

Detecção de Lesões de Câncer de Pele Utilizando Análise de Componentes Independentes e Análise Discriminante Linear

Luciano Marques Brito Reis, Lúcio Flávio de A. Campos
Universidade Estadual do Maranhão – Cidade Universitária Paulo VI
Caixa Postal 09 – São Luís – MA - Brasil
lucianomarquesreis@hotmail.com, lucioflavio@engcomp.uema.br

Resumo

O câncer de pele vem se configurando como o mais frequente em toda a população brasileira. A exposição excessiva à radiação solar é o fator principal de risco. Diante dessa situação, o método mais simples para a diminuição das taxas de incidência e mortalidade relativas ao câncer de pele é o diagnóstico precoce. Este trabalho tem como objetivo apresentar um método diagnóstico auxiliado por computador baseado em processamento de imagens para detecção de lesões de câncer de pele melanoma e não melanoma utilizando Análise de Componentes Independente (ICA) para extração de características, o algoritmo de máxima Relevância e Mínima Redundância (mRMR) para redução de dimensionalidade e Análise Discriminante Linear (LDA) para a classificação das imagens. A capacidade do método foi avaliada pela técnica de validação cruzada, e atingiu 100% de acurácia, 100% de sensibilidade e 100% de especificidade ao analisar todas as amostras.

Palavras-chave: Análise de Componentes Independentes, Análise Discriminante Linear, Câncer de Pele.

Title: Detection of Skin Cancer Injuries Using Independent Component Analysis and Linear Discriminant Analysis

Abstract

Skin cancer has become the most frequent in the brazilian population. Excessive exposure to solar radiation is the major risk factor. In this situation, the simplest method to reduce incidence and mortality rates related to skin cancer is early diagnosis. This work aims to present a computer-aided diagnostic method based on image processing for the detection of melanoma and non-melanoma skin cancer lesions using Independent Component Analysis (ICA) for feature extraction, the algorithm of maximum Relevance and Minimal Redundancy (mRMR) for reduction of dimensionality and Linear Discriminant Analysis (LDA) for image classification. The ability of the method was evaluated by the cross-validation technique, and reached 100% of accuracy, 100% of sensitivity and 100% of specificity when analyzing all samples.

Keywords: Independent component analysis, Linear Discriminant Analysis, Skin Cancer.

1. Introdução

Câncer é o nome dado a um grupo de doenças malignas caracterizadas pelo crescimento anormal e descontrolado de células que sofreram alteração em seu material genético, em algum momento do seu ciclo celular. Essas células geneticamente modificadas podem invadir os tecidos e órgãos, espalhando-se para outras regiões do corpo [Robbins *et al*, 2001], gerando diversos tipos de câncer. Os diferentes tipos de câncer são classificados em grandes categorias: os carcinomas, as leucemias, os linfomas e mielomas, e os tumores do sistema nervoso central [Prado, 2014].

Neste trabalho será abordado apenas o câncer de pele, que é um câncer do tipo carcinoma, isto é, são tumores que se originam nas células epiteliais ou glandulares (adenocarcinoma) com forte tendência a invadir tecidos vizinhos.

De acordo com Stewart *et al* (2016), em seu estudo realizado em 2012, houve mais de 232.000 novos casos de melanoma e cerca de 55.000 mortes. As taxas mais elevadas são em países com a população caucasiana, predominantemente, com mais de 80% desses novos casos, e cerca de 65% de morte ocorreram na Oceânia, Europa e América do Norte.

Segundo o INCA (2016a), no Brasil, embora o câncer de pele seja o mais frequente e corresponda a 30% de todos os tumores malignos registrados no país, o melanoma representa apenas 3% das neoplasias malignas do órgão, apesar de ser o mais grave devido à sua alta possibilidade de metástase. Enquanto o não melanoma apresenta altos percentuais de cura, caso seja detectado precocemente, e é entre os tumores de pele, o de menor incidência e mais baixa mortalidade.

Apesar da grande incidência do câncer de pele, se diagnosticado precocemente, verifica-se um alto índice de cura. A melhor maneira para se fazer um diagnóstico precoce é através do auto-exame da pele. O auto-exame da pele pode ser feito pelo paciente e é baseado na regra do ABCDE [INCA, 2016b], onde o paciente analisa a assimetria, a borda, a cor, o diâmetro e a evolução das manchas.

O objetivo geral deste trabalho é apresentar uma técnica denominada Diagnóstico Auxiliado por Computador (CAD), para o diagnóstico precoce do câncer de pele, através da Análise de Componentes Independentes (ICA) somada com o algoritmo de máxima Relevância e Mínima Redundância (mRMR) para a extração de características e encontrar o melhor conjunto das características obtidas, que por fim serão analisadas por um classificador, baseado em Análise Discriminante Linear (LDA), para decidir se as imagens de lesões de câncer de pele são do tipo melanoma ou não melanoma. E como objetivos específicos, esta técnica visa auxiliar o especialista, agindo como uma segunda opinião, para que seja aumentada a sensibilidade do diagnóstico, assim como melhorar a acurácia do diagnóstico.

2. Trabalhos Relacionados

A incidência do câncer de pele está ganhando, cada vez mais, proporções epidêmicas. Inúmeros especialistas de diversas áreas vem desenvolvendo técnicas auxiliares para o diagnóstico precoce do câncer de pele com o intuito de diminuir sua taxa de mortalidade.

O CAD é definido como um diagnóstico realizado pelo especialista, utilizando o resultado de análises quantitativas automatizadas de imagens como auxílio para tomada de decisões diagnósticas. A finalidade do diagnóstico auxiliado por computador é melhor a acuidade do diagnóstico, assim como a consistência da interpretação da imagem, mediante o uso da resposta do computador como referência [Azevedo-Marques, 2011], [Seixas e Saader, 2005].

Soares (2008) propôs a análise de classificação de imagens de câncer de pele por atributos de cor, forma e textura. Para a realização da técnica proposta, foi utilizada uma base de dados contendo 112 imagens obtidas através de exames clínicos, dermatoscopia e videodermatoscopia, previamente classificadas como melanoma, não melanoma e benigno por médicos dermatologistas. A extração de características de textura foi feita utilizando a transformada *Wavelet Packet*. Os descritores das imagens foram calculados para cada imagem com 84 canais de energia, 84 variâncias dos canais e ramos de frequência dominantes.

Para extrair os descritores de cor utilizou-se a imagem gerada pela segmentação realizada através do algoritmo *k-means*. Foram calculados de acordo com a média de *pixels* da cor da pele e a média dos *pixels* da cor da lesão. A partir dos descritores regionais e descritores de Fourier obteve-se as características de forma para a representação da lesão por meio de seu contorno independente das invariâncias da escala, rotação, translação e dispersão do sinal.

A classificação fez-se utilizando os dados extraídos como padrões de entrada para a Máquina de Vetor de Suporte (SVM). O conjunto de características foi dividido em dois subconjuntos, um para treinamento e outro para teste. A realização dos testes ocorreu em duas etapas: a primeira, por meio da estratégia tradicional um contra todos e a tomada de decisão por intermédio de uma Máquina de Comitê na forma de mistura de especialista; a segunda, a realização do teste se deu através da estratégia denominada de SVM especialistas, de tal forma que, novos atributos são adicionados às máquinas, em caso de empate, para a determinação da categoria pertencente.

Utilizando a transformada *Wavelet Packet* para a extração dos descritores de textura, para as três classes de imagens de lesões de pele, obteve-se uma taxa de acerto global de 92,73% para melanoma e 85,6% para não melanoma e lesões benignas. E a estratégia denominada de SVM Especialistas se mostrou mais eficiente que a estratégia tradicional, obtendo assim, uma taxa de acerto global de 100% para melanoma e 90% para lesões benignas e não melanoma.

Frutuoso *et al* (2013) propuseram uma técnica de reconhecimento de câncer de pele do tipo melanoma. A base de dados utilizada foi obtida através de [Silva, 2009] e [Rosado, 2009] e continha 41 imagens, sendo 12 classificadas como melanoma e 29 que não correspondem a melanoma. E, como neste sistema, não se tinha a intenção de realizar o acompanhamento da dimensão do câncer, a análise do diâmetro (regra D) foi eliminada do processo.

Na etapa de pré-processamento foram extraídas as Regiões de Interesse das imagens com o objetivo de remover todos os outros elementos que poderiam estar nas imagens. A extração de características foi baseada na regra do ABCD (Assimetria, Borda, Cor e Diâmetro).

A análise da variação de cor (regra C) foi realizada como intuito de quantificar a variação de cor que a região de interesse possui. Para esta etapa, primeiro calculou-se o histograma de cada componente (RBG) da imagem colorida. Logo em seguida, foi calculado o desvio padrão para cada histograma. E por fim, a quantidade de pontos maiores que um determinado limiar pré-estabelecido foi processada.

A análise da variação da borda (regra B), para quantificar as irregularidades das bordas das imagens foi realizada da seguinte forma: após a detecção da região de interesse, a imagem foi dividida em 4 partes e a primeira parte utilizada para fazer o processamento. Utilizando as operações de convolução com as máscaras visando destacar os *pixels* brilhantes circundados dos *pixels* mais escuros, realizou-se o cálculo de borda. Depois, o cálculo do histograma de projeção horizontal e vertical. E para cada histograma de projeção, foi calculado a quantidade de máximos e mínimos locais contidos em seu vetor.

A análise da simetria (regra A), feita para realizar a comparação entre as duas metades das regiões de interesse da imagem, e assim verificar as diferenças entre elas, seguiu os seguintes passos: detecção da região de interesse; divisão da imagem ao meio através do ponto médio da largura da imagem; e utilizando o momento *Hu* (invariantes), comparou-se as duas metades.

Para a etapa de classificação das imagens foram utilizados os classificadores *k-Nearest Neighbor* (KNN), *Naive Bayes* (NB), Distância Mínima ao Centróide (DMC). Para cada classe é assumido um centróide, isto é, um objeto pertencente 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.

Os resultados obtidos para o reconhecimento de lesões de pele do tipo melanoma, utilizando os 3 algoritmos de classificação foram superior a 70%, o classificador KNN obteve uma taxa de acerto de 81,81%, o classificador NB 90,90% e 71,03% para o DMC.

Mahmoud *et al* (2013) apresentaram um sistema para análise automática de melanoma. A base de dados utilizada continha 448 imagens de dermatoscopia classificadas como melanoma maligno e benigno. E foram obtidas através do *Sydney Melanoma Diagnostic Centre* no *Royal Prince Hospital* e através de sites da internet.

A etapa de pré-processamento foi realizada para remover objetos irrelevantes contidos em cada imagem. Os quadros pretos, que geralmente aparecem nas imagens durante o processo de digitalização, foram retirados. E as imagens foram suavizadas através dos filtros *Wiener* e mediana. Logo em seguida, no pós-processamento, com o objetivo de melhorar a forma da imagem, foi utilizado o algoritmo de equalização de histograma, técnica utilizada para realçar o contraste.

A segmentação foi implementada a partir da Região de Interesse Limiar (ROI), método que calcula o valor de intensidade a partir de imagens cinza. E da fusão de Região de Estatística (SRM), baseada na teoria da fusão de região que começa em um ponto de semente que é comparado com seus quatro pontos vizinhos ou *pixels*.

O procedimento de extração de características foi realizado através das transformadas *Wavelet* e *Curvelet*. A análise *Wavelet* utiliza os termos de aproximações e detalhes. As aproximações são os componentes de baixa frequência de alta escala, enquanto os detalhes são os componentes de alta frequência de baixa escala. A transformada *Curvelet* é uma pirâmide não padrão multi-escala com muitas direções e posições em cada escala de comprimento, e em escalas elevadas, a forma de onda *Curvelet* se torna tão fina que se parece com um elemento em forma de agulha. Dessa forma, com o aumento no nível de resolução, a *Curvelet* se torna mais fina e menor no domínio espacial e mostra mais sensibilidade nas bordas das curvas, o que lhe permite captar as curvas numa imagem com uma maior eficácia.

A classificação foi realizada através da Rede Neural de Retro-Propagação (BNN), ou seja, uma generalização da regra de aprendizagem de *Widrow-Hoff*, classe de filtro adaptativo usada para imitar um filtro desejado, para redes de camadas múltiplas e funções de transferências diferenciáveis não lineares.

A precisão do reconhecimento obtida pelo classificador foi de 58,44% para os coeficientes baseados em *Wavelet* e 86,57% para os baseados em *Curvelet*.

3. Método Proposto

A figura 1 mostra o método proposto por este trabalho, onde consistirá em extrair as características das imagens utilizando a Análise de Componentes Independente. Logo em seguida, a redução da dimensionalidade através do algoritmo de máxima Relevância e Mínima Redundância. E por fim, a classificação final será feita utilizando a técnica de Análise Discriminante Linear.

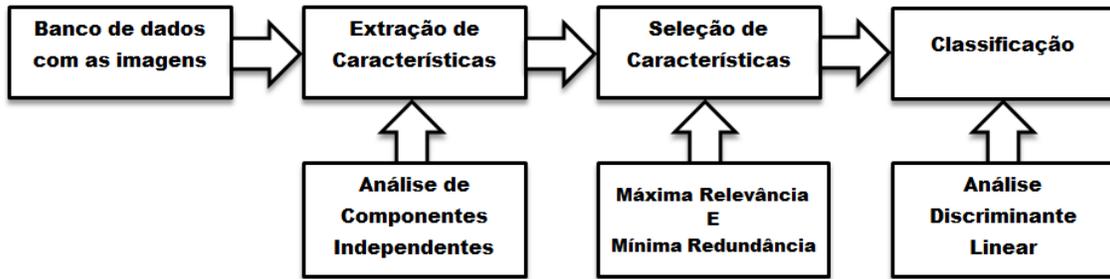


Figura 1: Diagrama de blocos do método proposto

3.1. Aquisição de Dados

As bases de dados utilizadas são bases de domínio público e podem ser adquiridas em [DermIS, 2012] e [DermQuest.com, 2012]. E são usadas para fornecer informações sobre diagnósticos, relatos de casos e informações adicionais sobre inúmeras doenças de pele.

Foram selecionadas 206 imagens de câncer de pele, sendo que através do [DermIS, 2012] adquiriu-se 43 imagens de câncer de pele diagnosticadas como melanoma e 26 como não melanoma. Já em [DermQuest.com, 2012] obteve-se 76 imagens de câncer de pele diagnosticadas como melanoma e 61 como não melanoma. No entanto, só foram utilizadas 204 imagens (duplicidade). A figura 2 apresenta quatro amostras utilizadas na aplicação do método proposto.



(a) DermIS - melanoma



(b) DermIS - não - melanoma



(c) DermQuest - melanoma



(d) DermQuest - não - melanoma

Figura 2: Imagens adquiridas

Como as imagens originais possuíam muitas informações desnecessárias e dimensões diferentes, elas foram contornadas de acordo com a região de interesse e redimensionadas para que ficassem com 25 *pixels* de altura e 25 *pixels* de largura, ou simplesmente, redimensionadas 25 x 25. Conforme a figura 3.

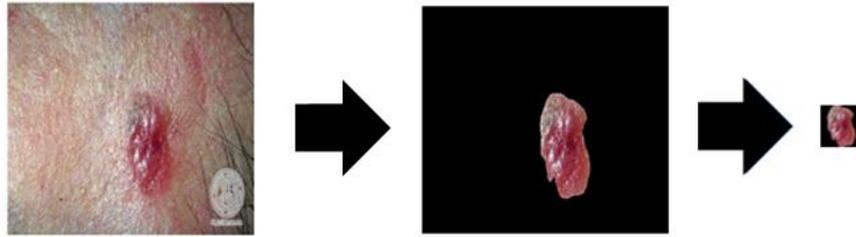


Figura 3: Contorno da região de interesse e redimensionamento da imagem

Após fazer o redimensionamento das imagens, foram eliminadas duas imagens, uma diagnosticada como melanoma e outra como não melanoma, pois verificou-se que estavam duplicadas. Dessa forma, obteve-se uma matriz X_m , dos casos de melanoma, com 118 linhas e 525 colunas, ou apenas com dimensão 118×525 . E uma matriz X_{nm} , dos casos de não melanoma, com dimensão 86×525 .

3.2. Extração de Características

Nesta etapa, o objetivo é extrair as características das regiões de interesse que representam de maneira mais significativa as diferenças entre os dados dos grupos analisados. Assim, essas características devem garantir que as regiões de interesse sejam classificadas corretamente como melanoma e não melanoma.

Para a extração destas características foi utilizada a técnica ICA, a imagem analisada é decomposta em uma combinação linear de imagens-base de modo que esta passa a ser representada por um vetor cujos elementos são os coeficientes de cada componente independente na mistura. Tais vetores são as características extraídas das imagens a partir da ICA [Leite, 2013]. Dessa forma, tem-se

$$x_i = a_{i1}s_1 + \dots + a_{in}s_n \text{ para todo } i = 1, \dots, n \quad (1)$$

sendo

x_i – imagem analisada

a_{ij} – coeficiente de mistura (reais)

s_i – imagens-base

ou representados na forma matricial

$$X = AS \quad (2)$$

onde teremos

$$X = [x_1 \quad \dots \quad x_n]^T \quad (3)$$

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} \quad (4)$$

$$S = [s_1 \quad \dots \quad s_n]^T \quad (5)$$

Portanto, a ICA permite derivar um algoritmo de seleção de variáveis independentes do modelo, baseado em um teste de dependência estatística. A estratégia é aplicar ICA para estimar a independência das entradas e então proceder a um teste estatístico para determinar o subgrupo desejado de variáveis de entrada.

A matriz X , de dimensão 204×525 , utilizada no modelo ICA foi obtida pela união da matriz X_m , dos casos de melanoma, de dimensão 118×525 , com a matriz X_{nm} , dos casos de não melanoma, de dimensão 86×525 . Assim, através da matriz X , e utilizando o algoritmo *fastICA*, obteve-se a matriz A , de dimensão 204×204 , com as características da amostra. Sendo que, cada coluna da matriz A corresponde a uma característica, e cada linha a uma amostra.

O algoritmo *fastICA* foi desenvolvido com o intuito de fornecer um método computacional mais rápido para a estimação das componentes independentes [Marchini *et al*, 2013]. Onde, as medidas em X tendem a ser mais gaussianas do que os componentes em S e por consequência, encontrar a matriz de separação W que maximiza a não-gaussianidade das fontes [Araujo *et al*, 2014]. Com isso, esta aproximação pode ser dada da seguinte forma:

$$j_{G(y)} = |E_y\{G(y)\} - E_v\{G(v)\}|^2 \quad (6)$$

Sendo v uma variável aleatória gaussiana normalizada, e y é assumido normalizado e com variância unitária.

3.3. Seleção das características mais significativas

Para a seleção de características utiliza-se a medida de máxima relevância, e é feita através da informação mútua (I), quantidade de informação que uma variável possui acerca de outra, entre as variáveis de cada característica v_i e de classe c .

$$\max D(v, c), \quad D = \frac{1}{|v|} \sum_{v_i \in v} I(v_i; c) \quad (7)$$

onde v é o vetor de características.

E como é provável que as características selecionadas apresentem uma grande dependência entre suas características [Ding e Peng, 2005], utiliza-se a medida de mínima redundância para selecionar as características mutuamente exclusivas, ou seja,

$$\min R(v), \quad R = \frac{1}{|v|^2} \sum_{v_i, v_j \in v} I(v_i; v_j) \quad (8)$$

A união dos métodos representados pelas equações (7) e (8) descritos acima, é denominada de máxima Relevância e Mínima Redundância [Peng *et al*, 2005]. E define-se o operador $\Phi(D, R)$ para fazer a combinação entre D e R . Logo em seguida, para otimizá-los simultaneamente

$$\max\Phi(D, R), \quad \Phi = D - R \quad (9)$$

Na seleção de características da matriz A foi aplicado, primeiramente, o algoritmo de máxima relevância, conforme a equação (7), que organizou cada característica, em ordem crescente de significância, ou seja, do mais significativo para o menos significativo. Logo após, como a dependência entre essas características ainda podem ser grande, aplica-se o algoritmo de mínima redundância, pela equação (8), para que a redundância, isto é, a similaridade entre essas características sejam a menor possível. E por fim, através da equação (9), fez-se a otimização para que a seleção das características obtivesse um resultado satisfatório.

Sendo assim, a redução do vetor de características de cada amostra foi realizada utilizando o algoritmo mRMR, ou seja, as características foram organizadas da mais significativa para as que menos se repetem.

Os dados da matriz A foram distribuídos em ordem crescente de representatividade, o que possibilita definir o número de características a serem utilizadas no classificador com a finalidade de encontrar o vetor de melhor desempenho [Linhares *et al*, 2016].

As características irrelevantes podem ser removidas sem comprometer o resultado da classificação, pois neste contexto, são consideradas redundantes, ou seja, implicam na presença de outra característica com a mesma funcionalidade, e não trazem nenhuma informação nova ao vetor de características [Araújo *et al*, 2014].

Os testes foram realizados incrementando 1 característica para cada iteração.

3.4. Métricas e Desempenho

A matriz de confusão foi utilizada para exibir uma medida eficiente de classificação. Uma vez que mostra o número de classificação correta para cada grupo determinado [Matos *et al*, 2009].

A matriz de confusão é uma matriz quadrada de números definidos em linhas e colunas que expressam o número de unidades da amostra atribuído a uma categoria particular relativa à categoria atual [Suarez e Candeis, 2012]. A diagonal principal da matriz de confusão exibe os dados classificados corretamente.

A qualidade dos testes apresentados foi feita através dos cálculos das medidas de acurácia, especificidade e sensibilidade. Onde a acurácia (Acc) é a capacidade do método acertar o diagnóstico, a sensibilidade ($Sens$) é a capacidade do método em reconhecer os doentes, e a especificidade ($Spec$) é a capacidade do método reconhecer os saudáveis.

Para tais medições foram utilizadas as seguintes fórmulas:

$$Acc = \frac{V_P + V_N}{V_P + V_N + F_P + F_N} \quad (10)$$

$$Sens = \frac{V_P}{V_P + V_N} \quad (11)$$

$$Spec = \frac{V_N}{V_N + F_P} \quad (12)$$

sendo,

V_P – o número de verdadeiros positivos

V_N – o número de verdadeiros negativos

F_P – o número de falsos positivos

F_N – o número de falsos negativos

3.5. Classificação

A etapa de classificação das amostras foi feita utilizando a técnica LDA, isto é a partir de um conjunto de imagens de treinamento gera-se um conjunto de vetores característicos que representa a imagem no espaço de características [Silva, 2016].

A LDA considera que a distribuição de probabilidade das amostras é conhecida e pode ser representada pela média de dispersão das amostras [Xavier *et al*, 2011]. Com isso, cria-se uma função discriminante linear para produzir, de acordo com suas características, uma máxima separação entre os subgrupos de imagens com melanoma e não melanoma.

$$D(X) = L' * X = [\mu_1 - \mu_2]' * \Sigma^{-1} * X \quad (13)$$

sendo que,

$$X = [X_1 \quad X_2 \quad \dots \quad X_n] \text{ e } \mu = [\mu_1, \mu_2] \quad (14)$$

onde,

L – é o vetor discriminante

X – é o vetor aleatório de características das populações

μ – é o vetor de médias p-variado

Σ - é a matriz comum de covariância das populações π_1 e π_2

Como a técnica LDA busca preservar o máximo de informação discriminatória possível dos dados originais, obtém-se a direção na qual a discriminação entre as classes se mantém a maior possível [Sbrissa Neto, 2015]. Visando encontrar um sistema de coordenadas ótimo, a LDA aplica uma transformação linear que garanta a máxima separabilidade entre as classes do conjunto de dados [Santos, 2005].

Por esse motivo, a técnica LDA foi utilizada na classificação das imagens. As características extraídas das imagens através da ICA e devidamente selecionadas pelo algoritmo mRMR serviram como os dados de entrada para a LDA.

Com base nesse conjunto de características, que equivalem às 204 imagens previamente classificadas por meio de análise visual, criou-se a função discriminante de Fisher, de acordo com a equação (13).

Após a classificação da LDA foi utilizada a técnica de validação cruzada *k-fold*. A validação cruzada *k-fold* baseia-se na divisão em *k* subconjuntos mutuamente exclusivos de um conjunto total de dados, onde são utilizados (*k* - 1) subconjuntos para a estimação dos parâmetros, ajuste do classificador, e 1 para avaliação [Junior, 2016].

Veja, pela figura 4, que o processo se repete *k* vezes, alternando os subconjuntos de tal forma que em cada iteração, um subconjunto seja utilizado como avaliação. A técnica tem como característica fazer uma melhor varredura da base de dados, atenuando dessa maneira problemas causados por diferenças na base [Severiano Jr, 2011].

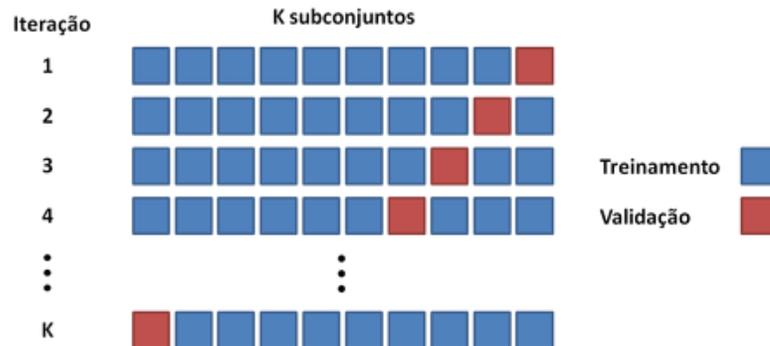


Figura 4: Técnica de validação cruzada

Portanto, as amostras foram divididas em 10 subconjuntos, onde 6 continham 20 amostras e 4 continham 21 amostras, com o objetivo de realizar o teste de validação cruzada *10-fold*, ou seja, 9 subconjuntos foram utilizados para o ajuste do classificador e 1 para a avaliação. A tabela 1 mostra os resultados obtidos no processo de classificação, para algumas das características analisadas.

Características	V _P	V _N	F _P	F _N	Acc(%)	Sens(%)	Spec(%)
20	101	37	17	49	67,65	67,33	68,52
40	101	48	17	38	73,04	72,66	73,85
60	100	55	18	31	75,98	76,34	75,34
80	100	60	18	26	78,42	79,36	76,92
100	104	70	14	16	85,29	86,67	83,33
150	114	83	4	3	96,27	97,44	95,40
185	118	86	0	0	100	100	100
204	118	86	0	0	100	100	100

Tabela 1: Desempenho do classificador

Dessa forma, analisando a tabela 1, pode-se verificar que de acordo com o aumento da quantidade de características, aumentam-se também as percentagens da acurácia, sensibilidade e especificidade. Para o vetor com 185 características, a exatidão é de 100%, a sensibilidade de 100% e a especificidade de 100%. E, que a partir do vetor de 185

características foi possível concluir que todas as 204 amostras, foram classificadas corretamente, sendo 118 melanomas (verdadeiro-positivo) e 86 não melanomas (verdadeiro-negativo), obtendo assim, 100% de acurácia, sensibilidade e especificidade.

4. Discussões

Com relação ao diagnóstico das imagens utilizadas, previamente classificadas por especialistas, o método proposto nesse trabalho se mostrou satisfatório, visto que, com o aumento das características analisadas a acurácia, a sensibilidade e a especificidade alcançaram uma taxa de acerto de 100%.

Para efeito de comparação com os trabalhos relacionados, pode-se dizer que o método proposto obteve resultados melhores. Mas não é possível afirmar que esta técnica é mais eficiente que as outras, tendo em vista que a utilização das bases de dados de imagens foram diferentes.

Sobre a relação com o diagnóstico clínico, esta técnica, é uma tentativa de encerrar as dificuldades encontradas pelos especialistas no diagnóstico do tumor. Uma vez que este diagnóstico é feito em várias etapas, onde o especialista deve fazer uma análise cuidadosa baseando-se apenas nas informações dos pacientes e no seu conhecimento sobre o caso.

Em relação ao diagnóstico dermatoscópico, o método proposto busca auxiliar o especialista na detecção precoce do câncer de pele, tendo em vista, em algumas cidades brasileiras não disponibilizam dos aparelhos necessários para a realização deste exame.

Como a realização de uma biópsia que é um procedimento invasivo necessário para o estudo histopatológico e causa inúmeros transtornos aos pacientes. A aplicabilidade desta técnica pode ser de extrema importância para que o diagnóstico do câncer de pele seja realizado sem danos extras.

Considerando que a detecção precoce é fundamental para se obter a cura do câncer de pele. Diante de todas as situações citadas acima, o método para o diagnóstico do câncer de pele apresentado neste trabalho, visa facilitar a detecção deste tipo de câncer no seu estágio inicial e assim, proporcionar o tratamento adequado aos pacientes.

5. Conclusão

Nesse trabalho foi utilizada uma técnica para auxiliar o diagnóstico precoce de lesões de câncer de pele denominada de Diagnóstico Auxiliado por Computador, onde foi utilizado um total de 204 imagens de lesões de câncer de pele previamente diagnosticadas como melanoma e não melanoma. Para tanto, empregou-se a análise de componentes independentes para a extração de características das imagens, para reduzir a dimensionalidade e selecionar as características mais significantes e menos redundantes foi utilizado o algoritmo de máxima relevância e mínima redundância. A classificação das imagens previamente diagnosticadas foi realizada pela Análise Discriminante Linear e a validação do método utilizado se deu a partir da validação cruzada *k-fold*.

Os resultados apresentados na seção 3 demonstram que a técnica utilizada alcançou um desempenho satisfatório. Uma vez que a técnica ICA se mostrou eficiente para a extração das características das imagens, o algoritmo de mRMR selecionou as características de forma que a classificação pela técnica LDA fosse de maneira mais precisa.

Apesar dos bons resultados previamente apresentados, a técnica proposta sugere um problema de sobre ajustamento, ou seja, o modelo estatístico ajustou-se demasiadamente ao conjunto de amostras. Em vista disso, outros testes deverão ser realizados em outras bases de dados com o propósito de aumentar a confiabilidade da técnica proposta. E, obtendo assim, uma melhor compreensão no diagnóstico e visando o desenvolvimento de softwares para que sejam testados em clínicas e hospitais auxiliando na diminuição da taxa de mortalidade para este e outros tipos de câncer.

Referências

ARAUJO, W.B.D.; Campos, L.F.A.; Furtado, A. S. (2004). Método de detecção de câncer de ovário utilizando padrões proteômicos, análise de componentes independentes e máquina de vetores de suporte. In: XIV Workshop de Informática Médica, vol. 14.

AZEVEDO-MARQUES, P. M. d. (2001). Diagnóstico auxiliado por computador na radiologia. Radiologia Brasileira, vol. 34, no. 5. p. 285–293.

DERMIS, S. D. I. (2012). Online atlases. Disponível em: <http://dermis.net/dermisroot/en/home/index.htm>. Acesso em 05 ago. 2016.

DERMQUEST. (2012). Online database. Disponível em: <https://www.dermquest.com/>. Acesso em 07 ago. 2016.

DING, C.; PENG, H. (2005). Minimum redundancy feature selection from microarray gene expression data. Journal of bioinformatics and computational biology. In: World Scientific vol. 3, no. 2. p. 185-205.

FRUTOSO, R. L.; SANTOS, J. R. V.; SIQUEIRA, R. S. (2013). Reconhecimento de cancer de pele do tipo melanoma. Instituto Federal de Educação, Ciência e Tecnologia do Ceará. Ceará, Brasil. Disponível em: <http://www.lbd.dcc.ufmg.br/colecoes/wim/2013/0013.pdf>. Acesso em 10 set. 2016.

INSTITUTO NACIONAL DO CÂNCER (2016a). Estimativa 2016: incidência de câncer no Brasil. Disponível em: <http://www.inca.gov.br/estimativa/2016/>. Acesso em: 10 ago. 2016.

INSTITUTO NACIONAL DO CÂNCER (2016b). Monitoramento das ações de controle do câncer de pele. Informativo Detecção Precoce, vol. 3. Disponível em: www1.inca.gov.br/inca/Arquivos/informativo_deteccao_precoce_03_2016.pdf. Acesso em: 11 set. 2016.

JUNIOR, E. E. R. (2016). Estratégias para classificação binária: um estudo de caso com classificação de e-mails. Universidade Federal do Paraná.

LEITE, I. C. C. (2-13). Análise de componentes independentes aplicada a avaliação de imagem radiográfica de sementes. Universidade Federal de Lavras.

LINHARES, J. d. N. et al. (2016). Método computacional para o diagnóstico precoce da granulomatose de wegener. *Revista de Informática Teórica e Aplicada*, vol. 23, no. 1. p. 277–292.

MAHMOUD, M. A.; AL-JUMAILY, A.; TAKRURI, M. (2013). Wavelet and curvelet analysis for automatic identification of melanoma based on neural network classification. *International Journal of Computer Information Systems and Industrial Management (IJCISIM)*. MIR Labs.

MARCHINI, J.; HEATON, C.; RIPLEY, B. (2013). Fastica: Fastica algorithms to perform ica and projection pursuit. R package version, 2013. p. 1–2.

MATOS, P. F. et al. (2009). Relatório técnico "métricas de avaliação". Universidade Federal de São Paulo, Universidade Federal de São Carlos, Universidade Metodista de Piracicaba, São Carlos, Brasil. Disponível em: <http://conteudo.icmc.usp.br/pessoas/taspardo/TechReportUFSCar2009a-MatosEtAl.pdf> Acesso em: 10 set. 2016.

PENG, H.; LONG, F.; DIRG, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*. Vol. 27, no. 8. p. 1226–1238.

PRADO, B. B. F. d. (2014). Influência dos hábitos de vida no desenvolvimento do câncer. *Ciência e Cultura*. Sociedade Brasileira para o Progresso da Ciência, vol. 66, no. 1. São Paulo. p. 21–24.

ROBBINS, S. L. et al. (2001). Fundamentos de robbins: patologia estrutural e funcional. In: *Fundamentos de robbins: patologia estrutural e funcional*. Guanabara Koogan.

ROSADO, L. (2009). Sistema automático para diagnóstico de lesões cutâneas baseado em imagens dermoscópicas. Tese (mestrado) - Instituto Superior Técnico-Engenharia Biomédica. Universidade Técnica de Lisboa, Lisboa.

SANTOS, A. R. d. (2005). Identificação de faces humanas através de PCA-LDA e redes neurais SOM. Tese (doutorado) - Universidade de São Paulo.

SBRISSA NETO, D. A. (2015). Análise da assimetria e irregularidade de borda entre lesões melanocíticas. Tese (doutorado) - Universidade de São Paulo.

SEIXAS, F. L.; SAADER, D. C. M. (2005). Diagnóstico auxiliado por computador. Universidade Federal Fluminense. Rio de Janeiro, Brasil. Disponível em: <http://www.midiacom.uff.br/~debora/fsmm/trab-2005-2/CAD.pdf>. Acesso em: 10 set. 2016.

SEVERIANO JR, C. (2011). Ajuste de análise discriminante linear semi-supervisionada através de validação cruzada ponderada. *Proceedings Seminário Interno da disciplina de Reconhecimento de Padrões*. p. 103.

SILVA, A. L. (2016). Redução de características para classificação de imagens de faces. Universidade do Estado do Rio Grande do Norte.

SILVA, S. F. d. (2012). Dermatology atlas. Disponível em: <http://www.atlasdermatologico.com.br>. Acesso em: 11 ago. 2016.

SOARES, H. B. (2008). Análise e classificação de imagens de lesões da pele por atributos de cor, forma e textura utilizando máquina de vetor de suporte. Tese (doutorado) - Universidade Federal do Rio Grande do Norte. Rio Grande do Norte, Brasil. Disponível em: http://www.repositorio.ufrn.br:8080/jspui/bitstream/123456789/15118/1/HelianaBS_TESE.pdf. Acesso em: 10 set. 2016.

STEWART, B. et al. (2016). World cancer report 2014. *World*.

SUAREZ, A. F.; CANDEIAS, A. L. B. (2012). Avaliação de acurácia da classificação de dados de sensoriamento remoto para o município de Maragogipe. In: IV Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação. Pernambuco.

XAVIER, A. C. et al. (2011). Análise discriminante e classificação de imagens 2d de ultrassonografia mamária. In: VII Workshop de Visão Computacional, Citeseer. 2011. p. 67–72.



Luciano Reis é licenciado em Matemática, Especialização em Matemática Financeira e Estatística pela Faculdade Venda Nova do Imigrante. Possui Mestrado em Engenharia da Computação e Sistemas pela Universidade Estadual do Maranhão. Atualmente é Professor de Matemática da Faculdade do Maranhão.



Lúcio Campos é doutor em Biotecnologia pela RENORBIO. Possui Mestrado em Engenharia Elétrica pela UFMA (2006). Atualmente é Professor da UEMA, atuando no curso de Engenharia da Computação, no Mestrado Profissional em Engenharia da Computação e Sistemas e coordenando o curso técnico à distância de Redes de Computadores. Tem experiência na área de Engenharia Elétrica, com ênfase em Telecomunicações e Processamento de Sinais Biológicos, atuando principalmente nos seguintes temas: Processamento de Imagens Médicas, Análise de Componentes Independentes, Redes Neurais Artificiais.