

Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging

Dani Silva

2013155412, Mestrado Integrado em Engenharia Biomédica

Filipa Costa

2013148125, Mestrado Integrado em Engenharia Biomédica

Abstract

Melanomas malignos, a maneira mais mortal do cancro da pele, tem um bom prognóstico se for tratado logo num estado inicial ainda curável. O diagnóstico precoce e a remoção do melanoma por cirurgia são os tratamentos mais eficazes do melanoma.

Este artigo apresenta as etapas computacionais do diagnóstico automatizado do melanoma. Estas etapas computacionais são: 1.) pré-processamento e segmentação de imagens para determinar as áreas fulcrais, 2.) extração e seleção de features para quantificar as propriedades, e 3.) classificar as áreas da lesão como malignas ou não. Neste artigo, detalhamos estas etapas computacionais e tentamos contornar os diversos desafios que o *data set* vai apresentando.

A melhor classificação final proporciona uma sensibilidade de 74% com uma especificidade de 75%.

Os resultados da classificação foram obtidos num conjunto de 764 imagens dermatoscópicas que pertencem ao arquivo do ISIC (*International Skin Imaging Collaboration*) fornecidas no Concurso *Skin Lesion Analysis Towards Melanoma Detections*.

1. Introdução

Melanoma é o tipo mais grave de cancro de pele, desenvolve-se nas células (melanócitos) que produzem melanina - o pigmento que dá cor à pele. A causa exata de todos os melanomas não é clara, mas a exposição à radiação ultravioleta (UV) da luz solar ou lâmpadas de bronzamento aumenta o risco de desenvolver melanoma.

O risco de melanoma parece vir a aumentar a cada ano, sendo responsável por mais de 9.000 mortes por ano apenas nos Estados Unidos, e mais de 10.000 mortes estimadas em 2016 [1]. Conhecer os sinais de alerta deste tipo de cancro da pele pode ajudar a garantir que as alterações cancerígenas são detetadas e tratadas antes que o cancro se desenvolva.

Como o melanoma ocorre na superfície da pele, é passível de deteção por um simples exame visual. Na verdade, a maioria dos melanomas são reconhecidos pela primeira vez pelos pacientes, não pelos médicos[2]. No entanto, a inspeção visual sem au-

xílio de dermatologistas especializados está associada a uma precisão diagnóstica de cerca de 60%, o que significa que muitos melanomas curáveis não são detetados até estágios mais avançados. Para melhorar o desempenho do diagnóstico e reduzir as mortes por melanoma, foi introduzida a dermoscopia, que é uma técnica de imagem que elimina a reflexão da superfície da pele, permitindo que camadas mais profundas possam ser visualizadas com maior pormenor. Assumindo níveis adequados de perícia pelo intérprete, a imagem dermatoscópica mostrou melhorar o desempenho de reconhecimento de melanomas, resultando numa melhoria da precisão absoluta para valores de 75% -84% [3]. No entanto, tem sido demonstrado que a dermoscopia pode realmente diminuir a precisão diagnóstica nas mãos de dermatologistas inexperientes [4]. Além disso, para o diagnóstico, os dermatologistas contam com sua experiência clínica e percepção visual. No entanto, o diagnóstico feito pela visão humana é um tanto subjetivo, faltando precisão e reprodu-

tibilidade.

Os sistemas de análise de imagens por dermatoscopia computadorizada não têm a limitação dessa subjetividade. Estes sistemas permitem a utilização de um computador como um segundo método de diagnóstico independente, que pode ser potencialmente utilizado para a pré-seleção de doentes, realizados por operadores não experientes e para a assistência dos clínicos. Embora as técnicas de análise computadorizada não possam fornecer um diagnóstico definitivo, estas podem ser usadas para melhorar a tomada de decisão de biópsia da mancha, o que alguns observadores acham que é o uso mais importante para a dermatoscopia [5]. Finalmente, a análise automatizada pode servir como uma ferramenta adicional para melhorar o acompanhamento.

À medida que o desenvolvimento em aplicações com base em imagens dermatoscópicas para smart phones estão a chegar ao mercado [6], aumenta também a investigação em algoritmos de avaliação dermatoscópica para influenciar positivamente o cuidado do paciente. Como resultado, o interesse da comunidade tem crescido, como muitos centros de investigação a reunirem os seus esforços de pesquisa em análise automatizada.

Neste trabalho, é apresentada uma abordagem metodológica para a classificação de imagens de dermatoscopia, com base no banco de imagens da ISIC. O resto do artigo está organizado da seguinte forma: Seção 2 descreve as várias técnicas e metodologias utilizadas até à atualidade. Seção 3 descreve o conjunto de dados utilizado na análise. A seção 4 explica o pré-processamento, a segmentação efetuada e os resultados obtidos pelas mesmas. A Seção 5 discute a extração de features. A seção 6 descreve a seleção de features. A Seção 7 apresenta as várias famílias de classificação utilizadas e os resultados obtidos. A Seção 8 apresenta as discussões consequentes dos resultados obtidos. Finalmente a Seção 9 apresenta a conclusão e os trabalhos futuros.

2. Estado de Arte

A abordagem padrão na análise automática de imagens dermatoscópicas tem normalmente três fases: A) Segmentação da imagem; B) Extração de features; C) Seleção de features e D) Classificação.

2.1. Segmentação de imagens

A segmentação é a fase mais importante para uma análise correta de imagens, uma vez que vai

influenciar a precisão das etapas subsequentes. No entanto, a fase de segmentação não é uma tarefa simples, devido à grande variedade de lesões, tipos de pele, presença de pelos [7]. Para resolver este problema, vários algoritmos têm sido propostos, podendo ser classificados como thresholding (o mais utilizado), edge-based ou region-based.

Em 1994, Green et al. [8] o algoritmo utilizado para a segmentação de imagens foi uma adaptação do thresholding para um tridimensional espaço de cor, obtendo 83,8% de segmentação correta (171 aproveitáveis imagens em 204 imagens de lesões).

Em 1998, Hance et al. [9] comparou 6 diferentes algoritmos de segmentação da cor (adaptive thresholding, fuzzy-c-means, SCT/center split [10], PCT/median, split and merge, and multiresolution segmentation). Numa gama de 66 imagens, os algoritmos que alcançaram uma melhor performance foram o adaptive thresholding (40 em 66) e o PCT/median (46 boas segmentações).

Em 2001, Ganster et al. [11] usaram 3 algoritmos diferentes para segmentar as imagens de melanomas, denominados por global thresholding, dynamic thresholding, e a 3-D color clustering concept. De um conjunto de 4000 imagens de lesões de pele, apenas 159 foram rejeitadas devido a erros de segmentação, nomeadamente quando a máscara não conseguia cobrir a lesão toda ou quando a diferença da cor entre a lesão e a pele saudável era mínima. Portanto, 96% das imagens das lesões foram corretamente segmentadas.

Em 2005, Erkol et al. [12] verificaram a qualidade do GVF-based method para a segmentação de 100 imagens de lesões da pele, usando outros 2 métodos para comparação - color histogram analysis technique e a segmentação manual por um segundo conceituado dermatologista. Os resultados obtidos foram baseados numa comparação com uma segmentação manual das lesões determinada por um primeiro dermatologista. A média dos erros obtidos pelo método GVF eram muito menores quer para as imagens benignas quer para as de melanomas (13.77% para benigno e 19.76% para melanoma) comparativamente aos erros obtidos pelo 2º método (19.87% e 91.96%). Ainda assim eram maiores do que os erros obtidos pelo segundo dermatologista (8,71% e 8,13%). Um dos maiores problemas do GVF aplicado a imagens dermatoscópicas são as bordas "fracas" em algumas das imagens.

Em 2009, Silveira et al. [13] avaliaram 6 métodos para a segmentação de lesões da pele em 100 imagens dermatoscópicas: Adaptive threshol-

ding (AT), Gradient Vector Flow (GVF), Adaptive Snake (AS), Level Set Method of Chan et al. (C-LS), Expectation-maximization level set (EM-LS) e fuzz-based split-and-merge algorithm (FBSM). Os resultados obtidos foram comparados com as imagens de “referência” (ground truth) e os erros de segmentação foram avaliados por 4 métodos diferentes de avaliação. O TDR com considerado dos 4 o mais relevante método de avaliação do ponto de vista da clínica, sendo que os melhores métodos de segmentação obtidos foram AS e EM-LS (que são semi-supervisionados, com TDR=95,47% e TDR=95,20% respectivamente) e o FBSM (que é um método completamente automático, com TDR=93,67%).

2.2. Extração de features

Existem inúmeros estudos e métodos que permitem extrair características relevantes e importantes de um determinado conjunto de funcionalidades, diminuindo a complexidade computacional dos algoritmos.

- **Forma:** é uma importante característica quando se trata de manchas na pele. A área da lesão deve ser obtida pelo método de quadriláteros de bits uma vez que, garante melhores resultados para objetos com bordas irregulares, o que não acontece por contagem do número de pixels dentro da borda [14]. É relevante também obter a proporção entre o comprimento do eixo maior e o comprimento do eixo menor; assimetria da lesão e a compactação da lesão [14] para desta forma adquirir resultados mais viáveis;
- **Cores:** Para quantificar as cores que a lesão apresenta pode-se utilizar a média e o desvio padrão, assimetrias severas na cor e histograma de distâncias [14];
- **Textura:** dependendo da técnica usada podem-se obter diversas características à cerca da textura da lesão. Um dos métodos que permite obter dezenas de dados estatísticos descritivos da textura é com base na matriz de ocorrência dos níveis de cinza (GLCM) [14].

Barata et al. [15] foi um pouco mais longe e analisou a importância da textura comparativamente às cores na classificação de imagens dermatológicas, concluindo pelos resultados obtidos que apenas os

dados fornecidos pela cor da lesão são mais assertivos do que apenas dados sobre a textura ou mesmo ambos.

2.3. Seleção de features

A seleção das características é um importante passo de pré-processamento em muitas tarefas de machine learning. O objetivo é reduzir o número de características consideradas para o estudo, eliminando features redundantes, irrelevantes ou ruidosos. Do ponto de vista de classificação, existem inúmeros benefícios associados com a seleção de características: (i) tempo de armazenamento, (ii) a redução da complexidade do classificador, (iii) aumento da precisão, (iv) redução dos tempos de teste, e (v) uma melhor compreensão e visualização de dados.

Os métodos de filtragem dependem de características gerais dos dados para selecionar um subconjunto de recursos sem envolver qualquer algoritmo de aprendizagem.

Na literatura proposta para lesões de pele existem muitos métodos para fazer a seleção de features, de seguida apresentamos alguns com bom desempenho em conjuntos de diversos dados:

- **ReliefF:** Neste algoritmo, um número de amostras é selecionado aleatoriamente a partir do conjunto de dados inicial e são determinados os vizinhos mais próximos, podendo ter em conta o ruído e dados multi-classes [14].
- **Mutual information based feature selection (MIFS):** avalia a informação igual entre as características individuais e os nomes das classes, e seleciona as features que tenham o máximo de informação em comum [14].
- **Correlation based feature selection (CFS):** Este algoritmo tenta encontrar um conjunto de recursos que individualmente se correlacionam bem com a classe [14].

O SVM pode ser utilizado para selecionar features pois apresenta uma ampla gama de aplicações e é um método simples que se provou ser robusto e preciso quando comparado com outras abordagens [16].

2.4. Classificadores

A classificação é o processo de extração de informação em imagens para reconhecer padrões e objetos homogêneos sendo que os estudos relacionados com a classificação automática de imagens de

lesões pigmentadas na pele têm aparecido na literatura desde 1987 [17]. Com base em imagens dermatoscópicas e após os processos de segmentação e extração de features podem ser aplicados diferentes tipos de classificadores:

1. Support vector machines (SVMs): é um dos classificadores que desperta mais atenção na comunidade de Machine Learning, devido à sua sólida formação teórica e excelente desempenho prático. SVMs têm várias vantagens sobre os classificadores mais clássicos, como árvores de decisão e redes neurais [14]. Como forma de melhorar o desempenho deste classificador houve recente uma pesquisa mais intensiva relativa ao desequilíbrio de classes na comunidade de Machine Learning. Esse problema geralmente ocorre quando uma ou mais classes se sobrepõem a outras. Em tais casos, a maioria dos classificadores concentra-se na análise das grandes classes, e em consequência diminui a precisão da classificação das classes mais pequenas. Em muitos domínios, tais como diagnósticos médicos as consequências de uma classificação incorreta, tal como classificar as amostras minoritárias (melanoma) como amostras benignas, são bastante graves. De acordo com Celebi et al. [14] para evitar erros provocados por classes desequilibradas deve-se analisar a área sob a curva ROC (gráfico de True Positive versus false positive). Com base na pesquisa de Celebi et al. [14] a abordagem efetuada utilizando SVM obteve ótimos resultados (especificidade de 92.34% e sensibilidade de 93.33%). Sendo que é necessário realçar que no número de features usadas na implementação do algoritmo (k) é um parâmetro importante que deve ser considerado de modo a obter-se o melhor desempenho possível do classificador.
2. KNN: Pela análise da investigação realizada por Barata [15], verifica-se que utilizando o classificador KNN obtêm-se bons resultados de especificidade e sensibilidade, fazendo variar o número de vizinhos (k) e a distância (SE = 100% e SP = 72%). De forma a garantir uma boa generalização do método comparou-se o KNN ao algoritmo SVM verificando-se que este apresenta uma performance inferior (SE = 92% e SP = 72%).
3. AdBoost: tem a capacidade de selecionar os recursos apropriados para um determinado problema e elimina features que não melhoram a

performance do classificador. Para melhorar o desempenho deste tipo de classificador é necessário também ter em conta o número de bins pois influenciará o número de features a considerar. Assim de acordo com o procedimento efetuado por Barata [15] alcançam-se valores de SE = 92% e SP = 85%.

4. C4.5: O algoritmo C4.5 constrói árvores de decisão a partir de um conjunto de dados de treino, utilizando o conceito de entropia. Em cada nó da árvore, o algoritmo C4.5 escolhe o atributo dos dados que mais particiona o conjunto de amostras em subconjuntos tendendo a uma categoria ou a outra. O critério de particionamento é o ganho de informação normalizado (diferença de entropia). Neste tipo de classificador é necessário definir corretamente o número mínimo de samples per leaf (M) e o fator de confiança [18]. O desempenho do algoritmo C4.5 na investigação de Celebi [18] foi então de 84,33% de sensibilidade e especificidade de 96,19%.
5. Artificial Neural Networks (ANN): são inspiradas pelo sistema nervoso central de um animal que são capazes de realizar uma aprendizagem a partir inputs e seguidamente reconhecer padrões entre eles para tomar uma decisão global, requerem um treino inicial para identificar o conjunto de parâmetros da rede de acordo com os referidos pesos [19]. James et al. [20], tal como outros métodos referidos anteriormente, para testar a validade deste classificador obteve a curva ROC, determinando a probabilidade de uma lesão de pele ser classificada como melanoma corretamente. Os resultados obtidos foram os seguintes: sensibilidade, 94,3% e com um valor semelhante de especificidade 93,8%. Rubegni et al. [21] também usou artificial neural network como classificador e obteve valores de SE = 96% e SP=93%.

3. Dataset

O dataset usado neste projeto consiste num conjunto de 3 grupos de imagens dermoscópicas de lesões na pele (melanomas ou não melanomas) que pertencem ao arquivo do ISIC (International Skin Imaging Collaboration) fornecidas no Concurso Skin Lesion Analysis Towards Melanoma Detection. Essas imagens foram recolhidas dos principais centros clínicos a nível internacional e adquiri-

das a partir de uma variedade de dispositivos dentro de cada centro.

1. *Training data*: conjunto de 900 em formato JPEG. Todas as imagens usam o esquema de nome ISIC_<image>_id.jpg, onde <image>_id é a identificação única de 7 dígitos. Estas imagens apresentam uma resolução espacial variável;
2. *Data*: conjunto de 379 imagens para testar o classificador;
3. *Ground Truth*: conjunto de imagens geradas por um grupo de especialistas em melanomas, utilizadas para comparação das imagens obtidas em cada fase do concurso. Estas imagens estão em forma de máscaras binárias.

4. Pré-processamento e Segmentação

Nesta etapa foi-nos fornecido no concurso um banco de imagens (ground truth) realizado manualmente por especialista e que utilizamos como padrão de comparação.

A etapa da segmentação e do pré-processamento são das mais importantes, uma vez que afeta a precisão dos passos subsequentes. No entanto, a segmentação é difícil por causa da grande variedade de formas de lesão, tamanhos (Fig. 1c) e cores (Fig. 1f), junto com diferentes tipos de pele e texturas. Além disso, algumas lesões têm limites irregulares (Fig. 1b) e, em alguns casos, há uma transição suave entre a lesão e a pele (Fig. 1e). Outras dificuldades estão relacionadas com a presença de objetos irrelevantes (Fig. 1d), presença de pelos sobre as lesões (Fig. 1a) e o mau enquadramento das mesmas na fotografia.

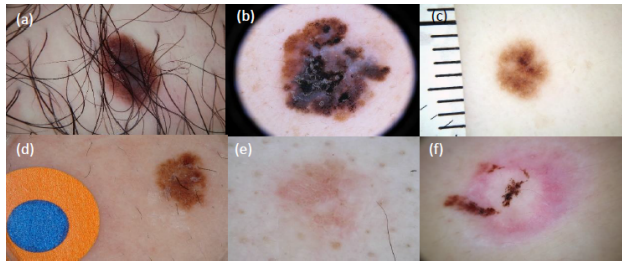


Figura 1: As figuras acima representam as mais diversificadas imagens do nosso *Dataset*. a) pelos, b) bordas pretas, c) régua, d) objetos circulares, e) manchas muito tênues, f) Manchas claras com pequenas manchas escuras.

Todo o tratamento das imagens do dataset nesta fase foi realizado com recurso ao MATLAB®, incluindo diversas funções intrínsecas do mesmo.

O pré-processamento passou por remover as regiões irrelevantes das imagens e ajustes do contraste da imagem para destacar as manchas da pele.

O algoritmo Otsu, foi executado para criar as imagens binárias, sendo a partir destas implementados diversos métodos para remover os pelos e os adesivos circulares que se encontravam em certas imagens sem nenhuma relevância para a segmentação.

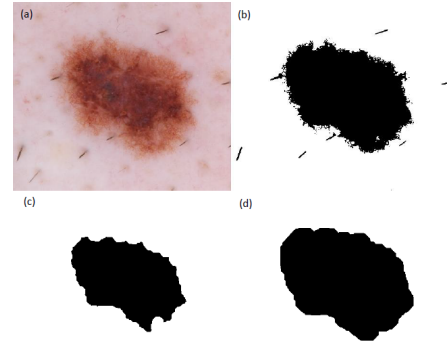


Figura 2: Comparação das imagens com implementação das funções strel, imrode e imdilate. (a) imagem original; (b) imagem tratada e após máscara binária; (c) imagem após aplicada a função imrode; (d) imagem com implementação da função imdilate.

4.1. Resultados da Segmentação

A verificação dos resultados da segmentação foi feita em comparação com o conjunto de dados do Ground Truth. As diferenças entre as imagens foram avaliadas por 4 métricas (pedidas no Concurso): Sensibilidade, Especificidade, Coeficiente de Dice e Jaccard (Tab. 1). A métrica que consideramos mais significativa foi a Jaccard, que tem um valor médio de 0,7206 (Fig. 3).

	Média	Desvio Padrão
Sensibilidade	0,8151	0,1990
Especificidade	0,9532	0,1111
Coeficiente de Dice	0,8165	0,1813
Jaccard	0,7206	0,2050
Accuracy	0,9045	0,1229

Tabela 1: Valores médios obtidos em cada método de avaliação para a segmentação.

No total existem 136 lesões que têm de ser rejeitadas devido às falhas na segmentação no conjunto das 900 imagens. O desempenho alcançado,

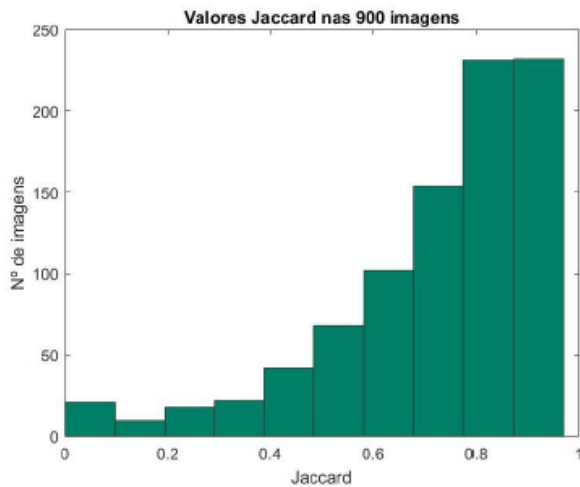


Figura 3: Valores do Jaccard nas 900 imagens

portanto, é de cerca de 84,8% de imagens de lesão cutânea corretamente segmentadas.



Figura 4: Exemplo de uma segmentação não satisfatória. A) imagem original da lesão. B) máscara binária resultante da nossa segmentação. C) máscara binária do Ground Truth Data resultante da segmentação manual por especialistas.

5. Extração de Features

Nesta secção são descritos os recursos que foram utilizados para caracterizar as imagens da lesão cutânea.

5.1. Features de forma

As características globais calculadas cobrem uma ampla gama de categorias, começando pelo tamanho simples (área, perímetro, medidas polares e retângulo delimitador) e de forma (redondeza, compactidade, extensão, solidificação, excentricidade, orientação...). Outros descritores de forma são derivados do retângulo delimitador: relação de aspeto de comprimento e largura, bem como a distância entre o centro do retângulo delimitador e o centro de massa da lesão.

5.2. Features de cor

Para calcular características de cor simples, as imagens originais foram transformadas do mapa de cores RGB (original) para o mapa de cores HSI, HSV e L^*a^*b . As características derivadas para cada mapa são o mínimo, o máximo, a média e a variância dos três canais dos diferentes tipos de sistemas de cor.

5.3. Features de textura

Neste estudo foram ainda extraídas features baseadas na análise de textura usando GLCM. O GLCM é uma poderosa ferramenta para a extração de features de imagem mapeando as probabilidades de co-ocorrência de nível de cinza com base nas relações espaciais de pixels em diferentes direções angulares. A característica extraída com base no GLCM são: autocorrelação, contraste, correlação, proeminência de cluster, cluster shade, dissimilaridade, energia, entropia, homogeneidade, máxima probabilidade, variância, média de soma, variância de soma, entropia de soma, diferença de variação, diferença entropia, medida de informação de correlação, diferença inversa homogênea (INV), diferença inversa normalizada (INN) e o momento de diferença inversa normalizada.

Após extração de todas as features deparamo-nos com alguns valores incoerentes e não válidos para definir a lesão e portanto foi necessário efetuar uma pré-seleção das que não apresentavam um valor possível e credível, procedendo-se à sua eliminação. Portanto no total resultaram 75 features, sendo que 11 definiam a forma da lesão, 43 definiam a cor tendo em conta os diversos colormaps e 21 definiam a textura da imagem segmentada.

6. Seleção de Features

O desempenho do sistema de classificação pode ser degradado usando todas as características dos dados de entrada, pois aumenta a complexidade. Otimizar os recursos é muito importante, já que alguns deles desempenham um papel mais importante no reconhecimento do que outros. Há uma série de métodos que são desenvolvidos e têm sido utilizados para selecionar características.

A seleção de features foi feita utilizando o software de análise de dados *Orange Canvas*. Selecionaram-se features com base nos seguintes

algoritmos: *PCA*, *Information Gain*, *Gain Ratio*, *Gini Decrease*, *Chi2*, *ReliefF* e *FCBF* (Tabela 2).

O procedimento estatístico PCA, utiliza uma transformação ortogonal para converter um conjunto de observações de possíveis variáveis correlacionadas num conjunto de valores de variáveis linearmente não correlacionadas. Este método tem uma natureza completamente diferentes dos restantes algoritmos e, portanto, foi analisado separadamente.

No prosseguimento deste trabalho foram utilizados, como termo de comparação, os resultados obtidos a partir do PCA e das dez features selecionadas pelos seis restantes algoritmos. Neste último utilizaram-se então as seguintes features: a média das cores da lesão verificada com um color map HSI, a homogeneidade, a média da componente verde do colormap RGB, o máximo da componente de intensidade no colormap HSI, a dissimilaridade, a energia, a soma da entropia, o mínimo da componente vermelha do colormap RGB, a média da componente azul do colormap RGB e a variância da componente de intensidade no colormap HSI.

7. Classificação

No geral, extraíram-se 75 features das 764 imagens, depois de remover as imagens que não foram corretamente segmentadas (136). Destas 764 imagens, 618 representam melanomas benignos e 146 melanomas malignos.

Os resultados foram classificados por duas maneiras diferentes. Recorrendo primeiro a classificadores convencionais, *Classification Tree*, *Logistic Regression*, *Naive Bayes*, *Nearest Neighbors*, *Random Forest Classification* e *SVM*, com recurso ao software de análise de dados *Orange Canvas* e, de seguida, a classificadores usando redes neuronais e redes baseadas em Machine Learning recorrendo a Toolboxes do Matlab para o efeito.

No final, para avaliar os vários classificadores, para posterior análise, calcularam-se os valores de especificidade e sensibilidade.

7.1. Resultados da Classificação

Todos os algoritmos referidos foram testados para as features selecionadas pelo método de PCA e as features obtidas a partir da média dos seis métodos. Para termos comparativos classificou-se ainda todas as features obtidas, sem qualquer seleção, desta forma podemos verificar as alterações que a otimização de features provoca.

	Total de Features		Média Rank		PCA	
	Sens.	Esp.	Sens.	Esp.	Sens.	Esp.
K-Nearest Neighbors	0,947	0,089	0,935	0,103	0,924	0,103
Decision Tree	0,854	0,329	0,830	0,253	0,867	0,178
SVM	0,866	0,281	0,626	0,425	0,435	0,486
Logistic Regression	0,977	0,089	0,992	0,034	0,994	0,014
Random Classification	0,945	0,151	0,951	0,068	0,861	0,158
Naive Bayes	0,571	0,712	0,586	0,685	1	0

Tabela 2: Resultados de sensibilidade e especificidade para os diferentes classificadores convencionais e de acordo com as features utilizadas.

Para avaliação dos diferentes classificadores calculou-se o valor da sensibilidade e especificidade. A sensibilidade reflete o quanto o método é eficaz em identificar corretamente, dentro todos os indivíduos avaliados, aqueles que realmente apresentam melanomas benignos. Já a especificidade de um método reflete o quanto ele é eficaz em identificar corretamente os indivíduos que apresentam melanomas malignos.

É bom ter valores altos de ambos os parâmetros mencionados em cima, mas o parâmetro mais relevante neste caso é a especificidade pois é essencial detetar corretamente todos os casos de melanoma malignos.

O classificador *Naives Bayes*, que assume que os atributos são condicionalmente independentes, foi de todos o que obteve melhores resultados, pois obteve o maior valor de especificidade entre todos e obteve um valor razoável de sensibilidade, exceto quando a seleção de features foi exetuada pelo PCA.

Os resultados obtidos pelos os outros dois algoritmos através das Toolboxes do Matlab estão representados em baixo.

	Sens.	Esp.
Image Category Classifier	0,74	0,75
Neural Network	0,91	1

Tabela 3: Resultados de sensibilidade e especificidade para os 2 classificadores recorrendo ao Matlab.

Estes 2 algoritmos apresentam resultados muito melhores comparativamente aos outros tipos de classificadores, sendo que a rede neuronal apresenta de longe os melhores resultados.

8. Discussão

O Data set fornecido para o concurso "*ISBI 2016: Skin Lesion Analysis Towards Melanoma Detection*" que pertencem ao arquivo do ISIC (*Interna-*

tional Skin Imaging Collaboration), composto por 900 imagens dermatoscópicas divididas em 2 classes, melanomas benignos e malignos.

No geral, podemos concluir que realizamos uma segmentação aceitável, pois obtivemos mais de 84% de imagens corretamente segmentadas. A performance da segmentação é conseguida através da fusão de vários algoritmos de segmentação, como por exemplo o *thresholding*. O valor de imagens corretamente segmentadas são é maior devido à grande diversidade de objetos e pelos no nosso *data set* o que torna a pré-processamento e a segmentação uma tarefa muito árdua.

Relativamente à extração de features, o principal foco das features extraídas a partir das imagens já segmentadas, foram os parâmetros de formas e cores.

Através da seleção de features efetuada, o número dos parâmetros mais relevantes para uma boa classificação foi reduzido de 75 para 10 parâmetros. Observou-se que a seleção de features não melhorou assim tanto os resultados, como era esperado acontecer.

Os resultados obtidos para a classificação utilizando classificadores convencionais não são assim tão bons, devido ao facto do *data set* utilizado não ser muito extenso, tal como a fração de imagens de melanomas malignos ser muito reduzida comparativamente ao total de imagens fornecidas (19%) o que faz com que os classificadores fiquem bem treinados para classificar os melanomas benignos.

Os 2 outros classificadores, em que o primeiro utilizou um conjunto de features próprios do método e o outro utilizou a totalidade das 75 features extraídas pelos métodos que realizamos, obtiveram resultados muito melhores, obtendo-se os melhores resultados nas redes neuronais. No entanto, este só utiliza 15% do *data set* para testar o classificador, sendo que por isso não se pode tirar conclusões muito fidedignas. De uma maneira geral e analisando todos os resultados obtidos concluímos que através do *Image Category Classifier* obtêm-se uma classificação mais fidedigna e real, obtendo uma *accuracy* de 75%.

Para finalizar, o nosso classificador é capaz de detetar corretamente uma ampla quantidade de lesões de pele em malignos e benignos. No entanto, há aspetos que devem ser tidos em conta. O nosso *data set* é relativamente pequeno. Para tirar conclusões mais fidedignas deveria-se aumentar o mesmo. Para obter melhores resultados na classificação, novos parâmetros devem ser incluídos nas features. Parâ-

metros que permitam a descrição entre a transição da área entre a lesão e a pele não estão incluídos nas features que obtivemos e testámos, e muito provavelmente é um bom ponto de partida para melhorar os resultados finais.

9. Conclusão

As imagens que compõem a nossa base de dados são bastantes difusas e heterogêneas visualmente e, para a vista normal sem experiência clínica, muito difíceis de distinguir os dois tipos de melanomas. Ao criar-se um método automático, que consiga classificar uma imagem dermatológica como maligno ou benigno, são várias as vantagens que daí advém nomeadamente não estar dependente de uma avaliação de dermatologistas cujo número de avaliações que consegue dar são limitadas ao seu tempo disponível. Considerando que neste trabalho a distinção é entre 2 tipos diferentes de melanoma, o valor de desempenho que conseguimos, 75% através do uso de uma rede baseada em *Machine Learning* é um valor bastante aceitável e encorajador para prosseguir e aperfeiçoar esta técnica em abordagens futuras.

- [1] H. W. Rogers, M. A. Weinstock, S. R. Feldman, B. M. Coldiron, Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the us population, 2012, *JAMA dermatology* 151 (2015) 1081–1086.
- [2] M. S. Brady, S. A. Oliveria, P. J. Christos, M. Berwick, D. G. Coit, J. Katz, A. C. Halpern, Patterns of detection in patients with cutaneous melanoma, *Cancer* 89 (2000) 342–347.
- [3] H. Kittler, H. Pehamberger, K. Wolff, M. Binder, Diagnostic accuracy of dermoscopy, *The lancet oncology* 3 (2002) 159–165.
- [4] M. Binder, M. Schwarz, A. Winkler, A. Steiner, A. Kaidler, K. Wolff, H. Pehamberger, Epiluminescence microscopy: a useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists, *Archives of Dermatology* 131 (1995) 286–291.
- [5] M. Vestergaard, P. Macaskill, P. Holt, S. Menzies, Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting, *British Journal of Dermatology* 159 (2008) 669–676.
- [6] Metaoptima, Molescope™, ???
- [7] T. Mendonca, A. R. Marcal, A. Vieira, J. C. Nascimento, M. Silveira, J. S. Marques, J. Rozeira, Comparison of segmentation methods for automatic diagnosis of dermoscopy images, in: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, pp. 6572–6575.
- [8] A. Green, N. Martin, J. Pfitzner, M. O'Rourke, N. Knight, Computer image analysis in the diagnosis of melanoma, *Journal of the American Academy of Dermatology* 31 (1994) 958–964.
- [9] G. A. Hance, S. E. Umbaugh, R. H. Moss, W. V. Stoecker, Unsupervised color image segmentation: with application to skin tumor borders, *IEEE Engineering in Medicine and Biology Magazine* 15 (1996) 104–111.
- [10] S. E. Umbaugh, R. H. Moss, W. V. Stoecker, G. Hance, Automatic color segmentation algorithms-with application to skin tumor feature identification, *IEEE Engineering in Medicine and Biology Magazine* 12 (1993) 75–82.
- [11] H. Ganster, P. Pinz, R. Rohrer, E. Wildling, M. Binder, H. Kittler, Automated melanoma recognition, *IEEE transactions on medical imaging* 20 (2001) 233–239.
- [12] B. Erkol, R. H. Moss, R. Joe Stanley, W. V. Stoecker, E. Hvatum, Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes, *Skin Research and Technology* 11 (2005) 17–26.
- [13] M. Silveira, J. C. Nascimento, J. S. Marques, A. R. Marcal, T. Mendonça, S. Yamauchi, J. Maeda, J. Rozeira, Comparison of segmentation methods for melanoma diagnosis in dermoscopy images, *IEEE Journal of Selected Topics in Signal Processing* 3 (2009) 35–45.
- [14] M. E. Celebi, H. A. Kingravi, B. Uddin, H. Iyatomi, Y. A. Aslandogan, W. V. Stoecker, R. H. Moss, A methodological approach to the classification of dermoscopy images, *Computerized Medical Imaging and Graphics* 31 (2007) 362–373.
- [15] C. Barata, M. Ruela, M. Francisco, T. Mendonça, J. S. Marques, Two systems for the detection of melanomas in dermoscopy images using texture and color features, *IEEE Systems Journal* 8 (2014) 965–979.
- [16] R. Archibald, G. Fann, Feature selection and classification of hyperspectral images with support vector machines, *IEEE Geoscience and remote sensing letters* 4 (2007) 674–677.
- [17] N. Cascinelli, M. Ferrario, T. Tonelli, E. Leo, A possible new tool for clinical diagnosis of melanoma: the computer, *Journal of the American Academy of Dermatology* 16 (1987) 361–367.
- [18] M. E. Celebi, H. Iyatomi, W. V. Stoecker, R. H. Moss, H. S. Rabinovitz, G. Argenziano, H. P. Soyer, Automatic detection of blue-white veil and related structures in dermoscopy images, *Computerized Medical Imaging and Graphics* 32 (2008) 670–677.
- [19] A. P. James, B. V. Dasarthy, Medical image fusion: A survey of the state of the art, *Information Fusion* 19 (2014) 4–19.
- [20] P. Rubegni, G. Cevenini, M. Burroni, R. Perotti, G. Dell'Eva, P. Sbrano, C. Miracco, P. Luzi, P. Tosi, P. Barbini, et al., Automated diagnosis of pigmented skin lesions, *International Journal of Cancer* 101 (2002) 576–580.
- [21] P. Rubegni, G. Cevenini, M. Burroni, R. Perotti, G. Dell'Eva, P. Sbrano, C. Miracco, P. Luzi, P. Tosi, P. Barbini, et al., Automated diagnosis of pigmented skin lesions, *International Journal of Cancer* 101 (2002) 576–580.