

Avaliação de Classificadores como Método de Segmentação de Lesões na Córnea

Pablo V. C. Lima¹, Rodrigo M. S. Veras¹,
Daniel A. Leite², Wallace F. Tales¹, Francisco M. C. Neto¹

¹Universidade Federal do Piauí (UFPI) - Campus Ministro Petrônio Portella
Centro de Ciências da Natureza (CCN) - Departamento de Computação
640649-550 - Teresina - PI - Brasil

²Universidade Federal do Piauí (UFPI) - Campus Ministro Petrônio Portella
Centro de Ciências da Natureza (CCN) - Departamento de Medicina Especializada
640649-550 - Teresina - PI - Brasil

{vinicius.pablo.18, oftalmodaniel, walacetalesfs, fmendescn}@gmail.com, rveras@ufpi.edu.br

Abstract. *As well as other areas of the body, the eye is the target of several kinds of diseases that directly affect vision, something essential to the human condition. This work proposes an application method for a Computer Aided Diagnosis system that uses corneal images marked by an ophthalmologist. The method seeks to segment the injured areas using Supervised Machine Learning algorithms, evaluating the performance of three different classifiers: Multi-Layer Perceptron, Random Forest and Support Vector Machine. Among the results obtained, the Random Forest algorithm performed best with Accuracy, Sensitivity and Specificity rates of 93.47%, 94.41% and 93.61%, respectively.*

Resumo. *Bem como outras áreas do corpo, o olho é alvo de diversos tipos de enfermidades que afetam diretamente a visão, algo essencial à condição humana. Esse trabalho propõe um método de aplicação para um sistema de Diagnóstico Auxiliado por Computador que utiliza imagens de córnea marcadas por profissionais oftalmologistas. O método busca segmentar as áreas lesionadas utilizando algoritmos de Aprendizado de Máquina Supervisionados, avaliando o desempenho de três classificadores distintos: Multi-Layer Perceptron, Random Forest e Support Vector Machine. Dentre os resultados obtidos, o algoritmo Random Forest se saiu melhor com taxas de Acurácia, Sensibilidade e Especificidade em 93,47%, 94,41% e 93,61%, respectivamente.*

1. Introdução

No seu recente estudo, [Flaxman and Bourne 2017] cita um alarmante dado: em 2015, 36 milhões de pessoas eram afetadas pela cegueira em todo o mundo. Além de enfermidades como catarata, erro de refração não corrigido, glaucoma, degeneração macular relacionada à idade, opacidade da córnea, tracoma, e retinopatia diabética, as causas socioeconômicas são fatores consideráveis para o aumento dessas estatísticas.

Essas enfermidades podem afetar diversas áreas no sistema ocular humano, dentre elas a córnea. Como visto na Figura 1, a córnea pode ser atingida por diversos tipos de lesões que, se não tratadas a tempo e de forma devida, podem ocasionar a cegueira. Essa

variedade de lesões é um desafio para um profissional oftalmologista, e ainda mais para sistemas computacionais que auxiliam os profissionais da medicina, chamados sistemas de Diagnóstico Auxiliado por Computador (CAD - *Computer Aided Diagnosis*).

Em [Tapan P. Patel and Woodward. 2017] é proposto um método de segmentação baseado em classificadores que busca verificar a proporção entre as áreas classificadas como lesão e as não lesão e comparar o desempenho do sistema CAD com métodos de segmentação manuais. Esse trabalho, por sua vez, tem como objetivo a avaliação de classificadores no procedimento de segmentação semiautomática das áreas afetadas, utilizando marcações de áreas lesionadas e não lesionadas feitas pelo profissional oftalmologista, buscando diminuir a sensibilidade a erros humanos ao lançar mão de técnicas de Processamento Digital de Imagens e Inteligência Artificial.

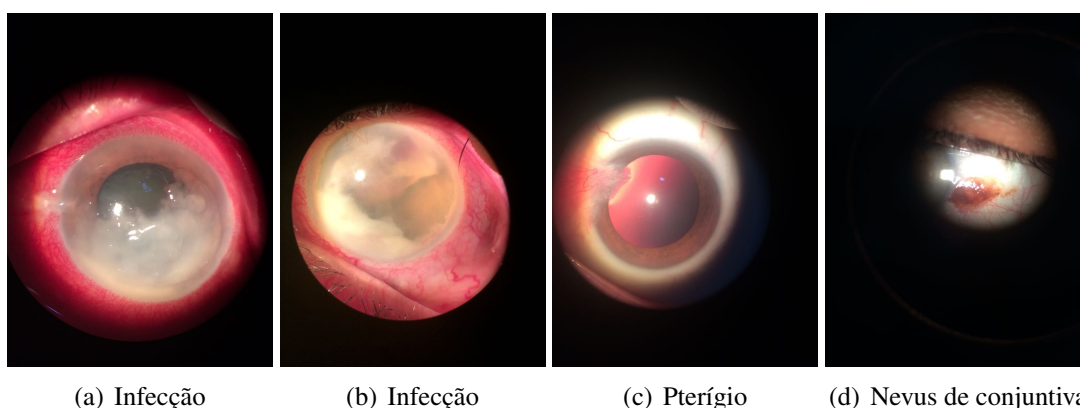


Figura 1. Diferentes tipos de lesões que podem acometer a córnea.

2. Método Proposto

A segmentação das áreas lesionadas da córnea é um processo de grande valia para o médico no decorrer do acompanhamento do paciente. Isso permite que a evolução de um tratamento, ou até mesmo da doença, seja avaliada, possibilitando que o profissional oftalmologista tome decisões baseadas na condição clínica em que o paciente se encontra.

O método proposto por este trabalho é uma forma semiautomática de segmentação que utiliza marcações feitas por um médico em imagens de córnea, de partes lesionadas e não lesionadas, para alimentar algoritmos de classificação.

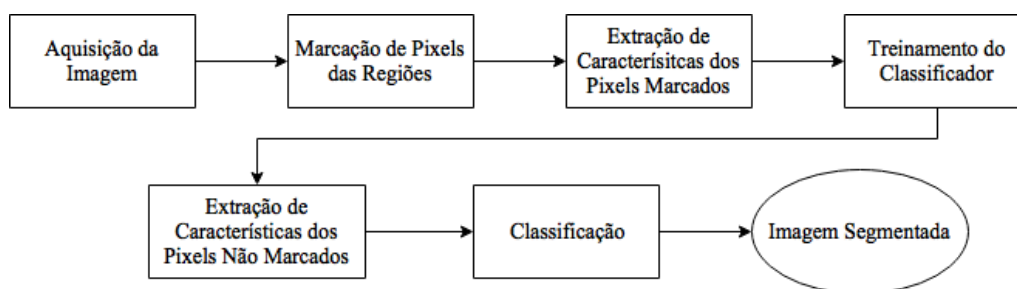


Figura 2. Fluxograma do método proposto.

A primeira etapa parte da captura e marcação das imagens de córnea que serão providas pelo médico oftalmologista. Nesse procedimento, o médico captura uma ima-

gem da região do olho e marca duas áreas: a primeira representa a área ocular, o que não inclui pálpebras e outras regiões que circundam o olho; e a segunda representa a área considerada lesão. Com as imagens marcadas, algoritmos auxiliares implementados segmentam essas áreas e geram duas máscaras de imagens binárias, a região lesionada e a região não lesionada. Essas máscaras serão utilizadas para dividir o conjunto de *pixels* marcados entre treino e teste, selecionando aleatoriamente uma porcentagem de *pixels* de cada área a fim de simular uma marcação menor do especialista em um sistema real. Aqueles selecionados para o conjunto de treino terão características de cor extraídas e utilizadas no procedimento de adaptação do classificador. Uma vez terminado o procedimento de treino, os dados do conjunto de teste são submetidos ao classificador, que irá prever os *pixels* entre lesão e não lesão, baseado nas mesmas características de cor utilizadas anteriormente.

3. Materiais e Métodos

3.1. Base de Dados

Contendo imagens da superfície da córnea, a base de dados foi colhida em procedimentos médicos em condições diversas de iluminação e equipamentos, aumentando a variabilidade de dados. Atualmente com 15 imagens, a base encontra-se em constante construção.

A captura das imagens foi realizada através de lâmpada de fenda modelo Topcon Sle1, com aumento de 10X, em iluminação difusa. As fotos foram obtidas por meio da câmera de um *smartphone* iPhone 5s, conectado na ocular da lâmpada. Após a captura das imagens, a área corresponde a lesão no segmento anterior do olho foi marcada em vermelho e a área total em análise em azul, com auxílio de uma caneta para toque em telas digitais.

Ademais, as imagens foram capturadas em diferentes clínicas da cidade de Teresina - PI e contêm diferentes tipos e níveis de lesão na córnea. A Figura 3 apresenta exemplos de uma imagem original, marcada e suas respectivas máscaras geradas a partir da marcação do médico.

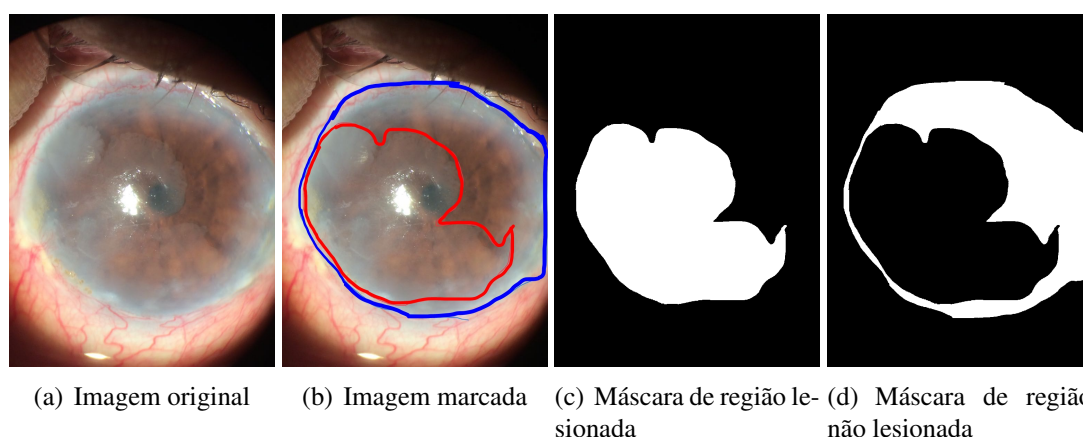


Figura 3. Exemplo de imagem utilizada.

3.2. Métricas de Avaliação

Para validar o desempenho do método proposto, uma vez utilizados algoritmos de Aprendizado de Máquina Supervisionados, calculou-se a matriz de confusão, essa pos-

sui a relação entre dados reais, a verdade terrestre, e aqueles preditos pelo classificador, permitindo avaliar seu desempenho.

As linhas representam os valores preditos pelo classificador, enquanto as colunas representam a verdade terrestre, gerando os conceitos de Verdadeiro Positivo (VP), para os *pixels* preditos corretamente como positivos; Verdadeiro Negativo (VN), para os *pixels* corretamente preditos como negativos; Falso Positivo (FP), para os *pixels* originalmente negativos, mas que foram preditos positivos; e Falso Negativo (FN), para os *pixels* positivos preditos como negativos. Sendo assim, os valores de VP, VN, FP e FN são utilizados para calcular as métricas de avaliação utilizadas, Acurácia (A), Sensibilidade (S) e Especificidade (E), de acordo com as equações descritas abaixo.

A primeira delas, Acurácia, representa a proporção entre as predições corretas, sem fazer distinção entre positivos e negativos; é a razão entre os valores preditos corretamente, positivos e negativos, e a soma de todos os dados. A Sensibilidade, por sua vez, representa a proporção dos verdadeiros positivos, ou seja, avalia se o algoritmo consegue prever que um elemento pertence a um dado grupo quando ele de fato pertence, é a razão dos VP e a soma dos VP e FN. E, por fim, a Especificidade representa a proporção dos verdadeiros negativos, ou seja, avalia se o algoritmo consegue prever que um elemento não pertence a um dado grupo quando ele de fato não pertence, é a razão entre VN e a soma de VN e FP.

4. Descritores e Classificadores

Buscando melhores resultados, testamos três classificadores distintos que são amplamente citados na literatura. Primeiramente utilizamos o *Multi-Layer Perceptron* (MLP). [Russel and Norvig 2013] afirmam que uma rede com todas as entradas conectadas diretamente com as saídas é chamada de rede neural de camada única ou rede *perceptron*. Assim, uma rede *perceptron* é formada por um único neurônio. O MLP é, então, uma rede formada por várias redes *perceptron* unidas, formando uma rede multicamadas. Na aplicação deste trabalho, o MLP foi configurado com 100 camadas ocultas, sendo a n -ésima camada oculta formada por n neurônios.

O segundo classificador escolhido, *Random Forest* (RF), também chamado de *Random Decision Forests*, é um método de classificação que constrói diversas *Random Trees*, outro classificador, ou seja, combina o resultado de diversos classificadores a fim de chegar em um resultado final. Além disso, têm a capacidade de corrigir o problema de sobre-treinamento das *Random Trees*, tornando-se bastante poderoso. Neste trabalho, a RF foi configurada contendo dez árvores na floresta.

Por fim, *Máquina de Vetor de Suporte* (SVM) é um método com treinamento supervisionado, usado tanto em problemas de classificação como de regressão [Vapnik and Cortes 1995]. Segundo [Pereira and Centeno 2013], SVM é um classificador supervisionado que busca encontrar um hiperplano que separa os dados em classes discretas, utilizando um processo iterativo e o conjunto de treino para ajustar esse hiperplano de maneira ótima. No classificador SVM deste trabalho utilizamos o *kernel* rbf, devido à sua maior adaptação a conjuntos de dados não linearmente separáveis e um *gamma* igual a 1/12 (número de características). Essa configuração foi escolhida empiricamente e mostrou o melhor resultado em testes.

Para efetuar o treinamento destes, é preciso organizar informações que represen-

tem cada classe existente. Na nossa realidade, isso se traduz na extração de características das partes de lesão e não lesão das imagens. Uma vez que escolhemos trabalhar com a informação de cada *pixel*, a forma definida para representá-los foi por meio dos canais de cor RGB, HSV, YCBCR e LAB, já que são formas diferentes de descrever um mesmo *pixel*, aumentando a variabilidade do vetor de características. De posse dessas informações, montamos um vetor de doze características para cada *pixel*, submetendo tais informações juntamente às classes dos *pixels* (lesão e não lesão) ao procedimento de treino dos classificadores.

5. Resultados e Discussão

Os algoritmos escolhidos foram executados cinco vezes para cada imagem da base de dados. Após a execução e cálculo de resultados para cada classificador, uma média dos cinco resultados foi obtida, bem como o desvio padrão (σ) desses dados.

A quantidade de dados do conjunto de treinamento foi determinada, inicialmente, com 50% da menor quantidade entre *pixels* de lesão e não lesão, o que garante que o classificador seria treinado com uma quantidade igual de dados para ambas as classes.

Como representado na Tabela 1, as métricas de avaliação variaram entre os diferentes algoritmos de classificação. Os melhores algoritmos, RF e SVM, atingiram resultados semelhantes, entre 92% e 95% de média para todas as métricas, com uma leve vantagem para o RF na Acurácia e Especificidade; este também obteve os menores, e portanto melhores, valores de desvio padrão, traduzindo uma boa classificação dos *pixels* lesionados. Apesar de ter os piores resultados, o MLP ainda atingiu valores de média entre 88% e 89% em todas as métricas, além dos mais altos valores de desvio padrão. Os dados destacados representam as melhores métricas atingidas.

Tabela 1. Resultados com 50% dos *pixels* para treino.

	Acurácia		Sensibilidade		Especificidade	
	Média	σ	Média	σ	Média	σ
MLP	88,02	6,56	88,57	6,24	88,05	7,58
RF	93,47	3,45	94,15	3,17	93,61	4,13
SVM	92,90	3,49	94,41	2,97	92,72	4,48

Ainda assim, é importante notar que se tratando de uma aplicação para auxiliar a segmentação de lesões de maneira semiautomática, 50% dos *pixels* é uma quantidade alta para entrada do algoritmo, uma vez que o objetivo é tornar o procedimento funcional mesmo com poucos *pixels* marcados pelo médico. Por isso, novos testes foram feitos com uma menor porcentagem de *pixels*, agora 30%, a fim de avaliar o comportamento dos algoritmos perante essa nova configuração.

A Tabela 2 mostra que apesar da diminuição do conjunto de treino e seu consequente efeito nas métricas de avaliação, os resultados continuam satisfatórios, indicando que os algoritmos funcionam bem com menos dados para treino. Na Figura 4 pode-se ver imagens resultantes do processo de segmentação para os três algoritmos, utilizando 50% de dados para treino.

Tabela 2. Resultados com 30% dos *pixels* para treino.

	Acurácia		Sensibilidade		Especificidade	
	Média	σ	Média	σ	Média	σ
MLP	87,58	6,50	88,79	6,81	87,22	7,28
RF	93,13	3,54	94,81	3,28	93,03	4,17
SVM	92,52	3,62	93,58	3,18	92,31	4,54

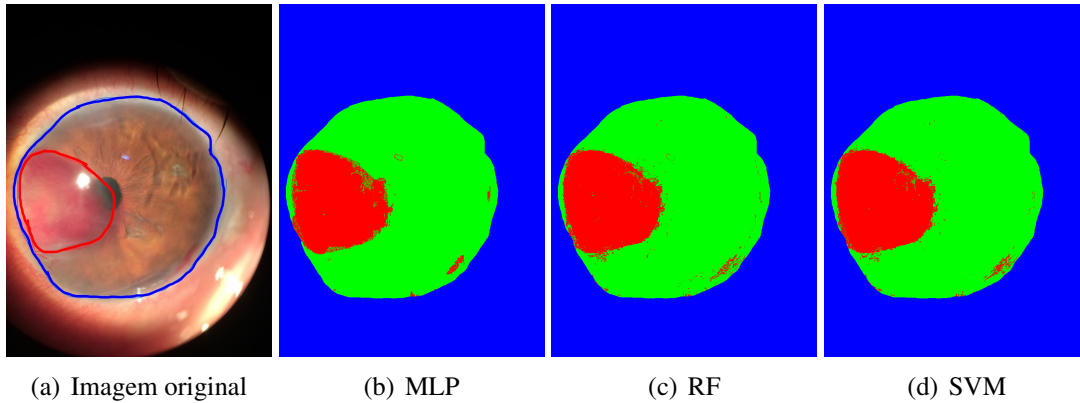


Figura 4. Resultados da classificação.

6. Conclusão e Trabalhos Futuros

De acordo com os resultados obtidos, apesar da existência de uma taxa de erro, o método proposto nesse trabalho mostra-se promissor como uma forma de auxílio aos profissionais oftalmologistas, abrindo um leque de melhorias possíveis a fim de aumentar a precisão e confiabilidade do sistema.

Deste modo, como trabalhos futuros propõe-se testar novos classificadores, incluindo aqueles do paradigma de Aprendizado de Máquina Semissupervisionado, assim como a adição de uma etapa de pós-processamento a fim de remover áreas erroneamente classificadas e distantes do centro da lesão. Como forma de diminuir a dependência do número de *pixels* marcados, é possível a aplicação de novas técnicas de extração de características para melhor representar as áreas lesionadas e não lesionadas.

Referências

- Flaxman, S. R. and Bourne, R. (2017). Global causes of blindness and distance vision impairment 1990-2020: a systematic review and meta-analysis.
- Pereira, G. H. A. and Centeno, J. A. S. (2013). Utilização de support vector machine para classificação multiclass de imagens landsat tm.
- Russel, S. and Norvig, P. (2013). *Inteligência Artificial*. Elsevier, 3th edition.
- Tapan P. Patel, N Venkatesh Prajna, S. F. N. G. V. L. M. N. L. D. K. H. K. and Woodward., M. A. (2017). Novel image-based analysis for reduction of clinician-dependent variability in measurement of the corneal ulcer size.
- Vapnik, V. and Cortes, C. (1995). Support-vector networks. machine learning. n. 20, p273–p297, 1995.