

EA006 - Trabalho de Fim de Curso (TFC)

Aluno:

Tiago de Lima Ferreira - RA 093120

tiago.de.lima.ferreira@gmail.com

+55 19 9 9533 1171

Orientador:

Prof. Dr. Eduardo A. do Valle Jr.

dovalle@dca.fee.unicamp.br

+55 19 3521-3769

Classificação de imagens para diagnóstico de melanoma

Introdução:

O melanoma é um tipo de câncer de pele, sendo o mais agressivo da categoria [1]. Sua incidência tem aumentado nos últimos anos [2], porém a chance de cura é bastante elevada se o mesmo é detectado cedo [3], tornando o processo de triagem crítico.

Triagem, na medicina, é a estratégia de detectar a doença ainda que a mesma não apresente sinais óbvios ou sintomas, proporcionando uma intervenção mais rápida e uma melhor esperança de reduzir a mortalidade [4].

Um dos exames feitos na triagem é através da imagem dermatoscópica, que, de forma não-invasiva, pode avaliar pintas e prevenir alguns tipos de câncer de pele, diferenciando as pintas benignas das lesões de risco [5]. Como é um método feito com imagens, ele está sujeito às dificuldades de interpretação humana, consequentemente, o diagnóstico assistido por computador, automatizado ou semi automatizado, tem atraído a atenção de cientistas e da indústria, sendo alvo de várias pesquisas [6][7] e sendo valioso para triar possíveis pacientes da doença, principalmente em locais onde a presença constante do dermatologista não é possível.

Neste trabalho de fim de curso, foram avaliados alguns modelos de aprendizado de máquina usados no diagnóstico do melanoma presentes na literatura, bem como seus resultados. Foi selecionado o método utilizado pelo SkinScan® [10] para ser implementada com base na descrição do mesmo, procurando reproduzir ao máximo seus resultados.

Objetivos:

O trabalho foi dividido em quatro etapas: revisão bibliográfica, seleção do método, implementação e comparação de resultados.

O objetivo da revisão é verificar os métodos existentes para diagnóstico de melanoma e seus resultados, visando possuir dados para determinar qual método possibilitaria melhores comparações com o método que está sendo desenvolvido pela equipe do RECOD (*Reasoning for Complex Data*) da Unicamp [9].

A seleção do método foi feita usando os critérios de facilidade de reprodução, ano de publicação e resultados. O objetivo era selecionar um *paper* recente, cuja

descrição fosse detalhada o suficiente para permitir a reprodução, e com bons resultados, de preferência testados em uma base de dados que já possuíamos.

A implementação tem como objetivo reproduzir o método utilizado no *paper* da forma mais fiel possível, sendo que a comparação dos resultados obtidos com os do *paper* é o indicador chave do sucesso deste trabalho. É preciso ressaltar, entretanto, que a reprodução de dados publicados em Ciência da Computação é um persistente desafio, devido à complexidade das técnicas, que nunca são descritas em todos os detalhes nos *papers*: muitas otimizações de parâmetros *ad hoc* tem forte impacto nos resultados obtidos, mas acabam varridas para fora dos artigos (às vezes, por insistência de revisores e editores!) sob o argumento de que se tratam de "detalhes de implementação". Soma-se a isso o fato de a maioria dos autores não disponibilizar nem código nem dados utilizados nos experimentos.

Metodologia:

Revisão bibliográfica e seleção do método:

Foram selecionados quatro *papers* [10][11][12][13] que apresentam métodos de aprendizado de máquina para o diagnóstico do melanoma. A tabela 1 contém um resumo da comparação entre os métodos utilizados.

Tabela 1: Quadro resumo dos métodos disponíveis na literatura

Trabalho	Método (resumo)	Resultados	Base utilizada	Facilidade de reprodução
Wadhawan(2011b)	Implementação do 7-point-checklist[14], método bastante utilizado para detectar melanoma	Sensibilidade de 87,27% e Especificidade de 71,31%	347 images selecionadas, consideradas por especialistas como de fácil diagnóstico e que tiveram um bom resultado com os algoritmos de segmentação automatizados.	Média-baixa - não são especificadas as imagens do banco, os meta parâmetros do classificador e da extração de características e também a forma de ativação de cada um dos 7-point que determinam o

				diagnóstico
Wadhawan(2011a)	Biblioteca C/C++ para uso em dispositivos móveis. Segmentação baseado em 3 métodos, seleção de características a partir de transformadas de wavelets e classificação usando SVM	AUC de 91,1%, sensibilidade de 80,76% e especificidade 85,57%	1300 imagens	Média - o método de segmentação utilizado possui implementação disponível na internet e há uma boa descrição do procedimento utilizado, porém algumas partes ficam pouco claras, como o tratamento que a imagem sofre pré e pós segmentação e na escolha dos parâmetros do SVM
Iyatomi(2008)	Modelo cliente servidor para classificação das imagens, usuário envia imagem para o servidor, segmenta a imagem, calcula as características, realiza a classificação usando redes neurais e retorna o diagnóstico.	Sensibilidade de 85.9% e especificidade de 86.0%	1258 imagens dermatoscópicas	Baixa - complexidade dos algoritmos utilizados e clareza do <i>paper</i>
Marques(2012)	Segmentação, uso das	Sensibilidade de 94.1%,	163 imagens dermatoscópicas	Baixa - não identifica o

	características de cor e textura, classificação binária. Analisa a importância da cor e textura para a classificação	especificidade de 77.4%		classificador utilizado e método de segmentação não especificado, apesar de mencionar os métodos de morfologia matemática utilizados.
--	--	-------------------------	--	---

Da tabela 1 é possível verificar que todos os papers selecionados são recentes e possuem bons resultados nas respectivas bases. A facilidade de reprodução foi determinada a partir da leitura dos *papers*, verificando o quanto o mesmo fornecia informações dos métodos utilizados de forma a tornarem possível sua reprodução.

É possível perceber que o Wadhawan(2011a) - chamado SkinScan® - e Wadhawan(2011b) são duas partes de um mesmo trabalho que busca implementar uma aplicação para dispositivos móveis para facilitar a triagem do melanoma. A parte “a” foca mais em avaliar o desempenho computacional da aplicação em um dispositivo móvel (no caso um iPhone 4), enquanto que a parte “b” é focada na explicação de um fluxo de classificação mais complexo.

A primeira parte descreve apenas a extração de características pelo método de Haar wavelet, que extrai informações de textura da imagem, porém não leva em consideração a cor da lesão.

Por ser um processo mais simples e ter uma descrição que facilita a reprodução (apesar de vários pontos serem omitidos pelo autor), foi selecionado o método da parte “a” para este trabalho. Entretanto o *paper* não dá base para identificar se o teste da aplicação foi realizado considerando apenas a característica obtida do Haar, logo a comparação de resultados levará em conta esse fato.

O método do SkinScan®:

O objetivo do *paper* é desenvolver uma biblioteca C/C++ que realize as operações de segmentação, extração de *features* e classificação de forma rápida o bastante para ser implementado em dispositivos móveis, que possuem memória e poder de processamento limitado.

Primeiramente, foi desenvolvida uma aplicação desktop em Matlab® [15] para classificação de imagens dermatoscópicas, porém com o advento dos smartphones, desenvolver uma aplicação portátil para esses aparelhos se tornou interessante tanto do ponto de vista da aplicação, já que ela poderá ser utilizada por mais pessoas de forma mais abrangente, quanto do ponto de vista médico, dada a importância da triagem na detecção do melanoma.

Este trabalho busca reproduzir da forma mais fiel possível a aplicação descrita para os dispositivos móveis, utilizando Matlab®, sendo que as decisões de implementação tomadas ao longo do projeto, principalmente referentes a pontos pouco claros do *paper*, são documentadas ao longo deste relatório.

Base de imagens:

Neste trabalho utilizamos uma base de imagens dermatoscópicas para realizar os testes. A mesma [16] possui 560 imagens classificadas como benignas e 187 como melanoma, todas com tamanho 512x512 pixels.

Tratamento e segmentação da Imagem:

A primeira etapa realizada é um tratamento na imagem. Todo o banco é convertido para escala de cinza e é aplicada uma filtragem mediana rápida bidimensional [17], para reduzir ruído e suprimir características físicas que podem afetar a etapa de segmentação. A seguir, é feita a segmentação propriamente dita, que consiste em criar uma máscara binária que separa a lesão do restante da imagem. Isso é importante, pois as imagens contêm uma grande área de pele saudável, logo é importante extrair a lesão para executar as análises posteriores.

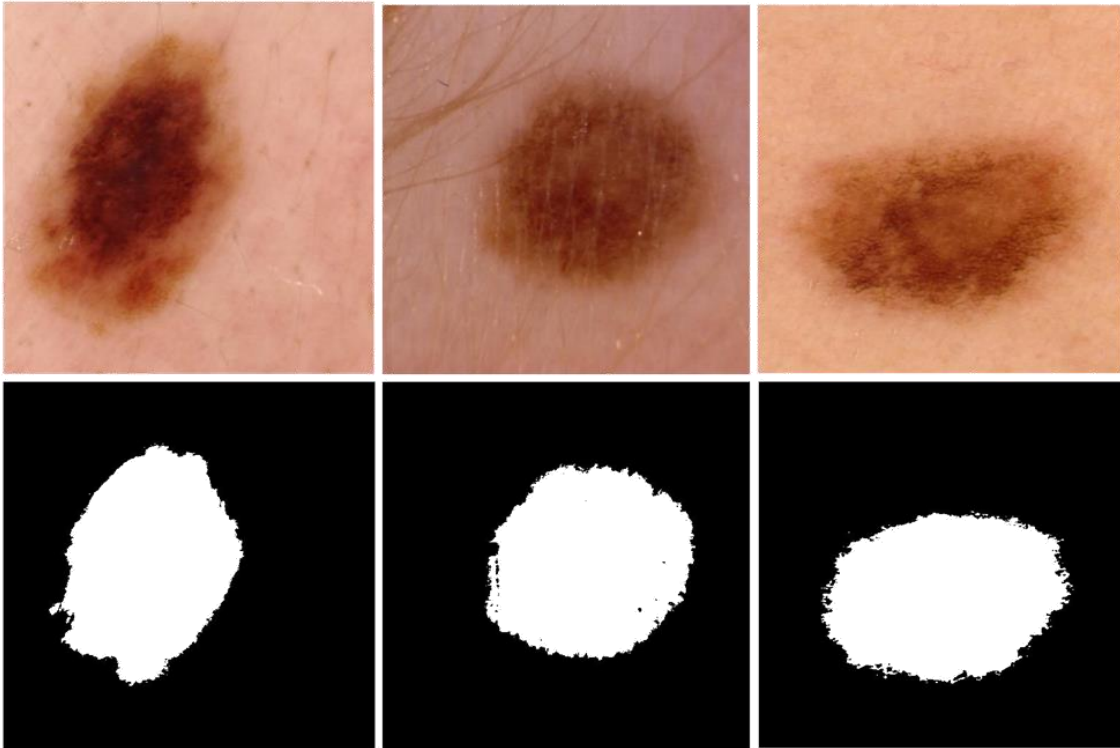


Figura 1: lesões benignas e suas respectivas máscaras de segmentação dadas pelo *Active Contours Without Edges*

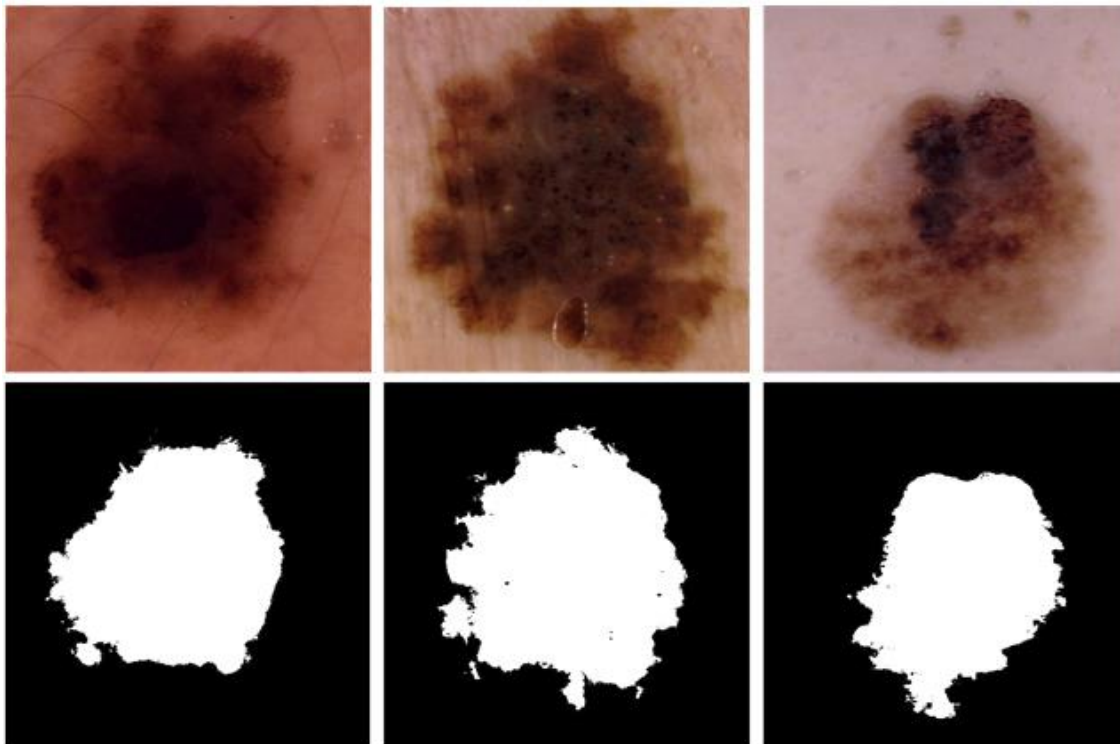


Figura 2: lesões malignas e suas respectivas máscaras de segmentação dadas pelo *Active Countours Without Edges*

No *paper* são utilizados três métodos de segmentação, ISODATA [18], *fuzzy c-means* [19] e *Active Contours Without Edges* [20]. O último obteve os melhores resultados na segmentação, portanto foi o escolhido para ser utilizado neste trabalho.

O *Active Contours Without Edges* possui uma implementação em Matlab® fornecida pelos autores do artigo [21]. Nessa implementação, o autor redimensiona a imagem a ser segmentado, o que pode acarretar na perda de informações relevantes para a classificação posterior da mesma, logo essa parte do código foi modificada para remover essa restrição. Isso acarreta num maior tempo de processamento por imagem.

Após a segmentação, a máscara passa por operações morfológicas, como abertura e fechamento. O autor não especifica quais operações realizou ou qual o elemento estruturante foi usado, nem em qual ordem as operações ocorrem, portanto foi utilizada para esse trabalho a função de fechamento (dilatação seguida de erosão) com elemento estruturante de uma matriz 3x3.

Extração das características (features):

Para a extração das características, foi utilizada a máscara binária criada no processo anterior para restringir a extração apenas à área da imagem que contém a lesão. Aplicada a máscara, um conjunto de pontos da lesão foi selecionado usando um grid ortogonal de tamanho M , então um conjunto de patches de tamanho $K \times K$ foi selecionado da imagem, sendo que os mesmos são centralizados nos pontos selecionados anteriormente do grid.

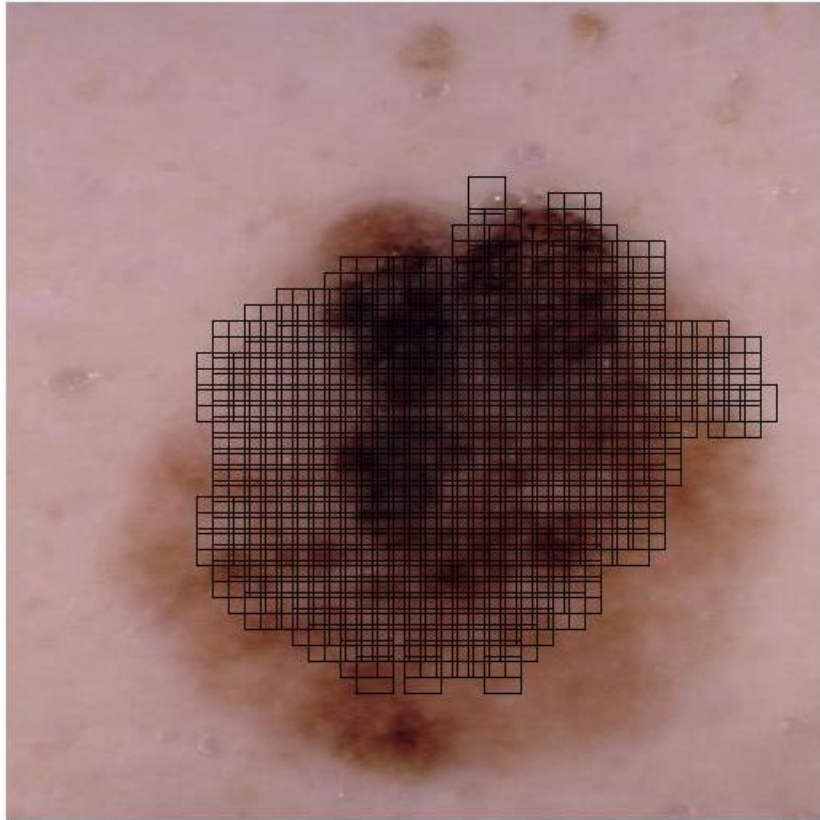


Figura 3: Patches sobre a imagem de uma lesão

O valor de M é uma troca entre velocidade computacional e desempenho da classificação: quanto maior o M , menor a acurácia, porém mais rápida será a extração das características. O valor de K sofre de um dilema parecido, valores grandes de K levam a um tempo de execução maior do algoritmo, enquanto valores pequenos resultam em características ruidosas, pouco relevantes para classificação. No *paper*, foram escolhidos os valores $M=10$ e $K=24$, portanto estes foram os valores utilizados nesse trabalho.

Feita a seleção dos patches, cada um deles passa por uma transformada de Haar de três níveis [22], que é um caso particular de uma transformada discreta de wavelet, obtendo assim 10 imagens de sub-bandas. De acordo com [15] o uso de coeficientes de wavelets é efetivo na detecção do melanoma.

Para cada sub-banda são então calculadas medidas estatísticas, que extraem a informação de textura da imagem, como média e desvio padrão. Esses valores são colocados de forma a formar um vetor de características, logo cada imagem passa a possuir uma matriz de $F \times 20$ características, sendo que F é o número de vetores e também o número de patches que ela possui (o que depende do tamanho original da imagem) e 20 correspondem às 10 médias e 10 desvios padrões calculados para cada uma das 10 sub-bandas geradas de cada patch.

Classificação das imagens:

Antes de aplicar a classificação em si, os vetores de características são normalizados para média zero e desvio padrão unitário, ou seja, é aplicado o Z-Score, calculado pela fórmula abaixo:

$$Z = \frac{x - \mu}{\sigma}$$

onde:

Z é o Z-score;

x é o valor a ser normalizado;

μ é a média da população;

σ é o desvio padrão da população;

Com os vetores normalizados, todos os vetores de características de todas as imagens de treinamento são clusterizados utilizando um k-means com L=200 clusters (número utilizado no paper do SkinScan®), resultando em 200 centroides.

Feita a clusterização, é construído para cada imagem um histograma, com 200 cestos em que, para cada vetor de características da imagem é incrementado o cesto j, em que j é obtido pela equação de minimização:

$$\min_j \|C_j - F_i\|$$

onde:

C_j é o centroide j;

F_i é o vetor de características i;

Com os histogramas prontos, foi feito o treinamento do classificador, no caso um SVM. Não é mencionado no *paper* os parâmetros de configuração do SVM, nem o *kernel* utilizado, portanto foi utilizado nesse trabalho um *kernel* Gaussiano em que os parâmetros C e gama foram obtidos para cada classificação a partir de um algoritmo de busca que varreu um conjunto de valores para C e gama e analisou a melhor combinação para o treino do SVM.

Teste:

Para o teste foi separado o conjunto de imagens em cinco *folds*, sendo que, a cada teste, um *fold* era reservado para teste e os outros utilizados para treino. O *fold* de treino passava pelas mesmas etapas de segmentação e extração de características, porém seus histogramas foram submetidos ao classificador SVM já treinado pelos outros quatro *folds* restantes.

Resultados:

Analisar os resultados da classificação com base na acurácia pode ser enganoso para a base de dados utilizada, que é altamente desbalanceada.

Tabela 2: Matriz de confusão

Resultado	Doença	
	Presente	Ausente
Positivo	Verdadeiro Positivo	Falso Positivo
Negativo	Falso Negativo	Verdadeiro Negativo

Para saber quão preciso é o teste então, duas medidas foram definidas: a sensibilidade e a especificidade. A sensibilidade mede a capacidade do teste em identificar corretamente a doença entre aqueles que a possuem, ou seja, o quão sensível é o teste. A especificidade mede a capacidade do teste em excluir corretamente aqueles que não possuem a doença, ou seja, o quão específico o teste é. Em outras palavras, alta sensibilidade implica em poucos falsos negativos; e alta especificidade, em poucos falsos positivos. Essas medidas são usadas em conjunto, pois há um compromisso entre as duas: uma técnica pode, sem grande esforço, aumentar sua sensibilidade simplesmente diminuindo sua especificidade, e vice-versa. O objetivo é ter ao mesmo tempo alta sensibilidade e especificidade.

A sensibilidade é a fração dos que obtiveram resposta positiva no teste entre aqueles que possuem a doença, ou seja:

$$\text{Sensibilidade} = \frac{VP}{VP + FN}$$

A especificidade é a fração dos que obtiveram resposta negativa no teste entre aqueles que não possuem a doença, ou seja:

$$Especificidade = \frac{VN}{VN + FP}$$

Valores altos de sensibilidade e especificidade indicam um bom teste, pois ele identificará corretamente aqueles que têm a doença e aqueles que não têm. Essas duas medidas permitem que seja construída a matriz de confusão, como mostrado na tabela 2.

Além desses critérios, outro bastante comum na verificação de desempenho de classificadores binários é a *Receiver Operating Characteristic* (ROC), que plota a taxa de verdadeiros positivos com a taxa de falsos positivos. Quanto melhor o classificador, maior é a AUC (*Area Under the Curve*) ou a integral da ROC.

Para este trabalho, foram calculadas a especificidade, sensibilidade e AUC do classificador, que são apresentados abaixo.

Base IRMA:

- **Teste 1:**
 - gama = 5, C = 0.25
 - Acurácia = 80.132450%
 - Sensibilidade = 69.230769%
 - Especificidade = 83.928571%
 - VP = 27, VN = 94, FP = 18, FN = 12
 - AUC = 0.7745
- **Teste 2:**
 - gama = 8, C = 0.25
 - Acurácia = 74.496644%
 - Sensibilidade = 70.270270%
 - Especificidade = 75.892857%
 - VP = 26, VN = 85, FP = 27, FN = 11
 - AUC=0.7907
- **Teste 3:**
 - gama = 8, C = 0.50
 - Acurácia = 75.838926%
 - Sensibilidade = 62.162162%
 - Especificidade = 80.357143%
 - VP = 23, VN = 90, FP = 22, FN = 14
 - AUC = 0.7822
- **Teste 4:**

- gama = 2, C = 0.5
- Acurácia = 77.852349%
- Sensibilidade = 67.567568%
- Especificidade = 81.250000%
- VP = 25, VN = 91, FP = 21, FN = 12
- AUC = 0.7732

- **Teste 5:**

gama = 8, c = 0.5

- Acurácia = 79.194631%
- Sensibilidade = 70.270270%
- Especificidade = 82.142857%
- VP = 26, VN = 92, FP = 20, FN = 11
- AUC = 0.7956

Média das AUCs = 0.7832

Sensibilidade média: 67.89%

Especificidade média: 80.71%

Discussão:

Os resultados obtidos diferem dos apresentados no *paper* do SkinScan®. Várias razões podem ser levantadas para explicar a diferença, como a base de imagens ser diferente, a não utilização de outras características a não ser textura na classificação, a falta de informação de quais operações morfológicas foram aplicadas e de quais meta parâmetros foram utilizados para o SVM.

Apesar dos bons resultados do SkinScan®, a comunidade médica ainda tem receio de utilizar softwares de auxílio ao diagnóstico médico, como pode ser visto, por exemplo, no artigo “*Skin scan: A demonstration of the need for FDA regulation of medical apps on iPhone*” [23] em que os autores testaram um conjunto de 93 imagens de melanoma (diagnosticadas por biópsia) e o resultado foi de que apenas 10 imagens foram apontadas pelo software como sendo de “alto risco”, o que levou os autores do artigo a questionarem sua eficácia e também a pedir ao FDA (*Food and Drug Administration* – órgão do governo americano que regula medicamentos e equipamentos médicos) uma regulação do uso de apps de natureza médica.

Dentro do contexto deste trabalho, porém o resultado é considerado satisfatório e possui uma ampla margem para melhorias futuras como melhorias na parte de segmentação e a inclusão de outros tipos de características na classificação, como a cor, e a utilização do 7-point-checklist, algo que pode aumentar drasticamente a desempenho do classificador.

Conclusão:

Os resultados deste trabalho demonstram a importância da característica de textura para classificação de imagens para o diagnóstico de melanoma. Sendo um software de auxílio médico, o mesmo precisa de constante evolução, integrando outras características que podem ser valiosas para a classificação e também obtendo feedback da comunidade médica de seu uso e oportunidades de melhoria.

Certamente o diagnóstico feito por software não eliminará a necessidade da validação do médico especialista, porém é um importante passo para facilitar e agilizar a tomada de decisão desses profissionais, buscando ser uma primeira ferramenta principalmente em locais em que o acesso ao exame dermatoscópico ou ao próprio dermatologista não é suficiente.

Bibliografia:

[1] -

<http://www.cancer.gov/cancertopics/pdq/treatment/melanoma/Patient/page1/AllPages> "General Information About Melanoma". *NCI*. 2014-04-17. Acessado em 06 de Dezembro, 2014.

[2]- W. Stolz, O.B. Falco, P. Bliet, M. Kandthaler, W.H.C. Burgdorf, A.B. Cagnetta Color atlas of dermatoscopy, 2nd enlarged and completely revised edition Blackwell publishing, Berlin (2002)

[3]- A. Jerant, J. Johnson, C. Sheridan, and T. Caffrey, "Early detection and treatment of skin cancer," *American Family Physician*, vol. 62, no. 2, pp. 357-368, 2000.

[4]- F. Nachbar, W. Stolz, T. Merkle, A. Cagnetta, T. Vogt, M. Landthaler, P. Bilek, O. Braun-Falco, and G. Plewig, "The ABCD rule of dermatoscopy: High prospective value in the diagnosis of doubtful melanocytic skin lesions," *Journal of the American Academy of Dermatology*, vol. 30, no. 4, pp. 551-559, 1994.

[5] - F. Nachbar, W. Stolz, T. Merkle, A. Cagnetta, T. Vogt, M. Landthaler, P. Bilek, O. Braun-Falco, and G. Plewig, "The ABCD rule of dermatoscopy: High prospective value in the diagnosis of doubtful melanocytic skin lesions," *Journal of the American Academy of Dermatology*, vol. 30, no. 4, pp. 551-559, 1994.

[6] - T. Wadhawan, N. Situ, H. Rui, K. Lancaster, X. Yuan, and G. Zouridakis, "Implementation of the 7-point checklist for melanoma detection on smart

handheld devices,” in International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2011, pp. 3180-3183.

[7]- H. Iyatomi, H. Oka, M. Celebi, M. Hashimoto, M. Hagiwara, M. Tanaka, and K. Ogawa, “An improved internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm,” *Computerized Medical Imaging and Graphics*, vol. 32, no. 7, pp. 566-579, 2008.

[9] - M. Fornaciali, S. Avila, M. Carvalho, and E. Valle, “Statistical Learning Approach for Robust Melanoma Screening” (aceito), in *27th Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2014.

[10] - Wadhawan, T., Situ, N., Rui, H., Lancaster, K., Yuan, X., and Zouridakis, G. (2011b). Implementation of the 7-point checklist for melanoma detection on smart handheld devices. In *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3180-3183.

[11] - Wadhawan, T., Situ, N., Lancaster, K., Yuan, X., and Zouridakis, G. (2011a). Skinscan : A portable library for melanoma detection on handheld devices. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 133-136.

[12] - Iyatomi, H., Oka, H., Celebi, M., Hashimoto, M., Hagiwara, M., Tanaka, M., and Ogawa, K. (2008). An improved internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm. *Computerized Medical Imaging and Graphics*, 32(7):566-579.

[13] - Marques, J., Barata, C., and Mendonca, T. (2012). On the role of texture and color in the classification of dermoscopy images. In *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4402-4405.

[14] - Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of the ABCD rule of dermoscopy and a new 7-point checklist based on pattern analysis. Argenziano G, Fabbrocini G, Carli P, De Giorgi V, Sammarco E, Delfino M *Arch Dermatol*. 1998 Dec; 134(12):1563-70.

[15] - Malignant Melanoma Detection by Bag-of-Features Classification - Ning Situ, Xiaojing Yuan, IEEE Member, Ji Chen, IEEE Member, and George Zouridakis, IEEE Senior Member

[16] - IRMA datasets - <http://ganymed.imib.rwth-aachen.de/irma/datasets>

[17] - Huang T, Yang G, Tang G. A fast two-dimensional median filtering algorithm. Acoustics, Speech and Signal Processing, IEEE Transactions on. 1979 Feb;27:13-18.

[18] - Ridler TW, Calvard S. Picture thresholding using an iterative selection method. IEEE transactions on Systems, Man and Cybernetics. 1978;SMC-8:630-632.

[19] - Dunn JC. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. Journal of Cybernetics.1973;3:32-57.

[20] - Chan TF, Vese LA, Active contours without edges. IEEE Trans Image Process. 2001; 10(2):266-7.

[22] - Chan Vese Active Contours without edges, by Yue Wu, 26 March 2009, disponível em <http://www.mathworks.com/matlabcentral/fileexchange/23445-chan-ve-se-active-contours-without-edges>

[22] - Stollnitz EJ, DeRose TD, Salesin DH. Wavelets for computer

[23] - Skin scan: A demonstration of the need for FDA regulation of medical apps on iPhone Natalie Anne Ferrero, BS, Dean S. Morrell, MD, Craig Nathaniel Burkhart, MD, Published Online: March 03, 2014

Resumo:

Neste trabalho é realizada a reprodução do SkinScan®, uma das aplicações disponíveis na literatura para auxílio no diagnóstico de melanoma. A implementação foi feita em Matlab®, e são utilizadas técnicas de tratamento de imagens e aprendizado de máquina. Foi utilizado um banco de dados de 560 imagens, 187 delas diagnosticadas como melanoma. Os resultados obtidos foram de 67.89% de sensibilidade, 80.71% de especificidade e 78.32% de AUC.