o tratamento da doença, contudo, quando ela já está espalhada em outros órgãos (metástase), ela é incurável na maioria dos casos. Assim, quando a doença se encontra em estágio avançado tem-se como objetivo aliviar os sintomas e melhorar a qualidade de vida do paciente .

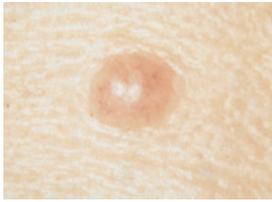
1.1.1. Regra ABCD

A aplicação clínica da regra ABCD pode melhorar a detecção de lesões suspeitas [Friedman, Rigel, e Kopf 1985]. A regra ABCD pode ser verificada da seguinte maneira [Ganzeli *et al.* 2011]:

- Assimetria (A): A forma de uma metade não é igual à outra;
- Bordas (B): As bordas da lesão são frequentemente imperfeitas, chanfradas, pouco nítidas ou irregulares no seu contorno, e o pigmento pode estender-se para a região circunvizinha;
- Coloração (C): A cor não é uniforme. Tonalidades escuras, marrons e bronzeadas podem estar presentes. Áreas brancas, cinzas, vermelhas, róseas ou azuis também podem ser encontradas;
- Diâmetro (D): Existe uma modificação no tamanho, geralmente um aumento. Os melanomas são geralmente maiores que 5 mm .

Na Tabela 1 mostra-se a diferença entre uma lesão de melanoma maligno e benigno.

Tabela 1. Regra ABCD. (Jerant *et al.* 2000)

Regra ABCD	Benigno	Maligno
A = Assimetria suspeita a melanoma: O tumor é dividido ao meio, e as metades não são semelhantes		
B = Borda irregular suspeita de melanoma: Bordas desiguais ou irregular.		
C = Variação de cor suspeita de melanoma: existe mais de uma cor de pigmento		
D = Diâmetro suspeito a melanoma: Caso o diâmetro seja maior que 6 mm		

Como o sistema desenvolvido não tem como meta realizar o acompanhamento da dimensão do câncer, a regra D do processo ABCD foi eliminada.

1.2. Reconhecimento de padrões

Entende-se por padrões as propriedades que possibilitam o agrupamento de objetos semelhantes dentro de uma determinada classe ou categoria, mediante a interpretação de dados de entrada, que permitam a extração das características relevantes desses objetos, [Tou e González 1981]. Contudo, este agrupamento pode ser muito complexo e até demorado para que um ser humano possa fazer sozinho, logo, o reconhecimento de padrões por computador é uma das mais importantes ferramentas utilizadas no campo da inteligência de máquina. Atualmente está presente em inúmeras áreas do conhecimento e encontra aplicações diretas em visão computacional [Nogueira *et al.* 2006].

Entretanto, para se fazer o reconhecimento é necessário a utilização de uma base de dados sobre o problema que se deseja classificar, no caso, o câncer de pele do tipo melanoma. Outro fator a citar é a utilização de algum algoritmo de reconhecimento que faça o uso da base de dados e possa determinar se a mancha é um melanoma ou não. Cabe mencionar que a imagem que será analisada passa antes por uma etapa de processamento de imagem para a extração de suas características. A partir das características extraídas serão utilizados os algoritmos de detecção.

Existem vários algoritmos para efetuar esta tarefa, mas neste trabalho foram utilizados o *DMC*, *KNN* e *Naive bayes*. Abaixo segue uma breve descrição sobre o funcionamento de cada um.

1.2.1. KNN

O KNN é um classificador que faz previsões de acordo com os rótulos dos K vizinhos mais próximos da instância de teste [Yan e Xu 2010]. Esta técnica pode ser muito efetiva, mas requer que seja armazenado o conjunto inteiro de treinamento. Portanto, o KNN pode usar uma grande quantidade de memória e torna-se muito lento. Muitas vezes agrupam-se o conjunto de treinamento para reduzir seu tamanho antes de usar este método [Bradski e Kaehler 2008]

1.2.2. NAIVE BAYES

O Naive Bayes é um classificador que faz previsões de acordo com o teorema de Bayes [Yan e Xu 2010]. É um dos mais eficientes e eficazes algoritmos indutivos de aprendizagem de máquina e mineração de dados [Zhang 2004].

1.2.3. DMC

Para cada classe é assumido um centro de massa (também conhecido como centróide). Um objeto pertence a essa classe quando a distância entre ele e o centróide for menor que todas as distâncias entre os outros centróides restantes do espaço de características. O primeiro passo do processo de classificação por distancia mínima é o calculo dos vetores médios (centróides) que representam cada classe por padrões [Scaranti, Bernardi e Plotze 2010].

2. Método proposto

O presente trabalho possui a estrutura demonstrada na Figura 1, onde cada módulo tem como finalidade processar a imagem de entrada e repassar seu resultado para o próximo. Nas seções a seguir será explicado com detalhes a função de cada módulo.

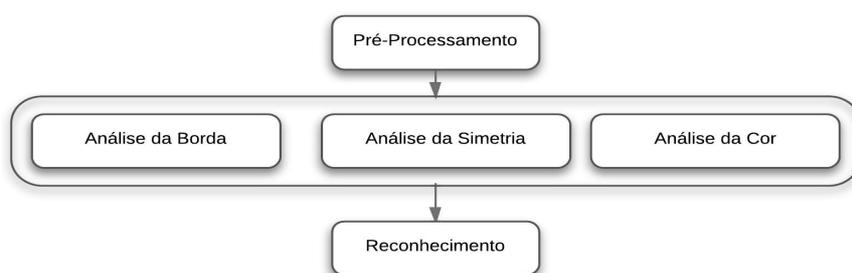


Figura 1: Estrutura do trabalho proposto: pré-processamento, análise dos atributos da imagem e reconhecimento da imagem.

2.1. Pré processamento

Esta etapa tem por finalidade a extração da região de interesse (mancha) da imagem, removendo qualquer outros elementos que possam estar presentes nela, como pêlos, pele e até mesmo alguns possíveis ruídos provenientes do ambiente de aquisição da imagem. Portanto a imagem submetida ao processo passa por uma série de etapas, como demonstra a Figura 2, até que seu objetivo seja alcançado.



Figura 2: Diagrama do funcionamento da etapa de Pré-Processamento.

Da Figura 2, temos:

1. Entrada de uma imagem colorida (RGB) ao pré-processamento;
2. Redimensionamento da imagem para um tamanho de 300 pixels de largura por 300 pixels de altura, para a redução de processamento das etapas seguintes;
3. Suavização da imagem utilizando filtro da mediana 3x3 [Filho e Neto 1999], tendo como objetivo reduzir a quantidade de ruídos na imagem, como exemplo, pêlos;
4. Binarização da imagem, utilizando faixa de limiar constante. [Carnelos e Zeliotto 2011];
5. Aplicação de um filtro morfológico, [Ren 2010], de abertura para redução de alguns ruídos que aparecem após a etapa de binarização;
6. Utilização da imagem da *Etapa3* como máscara com a imagem da *Etapa 5* [Bradski e Kaehler 2008], resultando em uma imagem com apenas o sinal em destaque, como mostra a Figura 3 (e).



Figura 3 : Exemplo do processo de pré-processamento: (a) Imagem submetida ao processo, já redimensionada; (b) Aplicação do filtro da mediana; (c) Binarização da imagem; (d) Uso do filtro morfológico, operador de abertura; (e) Resultado da aplicação da máscara (d) com a imagem (b).

2.2. Análise da variação de cor

Neste momento temos como objetivo quantificar a variação de cor que a região de interesse possui. Esse procedimento representa a regra C do método de detecção de melanoma. A Figura 4 mostra os passos que foram tomados para desenvolver essa etapa.

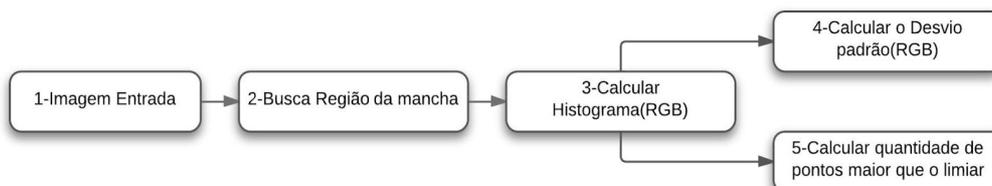


Figura 4 : Descrição da etapa de análise da variação de cor.

Para realizarmos a análise da variação de cor da imagem, foram efetuados os seguintes passos:

1. Encontrar a região de interesse (Figura 5 (b));
2. Calcular o histograma de cada componente (RGB) da imagem colorida (Figura 6 e Figura 7);
3. Efetuar o cálculo do desvio padrão, para cada histograma. Assim, temos uma métrica da variação de cor da região de interesse;
4. Processar a quantidade de pontos maiores do que um determinado limiar pré-estabelecido. Neste momento conseguimos outra métrica para a análise da variação de cor.

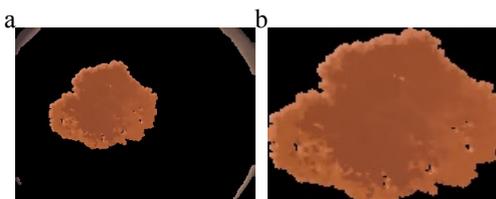


Figura 5 : Seleção da região onde se encontrada a região de interesse. (a) Imagem com a mascara aplicada. (b) Região onde ser encontra a mancha.

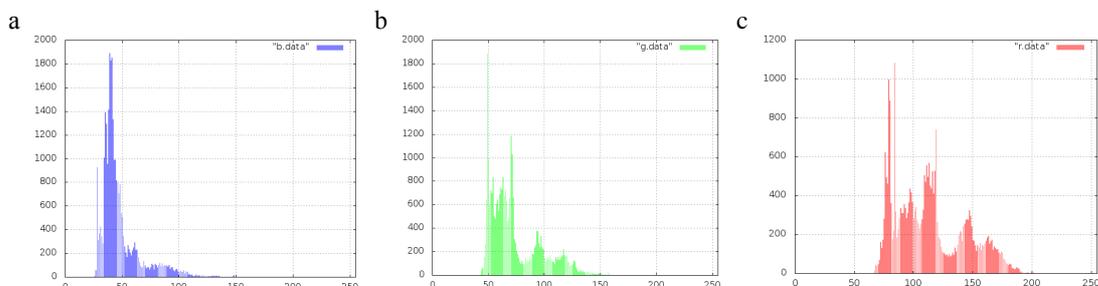


Figura 6 : Histogramas de uma imagem que possui melanoma. (a) Histograma do canal B (azul), desvio padrão = 27 e média = 72. (b) Histograma do canal G (verde), desvio padrão = 30 e média = 95. (c) Histograma do canal R (vermelho), desvio padrão = 27 e média = 128.

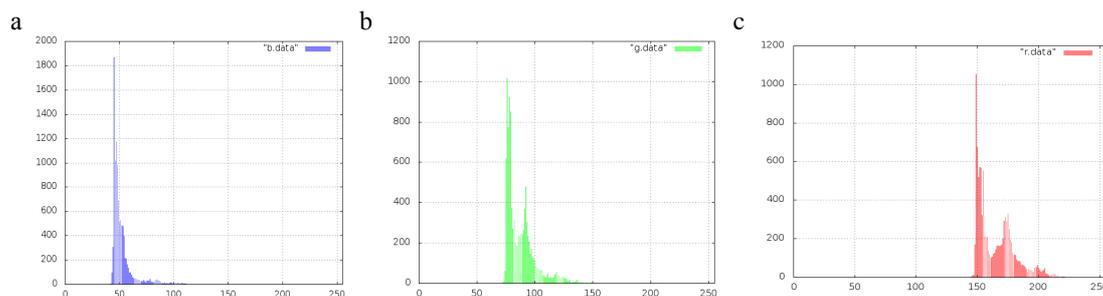


Figura 7: Histogramas de uma imagem que não possui melanoma. (a) Histograma do canal B (azul), desvio padrão = 19 e média = 180. (b) Histograma do canal G (verde), desvio padrão = 17 e média = 103. (c) Histograma do canal R (vermelho), desvio padrão = 15 e média = 68.

2.3. Análise da Variação de Borda

Essa etapa tem o objetivo de quantificar as irregularidades da borda da imagem da mancha, conforme apresentada a Figura 8.

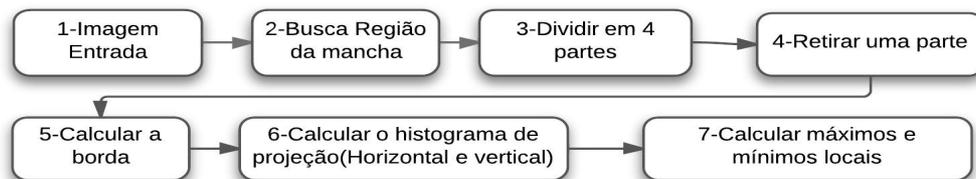


Figura 8: Etapas para a detecção de borda da imagem.

A análise da variação da borda foi efetuada da seguinte maneira:

1. Detecção da região de interesse, como descrito na seção 2.2 (Análise da variação de cor: Etapa 1);
2. Divisão da imagem em em 4 partes, com base no centro da região de interesse;
3. Recorte da primeira parte da imagem para fazer o processamento (Figura 9 (b));
4. Cálculo da borda, (Figura 9 (d)) utilizando a operação de convolução, com a máscara visando destacar os pixels brilhantes circundados por pixels mais escuros; [Filho e Neto 1999]
5. Cálculo do histograma de projeção horizontal e vertical; [Azevedo e Conci 2007]
6. Para cada histograma de projeção (horizontal e vertical), calculou-se a quantidade de máximos e mínimos locais contidos em seu vetor (Figura 10 e Figura 11).

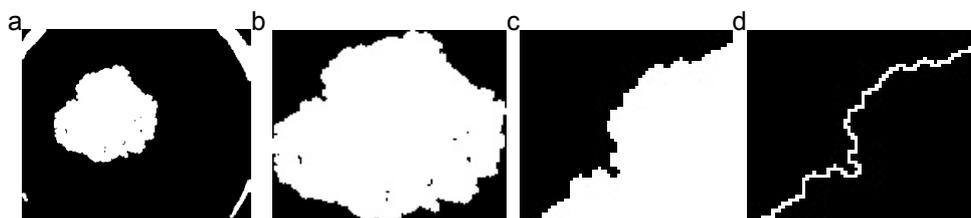


Figura 9: Etapas do processo de Análise da Borda. (a) Imagem binarizada, submetida ao processo. (b) Imagem com região encontrada. (c) Seleção de um quadrante de imagem. (d) Bordas do quadrante selecionado.

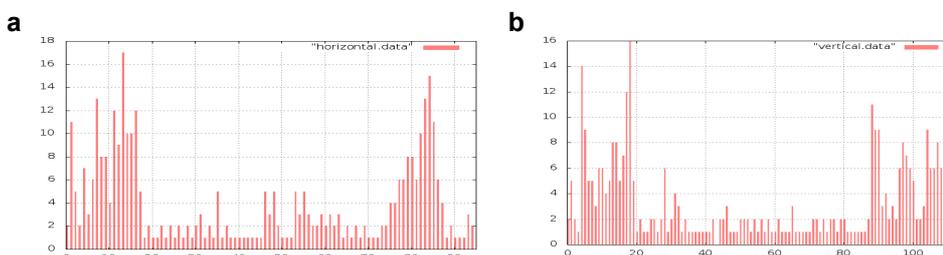


Figura 10: Histograma de projeção de uma imagem com melanoma. (a) horizontal. (b) vertical.

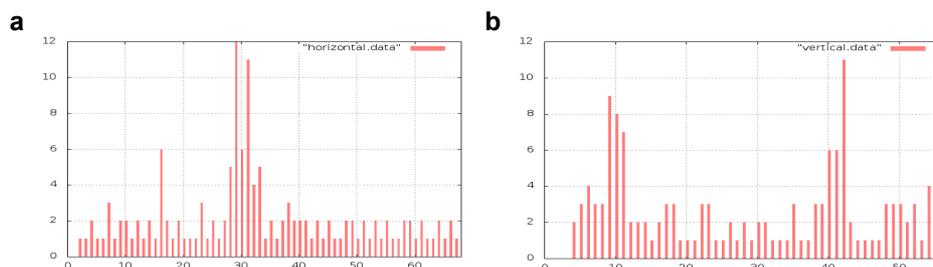


Figura 11: Histograma de projeção de uma imagem sem melanoma. (a) horizontal. (b) vertical.

2.4. Análise da simetria

Essa última etapa tem como objetivo comparar duas metades da imagem da região de interesse, e com isso verificar a diferença entre ambas. A Figura 12 mostra os passos tomados para efetuar esta etapa. Abaixo temos a descrição dos passos.

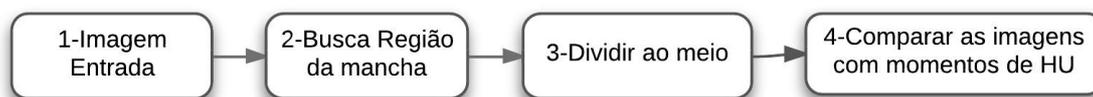


Figura 12: Etapas efetuadas na análise da simetria.

1. Detecção da região de interesse, como descrito na seção 2.2 (Análise da variação de cor: Etapa 1);
2. Divisão da imagem ao meio (Figura 13 (c) e Figura 13 (d)), selecionando o ponto médio da largura da imagem;
3. Utilizando *Momentos Hu* [Bradski e Kaehler 2008] [Corso, Almeida e Junior 2007], comparou-se as duas metades.



Figura 13. Etapas do processo de Análise da Simetria. (a) Imagem binarizada, submetida ao processo; (b) Imagem com região encontrada; (c) Metade direita da imagem (b); (d) Metade esquerda da imagem (b).

3. Base de dados e ambiente utilizado

Para este trabalho, a base de dados de 29 imagens de manchas que não correspondem à melanoma e 12 classificadas como melanoma, foram retiradas de [Silva s.d] e [Rosado 2009].

A fase de treinamento utilizou-se de 70% do total da base e 30% utilizado para testes.

O desenvolvimento da fase de extração de características foi feito utilizando a biblioteca OpenCV [Bradski e Kaehler 2008] e a parte de reconhecimento de padrões foi feito utilizando a linguagem de programação Java.

4. Resultados e discussão

Como demonstra a Tabela 2, nos três algoritmos de reconhecimento de padrões aplicados a detecção de melanoma, obtivemos taxa de acerto acima de 70%, demonstrando assim que a metodologia utilizada para a extração de características das imagens com base na regra ABCD e criação da base de dados foram bastante satisfatórias, destacando-se novamente que a regra D não foi avaliada por motivos anteriormente justificados.

Tabela 2 . Resultados obtidos pelos algoritmos de reconhecimento de padrões.

Algoritmo	Média de Acerto	Desvio Padrão
DMC	0.7103	0.035391033504621985
KNN	0.8181	$1.65 \times 10^{(-14)}$
Naive Bayes	0.9090	$6.88 \times 10^{(-15)}$

O algoritmo que obteve melhor taxa de acerto foi o Naive Bayes com média de acerto de 90% e desvio padrão $6.88 \times 10^{(-15)}$, sendo a menor taxa a do DMC com 71%.

5. Trabalhos futuros

Ampliar a base de dados, pois será possível melhorar a qualificação e a eficácia do método de reconhecimento, fazendo com que os algoritmos avaliem mais informações antes de fazer o reconhecimento da doença.

No que concerne ao reconhecimento, referente à regra D, pretende-se adicionar a informação do processamento da imagem, para assim, disponibilizar mais um forma de prever o quanto uma mancha é indicativa de ser melanoma.

Referências

- Azevedo, E. e Conci, A. (2007) Computação Gráfica, V.2 - Teoria e Prática. Rio de Janeiro, Editora Campus.
- Bradski, G. R. e Kaehler, A. (2008), A Learning OpenCV: Computer Vision with the OpenCV Library, O'REILLY, 1th edition.
- Carnelos, P. H. A. e Zeliotto, R. (2011) "Sistema de detecção de movimento em tempo real com kit DSP TMS320DM642". Universidade Federal do Paraná, Departamento de engenharia elétrica, Curitiba .
- Castro, A. A. M.; Prado P. P.L. (2002) Algoritmos Para Reconhecimento de Padrões. In *Revista Ciências. Exatas*, páginas 129-145, Publicação Impressa.
- Corso, D.A., Almeida, R.H.P. de e Junior, A.S.B. (2007) "Extração de características baseadas em forma para o reconhecimento de padrões em um sistema de visão computacional", In : Encontro de Engenharia e Tecnologia em Computação, <http://goo.gl/O4OHI>, Março.
- Filho, O.M. e Neto, H.V. (1999), Processamento Digital de Imagens, Brasport Livros e Multimídia.

- Friedman R. J., Rigel D. S., Kopf A.W. (1985) Early detection of malignant melanoma: the role of physician examination and self-examination of the skin. *CA Cancer J Clin.*
- Ganzeli, H. S., Bottesini, J. G., Paz L. O. and Ribeiro, M. F. S. (2011) “SKAN: Skin Scanner –System for Skin Cancer Detection Using Adaptive Techniques”, *IEEE Latin America Transactions*, vol. 9, no. 2, April 2011, pp:206-212
- Gonzalez, R.C. e Woods, R.E. (2010). *Processamento Digital de Imagens*, Pearson, 3th edição.
- INCA, Instituto Nacional do Câncer. (s.d) “Câncer de Pele : melanoma”, [:http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/pele_melanoma](http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/pele_melanoma), Fevereiro de 2013.
- Jerant, A. F., Johnson J. T, Demastes-Sheridan C. et al. (2000) Early detection and treatment of skin cancer. *Am Fam Physician* 62: 357-368.
- Montemayor-Quellenberg, M. (2012) “Skin cancer detection breakthrough: Missing pieces of DNA structure a red flag for deadly skin cancer”, <http://news.harvard.edu/gazette/story/2012/09/skin-cancer-detection-breakthrough/>, Março.
- Nogueira, A., Azevedo, J., Baptista, V., Siqueira, S. (2006) “ Um Overview Sobre Reconhecimento de Padrões”, In: II Simpósio de Excelência em Gestão e Tecnologia, Rio de Janeiro.
- Pereira, G.A. (2012) “Melanoma” [.http://cancerdepele.net.br/melanoma](http://cancerdepele.net.br/melanoma), Março de 2013.
- Ren, T. I. (2010). *Processamento de Imagem Morfológica*. Centro de Informática, Universidade Federal de Pernambuco. Recife.
- Rosado, L.F.C.M.G. (2009). *Sistema Automático para Diagnóstico de Lesões Cutâneas Baseado em Imagens Dermoscópicas*. Engenharia Biomédica, Universidade técnica de Lisboa. Lisboa
- Scaranti, A. ; Bernardi, R. ; Plotze, R. O.. (2010) “Identificação de Órgãos Foliares Utilizando as Wavelets de Daubechies”. In: WVC'2010 - VI Workshop de Visão Computacional.
- Silva, S.F. (s.d), “Dermatology Atlas”, <http://www.atlasdermatologico.com.br>, Dezembro de 2012.
- Sobieranski, A. C., Comunello, E., Coser, L. and Wangenheim, A. V. (2007). *Metodologia Computacional para Aplicação da Regra ABCD na Avaliação de Lesões Pigmentadas*. In: WIM 2007 - VII Workshop de Informática Médica, Porto de Galinhas, PE, Brasil.
- Tou, J. T., Gonzalez, R. C. (1981) *Pattern Recognition Principles*. Company, Massachusetts, . Addison-Wesley Publishing
- Yan, Z., Xu, C. (2010) “Combining KNN Algorithm and Other Classifiers”, In *Proceedings of 9th IEEE International Conference on Cognitive Informatics (ICCI'10)*, Beijing, China.
- Zhang, H. (2004) “The Optimality of Naïve Bayes”, In: *Proceedings of the Seventeenth*

International Florida Artificial Intelligence Research Society Conference. AAAI Press, Miami Beach.