

A Forensic Nudity Detector Based on Machine Learning

Danilo Coura Moreira¹, Joseana Macêdo Fechine¹

¹Universidade Federal de Campina Grande - UFCG

danilocoura@copin.ufcg.edu.br, joseana@computacao.ufcg.edu.br

Abstract. *Forensic examinations on computer devices become increasingly costly due to the large amount of files they can store. It is necessary to use algorithms that could detect pornographic images efficiently. We presented an approach that improved the work of Ap-Apid using a machine learning-based algorithm rather than using static rules. In addition, we used features extracted from a face detector. The AIIA-PID4 pornographic data set was used in the experimental phase. In the end, the proposed model reached an accuracy of 93.56%, outperforming the referenced works, which reached 79.1% and 85.05%, respectively.*

Resumo. *Os exames periciais em dispositivos computacionais tornam-se cada vez mais custosos devido a grande quantidade de arquivos que estes podem armazenar. Faz-se necessário utilizar algoritmos capazes de realizar a detecção de imagens pornográficas de uma maneira eficaz. Propusemos uma abordagem que aprimora o trabalho de Ap-Apid, utilizando um algoritmo baseado em aprendizagem de máquina ao invés de regras estáticas, além de utilizar características extraídas de um detector de faces. Foi utilizada na fase experimental a base de dados de imagens AIIA-PID4 pornographic data set. Por fim, o modelo proposto atingiu uma acurácia de 93,56%, superando os trabalhos referenciados, que atingiram 79,1% e 85,05%, respectivamente.*

1. Introdução

Atos preparatórios, de execução, ou até mesmo de consumação de um crime muitas vezes sequer eram registrados e, quando esse registro ocorria, se dava por meios analógicos de difícil propagação e alcance limitado. Em poucos segundos, devido o uso crescente dos microcomputadores e *smartphones*, aliado ao acesso cada vez mais fácil à rede mundial de computadores, dispositivos são capazes de capturar, armazenar imagens e facilmente compartilhá-las para outros dispositivos [Cappellari 2005].

Além do mais, paulatinamente os dispositivos de armazenamento conseguem manter cada vez mais arquivos de mídia. O trabalho de Eleutério e Machado [Eleutério and Machado 2011] mostrou a existência de mais de 300.000 arquivos de imagens em um dispositivo dessa natureza. Por fim, os autores afirmam que após a realização do exame pericial foram encontrados pouco mais de uma centena de arquivos que evidenciaram algum tipo de crime.

O estudo de Platzer et al. [Platzer et al. 2014] afirma que a realização de uma inspeção manual de imagens ilícitas, por um longo período de tempo, resulta na diminuição da capacidade cognitiva e de concentração humana. Esses efeitos se dão devido a lentidão e monotonia da referida tarefa. Dessa forma, esta atividade esta sujeita ao

erro humano, fazendo com que imagens ilícitas passem despercebidas, aumentando assim a taxa de falsos negativos. Por isso, é necessário que esta busca seja automatizada e que possa ser desempenhada em alta velocidade e com baixas taxas de falsos negativos.

Entretanto não se trata de uma tarefa trivial, principalmente quando se lida com a distinção entre imagens de nudez e não-nudez. Muitas vezes é difícil para o próprio ser humano julgar essa diferença, pois existem imagens muitas vezes com conteúdo subjetivo que poderá possuir classificações distintas de acordo com a observação de diferentes pessoas [Putro et al. 2015].

Portanto este trabalho apresenta uma abordagem para a distinção entre imagens com conteúdo de nudez e não-nudez, aumentando a capacidade de predizer se uma imagem é pornográfica ou não, tendo como principal objetivo subsidiar os exames periciais na busca por imagens que possam possuir pornografia infantil.

A primeira etapa do método proposto é responsável por realizar a extração das características da imagem por meio de um detector de pele baseado em regras. Esta fase atua nos espaços de cores RGB e YCbCr e possui um detector de face baseado no trabalho de Viola e Jones [Viola and Jones 2004]. A etapa posterior tem a função de classificar em que categoria se encontra determinada imagem. Nesta etapa foi utilizada a técnica de validação cruzada para encontrar os hiperparâmetros que apresentaram melhor comportamento nos modelos baseados em aprendizagem de máquina, selecionando o modelo detentor da melhor acurácia. Por fim, na etapa de testes, o modelo baseado em Floresta Aleatória conseguiu, em 93,56% das vezes, distinguir de maneira correta entre imagem contendo nudez e não-nudez.

Dessa forma, o método proposto melhorou a acurácia do algoritmo do trabalho de Ap-Apid [Ap-Apid 2005] (“Um Algoritmo Para a Detecção de Nudez”) em 18,28%. Foram utilizados detectores de pele e face considerados simples, associados ao uso de algoritmos baseados em aprendizagem de máquina para classificação.

Este artigo foi organizado da seguinte maneira: Toda fundamentação teórica encontra-se na Seção 2. Os trabalhos que nos motivaram estão contidos na Seção 3. Em seguida, na Seção 4, está descrito nosso método em detalhes. Na Seção 5, os experimentos e resultados são discriminados com o intuito de validar o método exposto. Por fim, na Seção 6, são apresentadas as conclusões e perspectivas de trabalhos futuros.

2. Fundamentação Teórica

Essa seção tem como objetivo apresentar toda a fundamentação teórica necessária para o entendimento da construção do modelo proposto, onde são apresentados conceitos de imagem digital, detecção de pele e de face.

2.1. Imagem Digital e Espaço de Cor

Uma imagem digital é composta por uma matriz de píxeis, em que cada píxel é representado por valores de cada canal de um espaço de cor. Existem diversos espaços de cores, sendo adotados de acordo com o tipo da aplicação. Segundo Xiong e Li [Xiong and Li 2012], os espaços de cores são formados por elementos de cromaticidade e luminância que, respectivamente, se referem aos valores de cores e de luz. Os espaços de cores podem ser divididos em 4 tipos: **Básicos** (e.g. RGB, nRGB), **Ortogonais** (e.g.

YCbCr, YDbDr, YPbPr, YUV, YIQ), **Perceptuais** (e.g. HSV, HSL, HSI) e **Perceptualmente Uniformes** (e.g. IE-Lab, CIE-Luv, CIE-LCH, CIE-CAT02 LMS).

2.2. Detecção e Segmentação de Pele

Mesmo com o constante desenvolvimento de novas técnicas na área da visão computacional, a detecção de pele ainda se mantém como parte integral de várias de suas aplicações, como detecção de face, detecção de pessoa, detecção de conteúdo ilícito, entre outras [Muhammad and Abu-Bakar 2015]. A detecção de pele pode ser considerada um problema de classificação binária, no qual seu classificador infere se um determinado píxel é considerado pele ou não [Xiong and Li 2012]. Apesar da definição simplória, não se trata de uma tarefa trivial, pois existem muitos fatores que alteram a padronização das cores relacionadas à pele humana (e.g. iluminação, plano de fundo, equipamento de captura, características pessoais, entre outros) [Mahmoodi and Sayedi 2016]. Atualmente existem diversos classificadores utilizados na detecção de peles em seres humanos (e.g. baseado em regras, baseado em histogramas, paramétricos, baseados em aprendizagem de máquina).

O classificador **baseado em regras** é o menos complexo, sendo altamente recomendado para aplicações em tempo real ou que necessitem de um ótimo tempo de resposta. Essa característica se dá pela menor necessidade de recursos computacionais por possuir cálculos mais simples que os demais classificadores. Além disso, levam vantagem por não possuírem a etapa de treinamento. Entretanto apresentam resultados inferiores comparado aos classificadores mais sofisticados [Vezhnevets et al. 2003].

O trabalho de Muhammad e Abu-Bakar [Muhammad and Abu-Bakar 2015], visando ser menos susceptível a heterogeneidade da luz, utilizou apenas os componentes de crominância dos espaços de cores HSV e YCgCr. Para chegar aos intervalos utilizados nas regras, os autores utilizaram inicialmente um histograma de frequência de píxeis de pele. A abordagem no estudo de Patil et al. [Patil et al. 2013] fez uso da mesma estratégia, contudo, utilizando os os componentes de crominância HCbCr dos espaços de cores HSV e YCbCr.

Com o objetivo de dirimir a dificuldade de detectar diferentes cores de pele, os autores encontraram uma nova faixa de valores no espaço de cor YCbCr para detectar pele humana independente de sua raça. Este experimento também foi realizado por meio da análise de histogramas dos píxeis das imagens [Basilio et al. 2011].

O trabalho de Kovac et al. [Kovac et al. 2003] utilizou diferentes regras no espaço de cor RGB para diferentes condições de luminosidade (i.e. luz matutina uniforme, luz artificial ou luz matutina leve). Já na pesquisa realizada por Platzer et al. [Platzer et al. 2014], percebe-se que foi feito o uso de regras para detecção de pele em condições heterogêneas de luminosidade nos espaço de cor RGB e HSV.

Outra categoria de classificador é o **baseado em histogramas**. Essa técnica não utiliza nenhuma espécie de densidade de probabilidade, requerendo assim uma quantidade elevada de dados para o treinamento do modelo. Funciona da seguinte maneira: são criados histogramas de frequência de píxeis, sendo todos contabilizados e a partir de então realiza-se um mapeamento dessa distribuição. Ao final, por meio do uso de um classificador Bayesiano, infere-se a probabilidade de um píxel ser de pele ou não [Mahmoodi and Sayedi 2016].

Foi realizado um comparativo no estudo de Khan et al. [Khan et al. 2012] entre 9 detectores de pele distintos (i.e. *AdaBoost*, Rede Bayesiana, Árvore de Decisão, Perceptron Multicamadas, *Naive Bayes* Gaussiano, Floresta Aleatória, Rede Neural RBF, Máquina de Vetores de Suporte e Histograma). O classificador baseado em histograma obteve apenas maior acurácia que o *AdaBoost*.

Já os classificadores **paramétricos** simulam uma distribuição real, como as baseadas em histogramas. Essa simulação se dá por meio do uso de funções de densidade de probabilidade. Dessa forma, são necessários menos dados para treinamento do modelo sem comprometer o resultado final. São fundamentalmente baseados em modelos gaussianos (e.g. Modelo Gaussiano Simples, Mistura de Modelos Gaussianos, Modelo Gaussiano Múltiplo) e elípticos (e.g. *Elliptical Boundary Model*) [Mahmoodi and Sayedi 2016].

Ficou claro no trabalho de Youtian et al. [Youtian et al. 2012] que o modelo baseado em Múltiplas Gaussianas apresentou uma melhor taxa de verdadeiros positivos em comparação com os modelos baseados em Gaussiana Simples, Misturas Gaussianas e *Elliptic Boundary*.

Por fim, os classificadores **baseados em aprendizagem de máquina**. Este tipo de classificador normalmente necessita de mais tempo nas fases de treinamento e de teste, pois apresentam uma maior complexidade comparada aos demais tipos de classificadores, principalmente na fase de treinamento para que o modelo possa aprender os padrões de cada classe (e.g. Regressão Logística, Árvore de Decisão, Floresta Aleatórias, Redes Neurais) [Oghaz et al. 2015].

Em seu estudo, Khan et al. [Khan et al. 2012] também comprovaram que os classificadores baseados em aprendizagem de máquina superaram os demais classificadores baseados em histogramas e os paramétricos.

Na pesquisa de Ma et al. [Ma et al. 2014] foram comparados dois classificadores baseados em aprendizagem de máquina (Árvore de Decisão e Floresta Aleatória) e um baseado em regras. Por mais uma vez, os classificadores baseados em aprendizagem de máquina apresentaram os melhores resultados.

2.3. Detecção de Face

A técnica utilizada para encontrar regiões de píxeis que representem uma face humana em uma imagem digital é denominada detecção de face [Lucena et al. 2017]. Hoje o estado-da-arte nesse segmento é dado pelo uso de algoritmos baseados em Redes Neurais Convolucionais Profundas. Apesar de alcançar excelentes resultados, seu desempenho de processamento ainda está aquém dos classificadores mais simples, como por exemplo os baseados na extração das características *Haar*. Esta técnica utiliza um conjunto de classificadores simples (*AdaBoost*), sendo o algoritmo aplicado por Viola e Jones [Viola and Jones 2004] o mais conhecido dessa família e extremamente recomendado para aplicações em tempo real ou que demandem alta velocidade [Zafeiriou et al. 2015].

3. Trabalhos Relacionados

É vasta a literatura sobre detecção de pornografia e nudez baseados nas características de pele que não é contemplada nessa seção. Entretanto os trabalhos considerados relevantes e que mais influenciaram nosso trabalho serão mencionados aqui.

Assim como o trabalho proposto, diversos trabalhos tiveram como maior inspiração o estudo de Ap-Apid [Ap-Apid 2005] (“Um Algoritmo Para a Detecção de Nudez”), que utiliza características extraídas de um detector de pele como os de [Polastro and Eleutério 2010, Platzer et al. 2014, Medina and Palladino 2017]. Este detector de pele é baseado em regras que foram encontradas por meio do uso de uma regressão linear nos espaços de cores nRGB e HSV. Estas características foram extraídas principalmente das três maiores regiões de pele. Por fim, foram obtidos limiares, de maneira empírica, que determinam se uma imagem é pornográfica ou não.

Em seu trabalho, Polastro e Eleutério [Polastro and Eleutério 2010] utilizaram outras técnicas em combinação ao algoritmo de Ap-Apid [Ap-Apid 2005]. São exemplos a análise dos nomes dos arquivos e o uso de uma lista negra contendo o código *hash* de arquivos previamente catalogados como ilegais.

O trabalho de Medina e Palladino [Medina and Palladino 2017] difere do de Ap-Apid [Ap-Apid 2005] principalmente em 2 aspectos: (i) no formato do delimitador (retangular) para segmentar as três maiores regiões de pele contidas na imagem e (ii) nos espaços de cores adotados (RGB e YCbCr) para a detecção de pele.

A pesquisa de Platzer et al. [Platzer et al. 2014] é a que mais difere da de Ap-Apid [Ap-Apid 2005]. Platzer et al. [Platzer et al. 2014] usaram uma grande variedade de características (e.g. porcentagem de pele, compacidade, elipticidade, retangularidade, excentricidade, orientação, *Hmean*). Além do mais, utilizaram as cinco maiores regiões de pele para extração das características, gerando um vetor de tamanho 43. Uma Máquina de Vetores de Suporte foi utilizada para classificar a imagem de entrada. Por fim, adicionaram etapas de pré e pós processamento para o detector de pele, além de implementar várias regras estáticas visando a redução de falsos positivos.

O estudo de Karavarsamis et al. [Karavarsamis et al. 2013] utilizou um algoritmo de decomposição planar geométrica para detectar regiões de interesse em uma imagem colorida. Tomou-se como base a pesquisa realizada por Yang et al. [Yang et al. 2007]. Foram utilizadas características que são baseadas nessas regiões de interesse, resultando em um vetor de características de tamanho 15. Uma Floresta Aleatória foi usada para treinar e classificar as imagens. Este trabalho utilizou sua própria base de dados de imagens, denominada *AIIA-PID4 pornographic data set*.

No estudo de Putro et al. [Putro et al. 2015] foram usados classificadores baseado em regras, tanto para o detector de pele, quanto para a classificação da natureza da imagem em si. Foram utilizadas apenas três características para diferenciar imagens pornográficas de não pornográficas: 1) a razão entre os píxeis das faces detectadas e de toda a imagem; 2) a posição da face detectada na imagem e 3) a proporção dos píxeis detectados das faces em relação à quantidade de píxeis de pele detectados.

Também com o objetivo de reduzir falsos positivos em retratos¹, Zhou et al. [Zhou et al. 2016] utilizaram a razão entre o número de faces detectadas e o número de píxeis de pele da imagem de entrada.

¹imagens em que a maior parte do seu conteúdo é uma face humana

4. Trabalho Proposto

A metodologia proposta, ilustrada pela Figura 1, é formada por duas etapas em sequência: A (i) Extração das Características e o (ii) Classificador. Mesmo sendo amplamente utilizada na literatura, nossa proposta se diferencia em três pontos principais: 1) O uso dos espaços de cores RGB e YCbCr; 2) a adoção de características de pele mais simples e em menor quantidade e 3) a utilização do número de faces detectadas e seus tamanhos, com o objetivo de minimizar os falsos positivos em retratos.

4.1. Extração de Características

Esta primeira etapa tem como objetivo obter as características da imagem por meio de dois módulos independentes: a (1) Detecção e Segmentação de Pele e a (2) Detecção de Face. As características extraídas dessa etapas atuam como a representação das imagens, contendo as singularidades de cada categoria e tornando-se subsídio para a etapa de classificação.

4.1.1. Detector e Segmentador de Pele

É imprescindível a necessidade de um detector de pele que, além de acurado, apresente bom desempenho computacional, tendo em vista a grande quantidade de imagens que são inspecionadas no exames forenses [Eleutério and Machado 2011].

Baseado nesses anseios, foi utilizado um detector de pele híbrido usado por Medina e Palladino [Medina and Palladino 2017]. Este classificador baseia-se nas regras propostas por Kovac et al. [Kovac et al. 2003], para detecção de píxeis de pele em condições heterogêneas de luminosidade no espaço de cor RGB e por Basilio [Basilio et al. 2011], para buscar por píxeis de pele humana de diferentes raças no espaço de cor YCbCr. Em suma, para ser considerado um píxel de pele, este precisa estar contido simultaneamente nos intervalos (1) e (2) dos espaços de cores YCbCr e RGB, respectivamente.

$$80 < Cb < 120 \quad \& \quad 133 < Cr < 173 \quad (1)$$

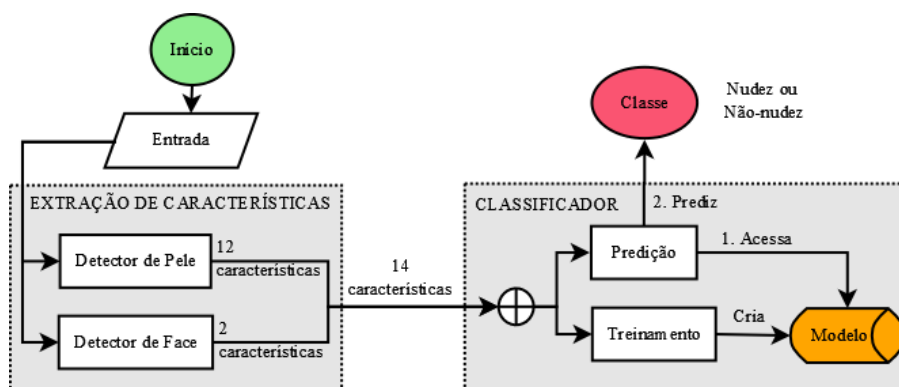


Figura 1. Diagrama de fluxo do algoritmo proposto mostrando suas duas fases em detalhes: (i) Extração das Características e (ii) Classificador.

$$\begin{aligned}
& (R > 95 \ \& \ G > 40 \ \& \ B > 20 \ \& \\
& \max(R, G, B) - \min(R, G, B) > 15 \ \& \\
& |R - G| > 15 \ \& \ R > G \ \& \ R > B)
\end{aligned} \tag{2}$$

Após a segmentação da pele, quatro regiões de interesse, como mostrado na Figura 2, foram escolhidas para a realização da extração das características: 1) a imagem completa (R1); 2) o menor retângulo que englobe as duas maiores regiões de pele (R2) e 3) as duas maiores regiões de pele, individualmente (R3 e R4). Para cada uma dessas regiões foram consideradas como características: 1) A área total; 2) o número de píxeis de pele e 3) a média da intensidade dos píxeis no espaço de cor RGB, resultando em 12 características. Essa metodologia de extração teve como embasamento maior o trabalho de Ap-Apid [Ap-Apid 2005], que utilizou regras estáticas baseadas principalmente nas maiores regiões de pele segmentadas, para definir se determinada imagem possuía nudez ou não.

4.1.2. Detecção de Face

Devido a existência de uma grande quantidade de píxeis de pele nas imagens de nudez e de retrato, as mesmas são facilmente confundidas pelos classificadores baseados estritamente em píxeis de pele, ocasionando um alto número de falsos positivos. Em seus trabalhos, Platzer et al. [Platzer et al. 2014] e Zhou et al. [Zhou et al. 2016] aplicaram regras estáticas baseados na razão entre os píxeis de pele existentes nas faces e em toda a imagem, na tentativa de eliminar esses falsos positivos.

Ao invés de utilizar regras estáticas, nosso trabalho propõe o uso desses dois valores como características: 1) o número de faces detectadas e 2) o somatório das regiões das faces detectadas. Dessa forma, objetiva-se aumentar a acurácia do modelo por meio da redução dos falsos positivos.

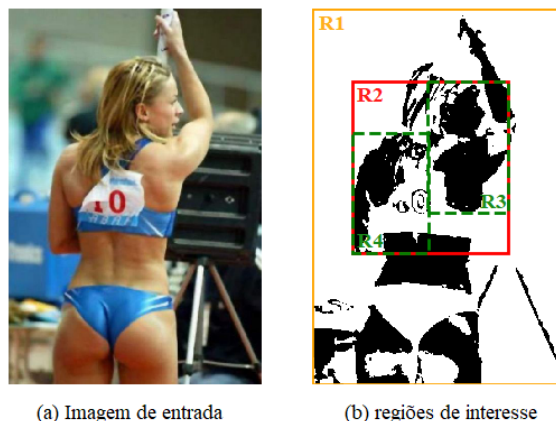


Figura 2. A imagem de entrada antes da extração das características de pele em (a). Imagem de entrada após a detecção dos píxeis de pele, estes representados pela cor preta e as quatro regiões de interesse para extração de características: R1, toda a imagem. R2, o menor retângulo que englobe as duas maiores regiões de pele. R3 e R4, as duas maiores regiões de pele em (b).

Foi utilizado o classificador baseado em características *Haar* proposto por Viola e Jones [Viola and Jones 2004], pelas mesmas razões da escolha do detector de pele. É recomendado para aplicações que demandam alta velocidade, devido a grande quantidade de imagens que são analisadas nos exames forenses [Eleutério and Machado 2011].

Sendo assim, esta etapa resultou em um vetor de características de tamanho 14. As Figuras 3 e 4 mostram exemplos das regiões de interesse da extração de características em uma imagem pornográfica e uma não pornográfica, respectivamente.

4.2. Classificação

Além de criar os padrões de cada categoria na fase de treinamento (exceto quando trata-se de um classificador baseado em regras que não dispõe dessa etapa), o classificador também assume o papel de categorizar as imagens em classes previamente determinadas (i.e. imagem com nudez, imagem sem nudez). Atualmente os algoritmos baseados em aprendizagem de máquina são amplamente utilizados para essa função [Platzer et al. 2014, Zhou et al. 2016]. Com o objetivos de alcançar os melhores resultados, fizemos uso de quatro tipos de classificadores dessa natureza.

O classificador baseado em **regressão logística** não é capaz de discriminar dados não separáveis linearmente, pois cria um modelo que assume apenas uma relação linear entre os dados de entrada e de saída. Entretanto é possível modificar os valores de entrada para que possa haver essa separação, por meio do uso da engenharia de recursos (e.g. multiplicação e potenciação dos valores). Apesar de possuir ótimo tempo de resposta, não apresenta bons resultados com dados de entrada muito complexos [Hosmer and Lemeshow 2000].

Já o classificador baseado em uma rede neural **perceptron multicamadas** consegue naturalmente distinguir dados não linearmente separáveis, devido a existência de



Figura 3. Imagem de nudez em (a) e suas regiões de interesse para extração de características em (b) e (c).

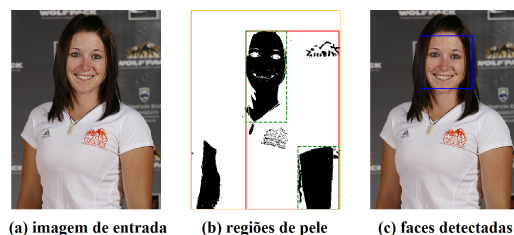


Figura 4. Imagem de não nudez em (a) e suas regiões de interesse para extração de características em (b) e (c).

múltiplas camadas e de funções de ativação não-lineares. Essas redes neurais são compostas por três ou mais camadas de neurônios conectados por pesos entre cada camada adjacente. Exceto nos nodos de entrada, cada neurônio alimenta (*feedforward*) a rede neural por meio de uma função de ativação não-linear (e.g. *Sigmoid*, *Tanh*, *ReLU*). Por fim, para otimizar os pesos em cada iteração da fase de treinamento, a rede utiliza uma técnica chamada *backpropagation* [Haykin 2009].

Os classificadores baseados em árvores utilizam regras em cada nodo não-folha. A classificação ocorre quando o processo atinge o nodo folha, este possuindo uma determinada classe [Breiman 2001]. O baseado em **árvore de decisão** é utilizado em diversas aplicações de visão computacional (e.g. detecção de face, análise de gestos manuais) e utiliza apenas uma árvore para realizar a classificação dos dados. Já a **floresta aleatória** utiliza de maneira simultânea um conjunto de árvores de decisão por meio do uso da técnica de *bootstrap aggregation*, obtendo uma acurácia alta e uma rápida fase de treinamento, entretanto, não é tão rápida quanto uma única árvore de decisão, por motivos óbvios [Oghaz et al. 2015].

5. Experimentos e Resultados

Todos os detalhes experimentais e seus resultados são discriminados nessa seção. A subseção 5.1 descreve como foram conduzidos nossos experimentos. Os resultados obtidos e a performance dos métodos propostos, além da sua análise estatística, são mostrados na subseção 5.2.

5.1. Experimentos

Visando avaliar a acurácia dos modelos propostos, foi utilizada a base de dados de imagens AIIA-PID4 pornographic data set [Karavarsamis et al. 2013]. Esta base de dados contém 12.740 imagens, divididas em 4 categorias: 1) pornográficas; 2) biquíni; 3) pele e 4) não pele. Baseado no trabalho de Karavarsamis et al. [Karavarsamis et al. 2013], foram criadas duas novas categorias: 1) Nudez: por meio do agrupamento das categorias pornográfica e biquíni e 2) Não nudez: por meio do agrupamento das categorias pele e não pele. As categorias Nudez e Não nudez detinham 6.642 e 6.128 imagens, respectivamente. Contudo foram removidas imagens que apresentaram problemas de leitura, sendo utilizadas 6.638 da classe Nudez e 6.102 da classe Não nudez. Os dados apresentavam-se relativamente balanceados, respectivamente apresentando 52,10% e 47,90% do montante total.

Os experimentos foram conduzidos com o objetivo de sobre-ajustar (*overfit*) cada modelo descrito na subseção 4.2 e, então, realizar o ajuste fino (*fine-tune*) dos hiperparâmetros com poder de regularização por meio do uso da técnica de validação cruzada, visando reduzir o erro de generalização. Esse processo acontece após a extração das características, descrito na subseção 4.1. Foi utilizada uma validação cruzada 5-fold em 90% da base de dados selecionada e foi reservado 10% para a fase de testes que acontece no final do experimento. Foi utilizada primordialmente a métrica de acurácia para avaliação dos modelos, visto que os dados são relativamente balanceados, entretanto, outras métricas amplamente utilizadas para dados desbalanceados foram utilizadas para aferir o desempenho dos modelos (i.e. F1-score, Precisão, Revocação).

Todos os classificadores foram utilizados com a configuração padrão da biblioteca Python Sci-kit Learn library². Visando realizar o ajuste fino dos hiperparâmetros de regularização, foi selecionado um conjunto de valores para cada classificador.

Tanto na regularização logística, quanto na rede neural perceptron multicamada, foram utilizados os valores empíricos em (3) para seus respectivos fatores de regularização. Na regressão logística há um regularizador C inversamente proporcional à regularização. Em suma, quanto menor for o regularizador, mais regularizado o classificador será. Já na rede neural perceptron multicamadas o fator de regularização é denominado $alpha$ e o mesmo possui comportamento proporcional à regularização.

$$C|\alpha = [0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000, 10000, 100000] \quad (3)$$

Nos classificadores baseados em árvores foi utilizado como fator de regularização o número mínimo de divisões em um nodo interno. Quanto maior for esta divisão, mais regularizado será o classificador. Empiricamente foram utilizados os valores em (4).

$$min_split_samples = [2, 4, 8, 16, 32, 64, 128, 256, 512, 1024] \quad (4)$$

Por fim, foram comparados os resultados do melhor classificador adotado e os trabalhos de Medina e Palladino [Medina and Palladino 2017] e Karavarsamis et al. [Karavarsamis et al. 2013]. Todos os experimentos foram realizados utilizando uma estação de trabalho HP Z820 equipada com um processador Intel Intel® Xeon® E5-2630 2.6 GHz (32 GB RAM), sistema operacional Windows 7 Home (64 bits). Os algoritmos foram implementados na linguagem de programação Python utilizando a biblioteca Sci-Kit Learn².

5.2. Resultados

Nos experimentos de validação cruzada foram computadas as acurácias de cada classificador, variando os valores pré-determinados dos hiperparâmetros de regularização. Essas métricas podem ser visualizadas nas Tabelas 1 e 2, assim como por meio das curvas das fases de treinamento e validação nos gráficos mostrados na Figura 5. Em seguida, foi selecionado em cada classificador o valor do fator de regularização que apresentou maior acurácia na fase de testes. A regressão logística apresentou melhor resultado utilizando seu fator de regularização com valor igual a 10. Já o perceptron multicamadas obteve melhor resultado quando seu fator de regularização possuía valor igual a 0,1. A árvore de decisão e a floresta aleatória apresentaram os melhores resultados quando seu fator de regularização foi igual a 16.

Na fase de teste também foram computadas as métricas de Precisão, Revocação (*Recall*) e F1-Score dos nossos classificadores para fins de comparação entre eles, como pode ser visto na Tabela 3.

²<http://scikit-learn.org>

Tabela 1. Acurácias resultantes da variação dos respectivos fatores de regularização (C e alpha) nas validações cruzadas da Regressão Logística e Perceptron Multicamada.

Classificador	C — alpha									
	0,0001	0,001	0,01	0,1	1	10	100	1000	10000	100000
REGRESSÃO LOGÍSTICA	89,24%	89,25%	89,34%	89,39%	89,37%	89,51%	89,26%	89,49%	89,23%	89,27%
PERCEPTRON MULTICAMADAS	85,34%	87,34%	87,28%	87,76%	86,31%	85,91%	87,35%	85,31%	81,80%	78,69%

Tabela 2. Acurácias resultantes da variação do fator de regularização (min_split_samples) nas validações cruzadas da Árvore de Decisão e Floresta Aleatória.

Classificador \ min_split_samples	2	4	8	16	32	64	128	256	512	1024
ÁRVORE DE DECISÃO	92,77%	92,77%	92,90%	93,14%	92,99%	92,92%	92,80%	92,35%	91,30%	89,43%
FLORESTA ALEATÓRIA	92,87%	93,11%	93,35%	93,45%	93,01%	92,94%	92,94%	91,87%	90,39%	89,45%

Tabela 3. Resultados nas métricas de precisão, revocação, F1-score e acurácia de cada classificador fatores de regularização mais bem ajustados.

Classificador \ Métrica	Precisão	Revocação	F1-score	Acurácia
Regressão Logística	93,45	86,26	89,71	88,70%
Perceptron Multicamadas	94,64	87,00	90,66	89,72%
Árvore de Decisão	93,15	93,29	93,22	92,86%
Floresta Aleatória	94,94	93,00	93,96	93,56%

Dessa forma, os classificadores baseados em árvores demonstraram os melhores resultados. A abordagem baseada em floresta aleatória superou a baseada em árvore de decisão, tendo apresentado 93,56% e 92,86% de acurácia, respectivamente. Os classificadores baseados em regressão logística e perceptron multicamadas apresentaram resultados menos favoráveis comparados aos demais, alcançando, de maneira respectiva, 87,70% e 89,72% na mesma métrica.

O trabalho de Medina e Palladino [Medina and Palladino 2017] foi reproduzido utilizando as mesmas imagens usadas na realização da nossa etapa de testes. Os resultados da pesquisa de Karavarsamis et al. [Karavarsamis et al. 2013] foram obtidos diretamente pelo seu artigo, no qual cita apenas a acurácia atingida na fase de validação, sem referenciar quais imagens foram utilizadas. Todos os modelos propostos superaram os referidos trabalhos. A Tabela 4 mostra um comparativo entre a acurácia do modelo proposto baseado em floresta aleatória, que atingiu os melhores resultados, e os trabalhos de Medina e Palladino [Medina and Palladino 2017] e Karavarsamis et al. [Karavarsamis et al. 2013], que atingiram as acurácias de 79,10% e 85,05%, respectivamente.

6. Considerações Finais

Este trabalho propôs a melhoria de uma consagrada técnica de detecção de nudez desenvolvida por Ap-Apid [Ap-Apid 2005] (“Um Algoritmo Para a Detecção de Nudez”). Ao invés de utilizar regras estáticas para classificação das imagens, foram utilizados modelos baseados em aprendizagem de máquina. O conjunto de características foi baseado em dados oriundos de detecção de faces, com o objetivo de minimizar falsos positivos em

Tabela 4. Comparação da acurácia dos métodos propostos e os trabalhos de Medina e Palladino [Medina and Palladino 2017] e Karavarsamis et al. [Karavarsamis et al. 2013].

Método	Acurácia
Floresta Aleatória	93,56%
Medina e Palladino [Medina and Palladino 2017]	79,10%
Karavarsamis et al. [Karavarsamis et al. 2013]	85,05%

retratos, e de detecção de pele, muitas dessas baseadas nesse mesmo estudo. Por fim, esses modelos foram treinados utilizando a técnica de validação cruzada, com o objetivo de diferenciar entre imagens contendo nudez e não nudez.

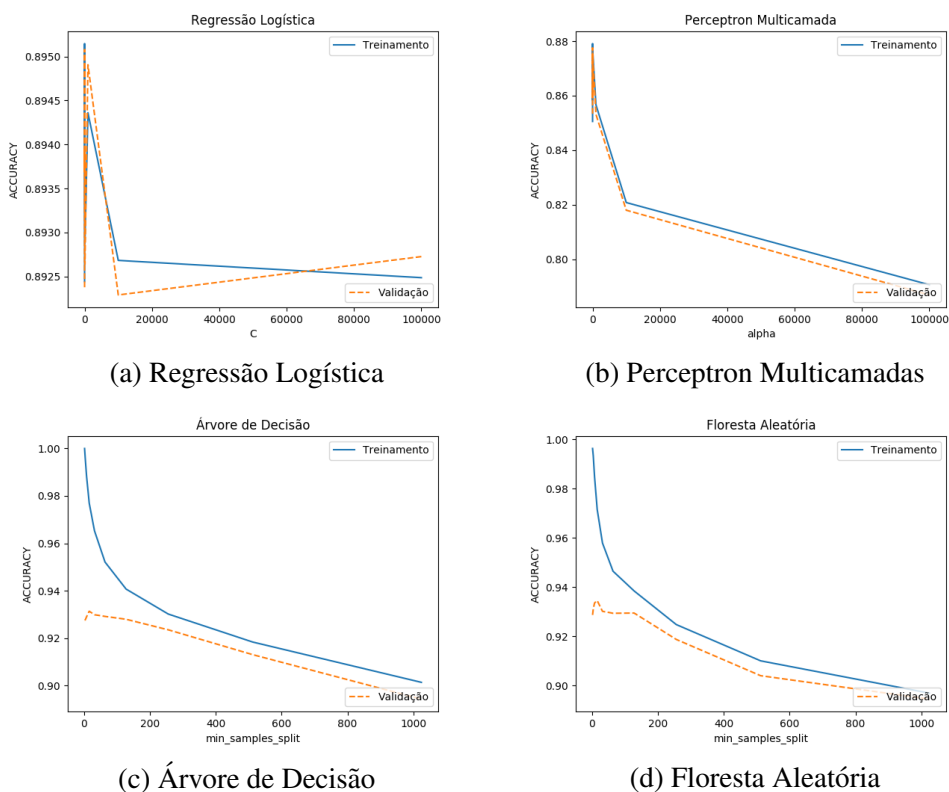
Os melhores resultados foram obtidos pelo classificador baseado em Floresta Aleatória, que atingiu 93,56% de acurácia e superando o trabalho de Medina e Palladino [Medina and Palladino 2017] em 18,28%, este tendo apresentado acurácia de 79,10%. Também foram atingidos melhores resultados em comparação ao trabalho de Karavarsamis et al. [Karavarsamis et al. 2013], criador da base de dados utilizada (*AIIA-PID4 pornographic data set*), que mostrou uma acurácia de 85,05%, sendo sido superado por nossa técnica em 10,01%.

Para trabalhos futuros, sugerimos utilizar esse modelo como uma etapa inicial de um processo mais sofisticado, utilizando-o para descartar imagens de não nudez com a maior precisão possível, repassando para uma segunda etapa baseada em Redes Neurais Convolucionais Profundas a responsabilidade de categorizar as imagens mais difíceis de serem classificadas.

Agradecimentos

Os autores gostariam de agradecer ao Laboratório de Inteligência Artificial por permitir e apoiar a implementação desse trabalho nas suas dependências, assim como à Universidade Federal de Campina Grande e ao Instituto de Polícia Científica da Paraíba.

Figura 5. Acurácias no treinamento e teste, variando os fatores de regularização dos classificadores na validação cruzada.



Referências

- Ap-Apid, R. (2005). An algorithm for nudity detection. In *Proceedings of the 5th Philippine Computing Science Congress*, Cebu City, Philippines.
- Basilio, J. A. M., Torres, G. A., Pérez, G. S., Medina, L. K. T., and Meana, H. M. P. (2011). Explicit image detection using ycbcr space color model as skin detection. In *Proceedings of the 2011 American Conference on Applied Mathematics and the 5th WSEAS International Conference on Computer Engineering and Applications*, AMERICAN-MATH'11/CEA'11, pages 123–128, Stevens Point, Wisconsin, USA. World Scientific and Engineering Academy and Society (WSEAS).
- Breiman, L. (2001). Random forests. *Mach. Learn.*, 45(1):5–32.
- Cappellari, M. S. V. (2005). A pedofilia na pós-modernidade: um problema que ultrapassa a cibercultura. *Em Questão*, 11:67–82.
- Eleutério, P. M. D. S. and Machado, M. P. (2011). *Desvendando a computação forense*. Novatec, São Paulo, 1 edition.
- Haykin, S. S. (2009). *Neural networks and learning machines*. Pearson Education, Upper Saddle River, NJ, third edition.
- Hosmer, D. W. and Lemeshow, S. (2000). *Applied logistic regression*. John Wiley and Sons.
- Karavarsamis, S., Ntarmos, N., Blekas, K., and Pitas, I. (2013). Detecting pornographic images by localizing skin rois. *International Journal of Digital Crime and Forensics (IJDCF)*, 5:39–53.
- Khan, R., Hanbury, A., Stöttinger, J., and Bais, A. (2012). Color based skin classification. *Pattern Recognition Letters*, 33(2):157 – 163.
- Kovac, J., Peer, P., and Solina, F. (2003). Human skin color clustering for face detection. In *The IEEE Region 8 EUROCON 2003. Computer as a Tool.*, volume 2, pages 144–148 vol.2.
- Lucena, O., Oliveira, . D. P., Veloso, L., and Pereira, E. (2017). Improving face detection performance by skin detection post-processing. In *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 300–307.
- Ma, B., Zhang, C., Chen, J., Qu, R., Xiao, J., and Cao, X. (2014). Human skin detection via semantic constraint. In *Proceedings of International Conference on Internet Multimedia Computing and Service*, ICIMCS '14, pages 181:181–181:184, New York, NY, USA. ACM.
- Mahmoodi, M. R. and Sayedi, S. (2016). A comprehensive survey on human skin detection. *International Journal of Image, Graphics and Signal Processing*, 8:1–35.
- Medina, M. R. and Palladino, P. (2017). Pornographic images jacking algorithm. <https://github.com/alcuadrado/pija>. Accessed: November 17, 2017.
- Muhammad, B. and Abu-Bakar, S. A. R. (2015). A hybrid skin color detection using hsv and ycgcr color space for face detection. In *2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pages 95–98.

- Oghaz, M. M., Maarof, M. A., Zainal, A., Rohani, M. F., and Yaghoubyan, S. H. (2015). A hybrid color space for skin detection using genetic algorithm heuristic search and principal component analysis technique. *PLOS ONE*, 10(8):1–21.
- Patil, H. Y., Bharambe, S. V., Kothari, A. G., and Bhurchandi, K. M. (2013). Face localization and its implementation on embedded platform. In *2013 3rd IEEE International Advance Computing Conference (IACC)*, pages 741–745.
- Platzer, C., Stuetz, M., and Lindorfer, M. (2014). Skin sheriff: A machine learning solution for detecting explicit images. In *Proceedings of the 2Nd International Workshop on Security and Forensics in Communication Systems, SFCS '14*, pages 45–56, New York, NY, USA. ACM.
- Polastro, M. d. C. and Eleutério, P. M. d. S. (2010). Nudetective: A forensic tool to help combat child pornography through automatic nudity detection. In *2010 Workshops on Database and Expert Systems Applications*, pages 349–353.
- Putro, M. D., Adji, T. B., and Winduratna, B. (2015). Adult image classifiers based on face detection using viola-jones method. In *2015 1st International Conference on Wireless and Telematics (ICWT)*, pages 1–6.
- Vezhnevets, V., Sazonov, V., and Andreeva, A. (2003). A survey on pixel-based skin color detection techniques. In *IN PROC. GRAPHICON-2003*, pages 85–92.
- Viola, P. and Jones, M. J. (2004). Robust real-time face detection. *Int. J. Comput. Vision*, 57(2):137–154.
- Xiong, W. and Li, Q. (2012). Chinese skin detection in different color spaces. In *2012 International Conference on Wireless Communications and Signal Processing (WCSP)*, pages 1–5.
- Yang, J., Shi, Y., and Xiao, M. (2007). Geometric feature-based skin image classification. In Huang, D.-S., Heutte, L., and Loog, M., editors, *Advanced Intelligent Computing Theories and Applications. With Aspects of Theoretical and Methodological Issues*, pages 1158–1169, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Youtian, D., Zhongmin, C., Xiaohong, G., and Qian, L. (2012). Almost optimal skin detection approach within the gaussian framework. *Optical Engineering*, 51:51 – 51 – 10.
- Zafeiriou, S., Zhang, C., and Zhang, Z. (2015). A survey on face detection in the wild: Past, present and future. *Computer Vision and Image Understanding*, 138:1 – 24.
- Zhou, K., Zhuo, L., Geng, Z., Zhang, J., and Li, X. G. (2016). Convolutional neural networks based pornographic image classification. In *2016 IEEE Second International Conference on Multimedia Big Data (BigMM)*, pages 206–209.