

**RAFAEL ESTÉFANO VICENTINI**

**USO DE REDES NEURAS ARTIFICIAIS PARA DETECÇÃO DE PELE EM  
IMAGENS DIGITAIS**

Dissertação apresentada à Faculdade de Engenharia do  
Campus de Ilha Solteira – UNESP como parte dos  
requisitos para obtenção do título de Mestre em  
Engenharia Elétrica. Especialidade: Sistemas  
Inteligentes.

**Profa. Dra. Anna Diva Plasencia Lotufo**

Orientadora

Ilha Solteira

2017

FICHA CATALOGRÁFICA

Desenvolvido pelo Serviço Técnico de Biblioteca e Documentação

V633u      Vicentini, Rafael Estéfano.  
            Uso de redes neurais artificiais para detecção de pele em imagens digitais /  
            Rafael Estéfano Vicentini. -- Ilha Solteira: [s.n.], 2017  
            79 f. : il.

            Dissertação (mestrado) - Universidade Estadual Paulista. Faculdade de  
            Engenharia de Ilha Solteira. Área de conhecimento: Automação, 2017

            Orientador: Anna Diva Plasencia Lotufo  
            Inclui bibliografia

            1. Redes neurais artificiais. 2. Processamento de imagens. 3.  
            Reconhecimento de padrões. 4. Resilient backpropagation.



UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Ilha Solteira

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: Uso de redes neurais artificiais para detecção de pele em imagens

AUTOR: RAFAEL ESTÉFANO VICENTINI

ORIENTADORA: ANNA DIVA PLASENCIA LOTUFO

Aprovado como parte das exigências para obtenção do Título de Mestre em ENGENHARIA ELÉTRICA, área: AUTOMAÇÃO pela Comissão Examinadora:

Profa. Dra. ANNA DIVA PLASENCIA LOTUFO

Departamento de Engenharia Elétrica / Faculdade de Engenharia de Ilha Solteira

Profa. Dra. MARA LUCIA MARTINS LOPES

Departamento de Matemática / Faculdade de Engenharia de Ilha Solteira

Prof. Dr. KENJI NOZÊ FILHO

Curso de Engenharia da Informação / Universidade Federal do ABC

Ilha Solteira, 20 de outubro de 2017

## RESUMO

Na última década, o aumento da capacidade de processamento de informação em computadores e dispositivos de uso pessoal possibilitou o desenvolvimento de filtros e classificadores automatizados que operam em tempo real, aplicados em diversas áreas. No âmbito do Processamento Digital de Imagens associado às Redes Neurais Artificiais, os filtros emulam a percepção humana buscando por padrões para identificação de características de interesse.

Filtros que têm por objetivo restringir o acesso a conteúdo impróprio partem da identificação de pele - principal indício de presença humana em uma imagem. Independentemente de sua complexidade e/ou robustez, caso o classificador não seja capaz de identificar as diferentes tonalidades de pele sob diferentes condições de captura, sua eficácia é prejudicada. Frente à diversificada forma de descrever uma tonalidade de pele usando diferentes espaços de cor, neste estudo foram destacados os espaços de cor RGB, YCbCr e HSV, amplamente utilizados em equipamentos de captura (por exemplo câmeras fotográficas e filmadoras digitais).

A partir de exemplos apresentados durante a etapa de treinamento, as RNAs devem estar aptas para classificar as tonalidades em dois grupos distintos: *pele* e *não pele*.

Dentre os espaços de cores indicados, seja utilizando ou descartando o aspecto da iluminação (critério amplamente discutido na literatura), este trabalho busca avaliar qual possui a maior taxa de detecção de pele em uma imagem.

**Palavras-Chave:** Redes neurais artificiais. Reconhecimento de padrões. Resilient backpropagation. Processamento digital de imagens.



## ABSTRACT

Over the last decade, the increasing capacity of data processing in personal computers and devices could develop filters and automatic classifiers working in real time and applied in several areas. Considering Digital Image Processing and Artificial Neural Networks, these filters emulate the human perception searching for patterns to identify specific features.

Filters which the main goal is to restrict the access to inappropriate content starts identifying skin tones - the main evidence of human presence in a picture. Although being complex and robust, if the classifier is not able to identify distinct skin tones under random capture conditions, the accuracy is minimal. Facing several ways on describing skin tones over different color spaces, this work uses the RGB, YCbCr and HSV color spaces which are widely applied in recording devices (photographic and digital cameras for example).

Based on the examples shown during the training phase, the ANNs must be able to classify skin tones into two distinct groups: *skin* and *non skin*.

Among the different color spaces used, considering or not the luminance aspect (widely discussed on papers), this work intends to evaluate which one has the highest detection accuracy to identify skin tone in such a picture.

**Keywords:** Artificial neural networks. Pattern recognition. Resilient backpropagation. Digital image processing.

## LISTA DE SÍMBOLOS

$rgb \vee \text{skin}$	Pixel pertencente à classe <i>pele</i>
$rgb \vee \neg \text{skin}$	Pixel pertencente à classe <i>não pele</i>
$\text{Hist}_{\text{skin}}$	Histograma correspondente à região de <i>pele</i>
$\text{Hist}_{\text{non-skin}}$	Histograma correspondente à região <i>não pele</i>
$\Theta$	Limiar definido empiricamente no SPM - <i>Skin Probability Maps</i> , para definir se o pixel possui tonalidade de pele
C1 e C2	Saídas da Rede Neural Artificial criada por Bohyar
$\tau$	Taxa de Aprendizagem
$w$	Peso (parâmetro livre) ajustável da Rede Neural Artificial
$\frac{\partial E}{\partial w}$	Derivada Parcial
$\Delta w$	Tamanho do ajuste do peso $w$
$\Delta$	Valor de Atualização do peso $w$
$\eta^+$	Variável utilizada para incrementar o ajuste de pesos no <i>Resilient Backpropagation</i>
$\eta^-$	Variável utilizada para minimizar o ajuste de pesos <i>Resilient Backpropagation</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>09</b>
1.1	Motivação e Trabalhos Relacionados.....	12
1.2	Objetivos e Organização do Trabalho.....	17
<b>2</b>	<b>ESTADO DA ARTE.....</b>	<b>18</b>
2.1	Imagem Digital.....	18
2.2	Processamento Digital de Imagens.....	20
2.3	Redes Neurais Artificiais.....	22
2.3.1	<i>Modelo de um neurônio.....</i>	<i>24</i>
2.3.2	<i>Topologia de Redes Neurais Artificiais.....</i>	<i>24</i>
2.3.3	<i>Processo de aprendizagem.....</i>	<i>25</i>
2.3.4	<i>RNA usada no reconhecimento de padrões.....</i>	<i>26</i>
2.4	Resumo e discussão.....	27
<b>3</b>	<b>ESPAÇOS DE COR.....</b>	<b>29</b>
3.1	Espaço De Cor Rgb.....	31
3.1.1	<i>Espaço de cor RGB normalizado.....</i>	<i>32</i>
3.2	Espaço de Cor YCbCr.....	33
3.3	Espaço de Cor HSV.....	34
3.4	Resumo e discussão.....	35
<b>4</b>	<b>MÉTODO PROPOSTO E RESULTADOS OBTIDOS.....</b>	<b>37</b>
4.1	Desenvolvimento do Classificador Neural.....	37
4.1.1	<i>Coleta De Dados.....</i>	<i>37</i>
4.1.2	<i>Desenvolvimento da Redes Neurais Artificiais.....</i>	<i>39</i>
4.1.3	<i>Treinamento das Redes Neurais Artificiais.....</i>	<i>41</i>

4.1.3.1	<i>Resilient Backpropagation</i> .....	41
4.1.3.2	<i>Critérios de parada do treino</i> .....	45
4.1.4	<i>Uso da Rede Neural Artificial</i> .....	53
4.1.4.1	<i>Máscara Binária Produzida Pelas RNAs</i> .....	55
4.2	<b>Estudo de Caso: Classificador de nudez para selfies</b> .....	66
5	<b>CONCLUSÃO</b> .....	71
5.1	<b>Observações e tópicos para próximas pesquisas</b> .....	73
	<b>REFERÊNCIAS</b> .....	75

## 1 INTRODUÇÃO

A popularização das mídias digitais e o aumento do número de pessoas que acessam a *world wide web* proporcionaram, na última década, a mudança por completo no modo de como acessamos a informação. Expressando em números, há mais de um bilhão de *websites* ativos (abril de 2017) e este número continua a crescer, segundo o *Internet live stats* (<http://www.internetlivestats.com/total-number-of-websites>). Dada a significativa quantidade de informação disponível na *web*, é humanamente impossível o controle manual para restringir o acesso caso tal informação seja ofensiva às crianças. Santos (2012) relaciona o conteúdo ofensivo à nudez humana cuja utilização está, na maioria das vezes, associada ao conteúdo sexualizado.

Aplicações automáticas têm sido desenvolvidas buscando a classificação da informação. Muitas delas utilizam recursos como mineração de dados (*datamining*), processamento digital de imagens (*digital image processing – DIP*) e aprendizagem de máquina em conjunto para obtenção do resultado desejado.

Belém (2006) distingue os métodos de filtragem automática em três categorias: reconhecimento baseado em texto, reconhecimento baseado em imagem e reconhecimento baseado na combinação de texto e imagem.

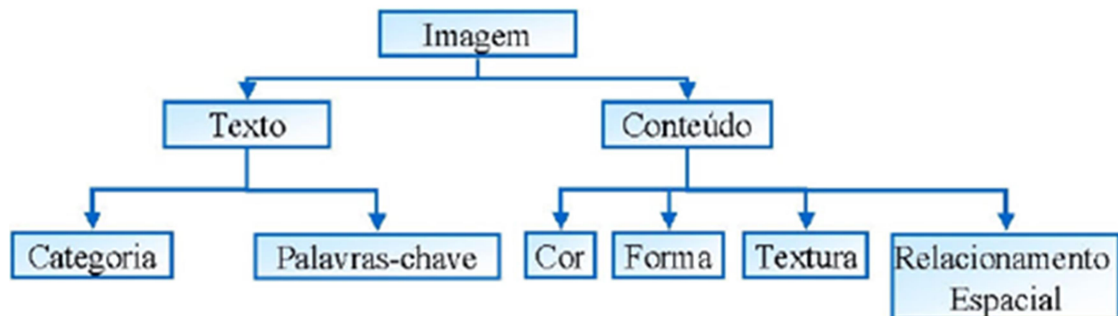
Métodos do primeiro grupo baseiam-se em uma *blacklist* que contém expressões comumente utilizadas em *websites* com conteúdo impróprio. Neste método cada expressão recebe uma pontuação (*score*). Ao acessar uma página na *internet*, antes que seu conteúdo seja exibido ao usuário final, a busca por estas expressões é realizada e caso sejam encontradas com elevada frequência fazendo com que a página receba uma pontuação alta, seu acesso é negado. Dentre as aplicações que utilizam esta técnica, uma das mais conhecidas é o *SQUID* <<http://www.squid-cache.org/>>, que age como uma camada intermediária entre o usuário e a *web*. Ele recebe as solicitações de acesso e verifica a informação contida na página *web* antes de retorná-la ao usuário com baixo custo de tempo e processamento. Contudo as páginas *web* são compostas por textos e imagens e, neste processo de busca por expressões, somente o nome do arquivo correspondente à imagem é avaliado. Caso uma imagem com conteúdo ofensivo possua o nome '*foto1.jpeg*' e a expressão *foto*, por não se tratar de uma expressão ofensiva não estando listada na *blacklist*, o *SQUID* autorizará seu acesso sem qualquer outro processo de análise.

No segundo grupo, métodos de filtragem baseados em imagens buscam extrair e identificar a informação contida em uma imagem digital. A Figura 1 mostra diferentes abordagens para representação de tal informação, divididas em dois principais subgrupos: texto e conteúdo (ERPEN, 2004).

Na representação textual é gerado um banco de dados de palavras que descreve o conteúdo da imagem. Porém, como qualquer objeto ou cenário pode ser descrito de inúmeras formas, este tipo de interpretação possui baixa eficácia de identificação por conta da subjetividade do processo. Erpen (2004) destaca que outro aspecto desfavorável está no fato de que a aplicação que está recuperando a informação não é a mesma que realizou o armazenamento. Ou seja, caso a representação da imagem seja armazenada contendo a descrição dos objetos e no futuro for realizada uma busca pelo contexto da imagem (por exemplo, o pôr do sol), mesmo que ela contenha os elementos da busca, há uma grande possibilidade de que não sejam corretamente identificados e a imagem classificada erroneamente.

A descrição da imagem através do seu conteúdo utiliza de técnicas de processamento digital de imagens para processar a informação armazenada.

Figura 1 – Métodos de descrição de uma imagem.



Fonte: Erpen (2004)

Diferente da percepção humana em que as formas, cores e texturas são distinguidas automaticamente com base no conhecimento, o reconhecimento através de meios computacionais está associado à classificação de padrões. Haykin (1999) define a classificação de padrões como “o processo em que dado um padrão/sinal recebido, de acordo com suas características, este é atribuído a uma específica classe dentro de um número predeterminado de classes”.

Em uma imagem digital atribui-se a uma mesma classe regiões que possuam

características homogêneas. A homogeneidade pode ser caracterizada pela cor, profundidade da camada, níveis de cinza, textura, etc. (ALMOHAIR et al., 2007). A extração dessas regiões é realizada no processo de segmentação que divide a imagem em regiões não sobrepostas constituídas por grupos de pixels homogêneos. O processo de segmentação é exemplificado na Figura 2 (SENA JR, 2002). A imagem original de um recipiente com uma planta (esquerda) e a segmentação (direita) obtida através da definição de duas classes: *vegetal* e *não vegetal*. Inicialmente são identificadas regiões que contém as características pertinentes à primeira classe e as rotula pela cor branca. Caso contrário, é atribuída para a segunda classe e rotulada pela cor preta.

Figura 2 - Segmentação dos corpos vegetais. Imagem original (esquerda) e segmentação da vegetação (direita).



Fonte: Sena Jr (2002)

Aplicações que buscam filtrar imagens com conteúdo ofensivo partem do pressuposto de que as imagens que contém nudez apresentam elevada porcentagem de tons de pele (JONES; REHG, 2002 citados por SANDOVAL, 2014).

A identificação de pixels com cor de pele é um forte indício de presença humana em uma imagem. A partir da segmentação dos pixels de pele é possível a identificação de características (padrões) comuns em imagens com teor sexual. Pesquisas têm sido amplamente realizadas utilizando a detecção de pixels de pele como o primeiro passo para reconhecimento de faces (THAKUR et al., 2012), (GOUVEIA, 2010), análise de gestos (TERRILLON et al., 2000), sistemas de segurança (YANG et al., 2010), e filtro de conteúdo pornográfico (BELÉM, 2006), (JONES; REHG, 1998), (JEONG; KIM; HONG, 2004), (SANTOS, 2012).

## 1.1 Motivação e Trabalhos Relacionados

O desenvolvimento de um método para classificação de pixels de pele possui duas etapas: treinamento e detecção. Durante o treinamento são cometidos os seguintes passos (ELGAMMAL; MUANG; HU, 2009):

1. Coleta de dados com exemplos de pele em diferentes imagens. Tais imagens devem conter diversas tonalidades de pele sob variadas condições de iluminação.
2. Escolha do método de representação das cores (espaço de cor).
3. Definição de parâmetros que determinam a distinção entre as classes de *pele* e *não pele*.

Na segunda etapa, em sua utilização, ao receber uma nova imagem, o classificador deve identificar e rotular cada pixel como pele ou não pele. Após a classificação, comumente são utilizadas práticas de pós-processamento para estabelecer a homogeneidade nas áreas detectadas.

Um dos maiores desafios no desenvolvimento de um classificador de pele está na escolha do espaço de cor (RGB, YCbCr, HSV, etc.) a ser utilizado dado que a cor da pele em uma imagem é sensível e variável diante de fatores como condições de iluminação, etnia e até mesmo a escolha do equipamento utilizado para a captura da imagem (SANDOVAL, 2014). Yang, Lu e Waibel (1998) atestaram em seus experimentos que a diferença da tonalidade na pele entre pessoas é maior na luminosidade se comparada com diferença na cromaticidade. Portanto, a escolha do espaço de cor que possibilita a separação entre estas duas características pode conduzir a melhores resultados. Os espaços de cor representam uma mesma imagem utilizando atributos variados que podem ser benéficos ou prejudiciais no processo de classificação do pixel de pele. Em seus trabalhos, cada autor justifica a escolha que obteve melhores resultados em sua aplicação, seja no espaço YCbCr (AL-MOHAIR; MOHAMAD-SALEH; SUANDI, 2012), RGB (JONES; REHG, 2002), (BHOYAR; KAKDE, 2010) e até mesmo a combinação de dois ou mais espaços de cores como RGB-H-CbCr (RAHMAN; SEE, 2006). Porém em nenhum dos estudos o espaço de cor escolhido é tratado como solução ótima para qualquer modelo de classificador.

Desde o surgimento de pesquisas voltadas para classificação de pixel de pele diversos métodos foram desenvolvidos. Inicialmente surgiram aplicações que utilizam de algoritmos



estáticos como o *classificador explícito de grupos de pele* que utiliza métodos de limite (*threshold*) para definir se o pixel contém pele. Ele define, dentro do espaço de cor escolhido, uma fronteira entre os grupos de pixels que pertencem a diferentes classes. Kovac, Peer e Solina (2003) apontaram as seguintes condições para classificar a cor como *pele* no espaço de cor RGB:

$$R > 45 \text{ e } G > 40 \text{ e } B > 20 \quad (1)$$

$$\max[\text{RGB}] - \min[\text{RGB}] > 15 \quad (2)$$

$$|R - G| > 15 \text{ e } R > G \text{ e } R > B \quad (3)$$

Aplicações desta natureza possuem baixa flexibilidade e resultam em pouca eficácia de detecção diante de imagens sob diferentes condições de iluminação (WADUD, 2008).

Outros trabalhos desenvolvem aplicações com classificadores não-paramétricos. Estes estimam a distribuição da cor de pele a partir de um histograma gerado através de vários exemplos de imagens.

O algoritmo SPM - *Skin Probability Maps* (JONES; REHG, 1998) (BRAND et al. 2000) utiliza dois histogramas 3D com a distribuição de cores no espaço RGB, sendo um para *pele* e outro para *não pele*. A partir destes histogramas a probabilidade de um pixel com valores RGB ser pele ou não pele é expressa respectivamente por

$$P(\text{rgb} \vee \text{skin}) = \frac{\text{Hist}_{\text{skin}}[\text{rgb}]}{\text{Total}_{\text{skin}}} \quad (4)$$

e

$$P(\text{rgb} \vee \neg\text{skin}) = \frac{\text{Hist}_{\text{non-skin}}[\text{rgb}]}{\text{Total}_{\text{non-skin}}} \quad (5)$$

Calculadas as probabilidades, o pixel será classificado como *pele* caso satisfaça o limiar

$$\frac{P(\text{rgb} \vee \text{skin})}{P(\text{rgb} \vee \neg \text{skin})} > \Theta \quad (6)$$

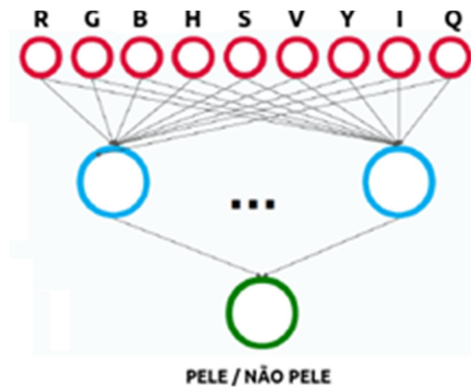
onde  $\Theta$  é obtido empiricamente.

Tais aplicações de natureza estatística possuem como vantagem a rapidez no treinamento e utilização porém apresentam baixa eficácia diante de variações, restringindo sua utilização somente para imagens que possuam condições de captura similares às imagens utilizadas para treinamento.

Com o aumento do poder de processamento computacional nos computadores atuais, classificadores adaptativos e dinâmicos têm sido a nova proposta. Este grupo de classificadores visam a otimização da performance com maiores taxas de detecção de pele e baixo índice de falso positivo (AL-MOHAIR; MOHAMAD-SALEH; SUANDI, 2012). A utilização de um modelo dinâmico possibilita a sua atualização constante para lidar diante das mudanças de condições (iluminação, espaço de cor, características da câmera).

A exemplo, a Figura 3 ilustra a Rede Neural Artificial (RNA) Perceptron Multicamadas empregada para classificação dos pixels de pele. A rede constituída por três camadas recebe como entrada as coordenadas do pixel em três espaços de cor: RGB, HSV e YIQ. Os vetores são passados adiante para a segunda camada que processa a informação e as classifica de acordo com o conhecimento adquirido com exemplos apresentados anteriormente durante o treinamento. A última camada, que contém um único neurônio, produz a saída classificando o pixel como pele ou não pele (BASTOS; ANGELO, 2014).

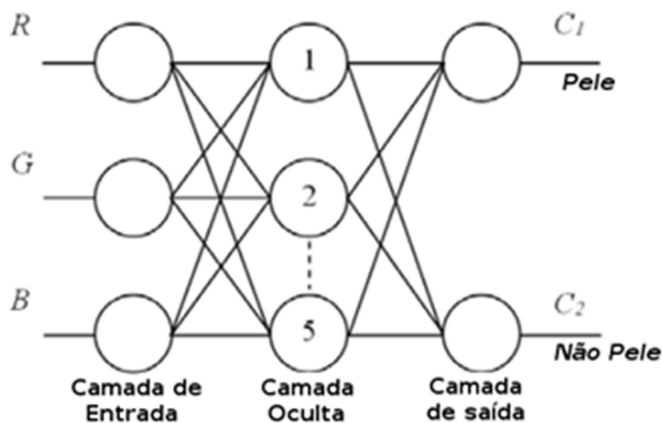
Figura 3 - Rede Neural Artificial Perceptron com três camadas desenvolvida para classificação de pixels de pele.



Fonte: Bastos e Angelo (2014)

Bhoyar e Kakde (2010) propuseram uma RNA que, ao invés de uma, produza duas saídas: o primeiro para representar a classe *pele* e o segundo, a classe *não pele*. A proposta da tal divisão aumenta a separabilidade entre as duas classes. A RNA (ilustrada na Figura 4) recebe como entrada os vetores no espaço de cor RGB que são processados nas camadas adiantes até que as saídas sejam produzidas.

Figura 4 - Rede Neural Artificial com duas saídas para aumentar a separabilidade entre classes.



Fonte: Bhoyar e Kakde (2010)

Obtidas as saídas, aplica-se o seguinte limiar:

Se,  $C1 - C2 \geq \Theta$  pixel *pele* (7)

Senão pixel *não pele*

onde  $C1$  e  $C2$  representam as classes pele e não pele respectivamente.

Os resultados são aceitáveis com taxas de erros em 7% quando o limiar  $\Theta$  é otimizado.

Métodos dispostos de RNAs utilizam um *dataset* com amostras de pixels *pele* e *não pele* como exemplos (BHATT et al., 2009). Este *dataset* recebe o nome de conjunto de treinamento. Enquanto aprende através destes exemplos, a RNA ajusta seus parâmetros na camada oculta mapeando uma relação entre as características de cada classe. Este processo de aprendizagem utiliza um algoritmo que objetiva a otimização do resultado. Dizemos que a RNA “aprendeu” quando ela for capaz de generalizar o resultado da classificação, ou seja, a partir de um pixel nunca apresentado antes, seja capaz de classificá-lo como pele ou não pele corretamente (HAYKIN, 1999).

O emprego de RNAs para detecção de pele tem como propósito superar as inconveniências presentes nos classificadores estáticos. O maior benefício de sua aplicação está em sua flexibilidade e habilidade de se adaptar às condições variadas (ELGAMMAL; MUANG; HU, 2009).

Sua utilização em conjunto de métodos estáticos (como o limiar aplicado por Bhoyar) podem maximizar a performance sem custo de processamento.

Em análise às mais recentes pesquisas que empregam RNAs para classificação de pixels de pele, Al-mohair, Mohamad-saleh e Suandi (2012) observaram que:

a) Todos os algoritmos utilizam alguma técnica limiar cuja definição é empírica ou dada após inúmeros experimentos de tentativas e erros. O tempo de processamento (característica importante para aplicações que trabalham em tempo real) não é mencionado em nenhum dos trabalhos;

b) A utilização de diferentes espaços de cor e *datasets* para avaliação dificultam a comparação justa entre os resultados dos métodos;

c) Todo método proposto deve ter como escopo a alta precisão em baixo custo de tempo, critérios essenciais para qualquer aplicação que trabalhe em tempo real.

## 1.2 Objetivos e Organização do Trabalho

O presente trabalho propõe o desenvolvimento de uma RNA *feedforward* de arquitetura multicamadas que utiliza o *Resilient Backpropagation* como algoritmo de aprendizagem e definir, dentre os espaços de cores descritos, qual espaço de cor apresenta maior precisão na detecção de pixels de pele que viabilize a extração de características a partir da máscara binária gerada pela segmentação de pixels de pele. Dados relevantes como a duração da etapa de treinamento e tempo de processamento para classificação serão apontados para referência para futuras pesquisas.

A organização do documento é expressa da seguinte forma:

O Capítulo 2 apresentará os conceitos básicos para o entendimento da pesquisa: definição de imagem digital, técnicas de processamento digital de imagens para segmentação e extração de características e Redes Neurais Artificiais aplicadas no reconhecimento de padrões.

No Capítulo 3 serão listados os espaços de cor que foram objetos de estudo nesta pesquisa: suas propriedades, aplicações e resultados obtidos em trabalhos anteriores.

O Capítulo 4 seguirá com a metodologia empregada: Aquisição do *dataset* para treinamento através de exemplos, desenvolvimento e utilização da rede neural.

Por fim, os resultados e conclusões serão dissertados no Capítulo 5.

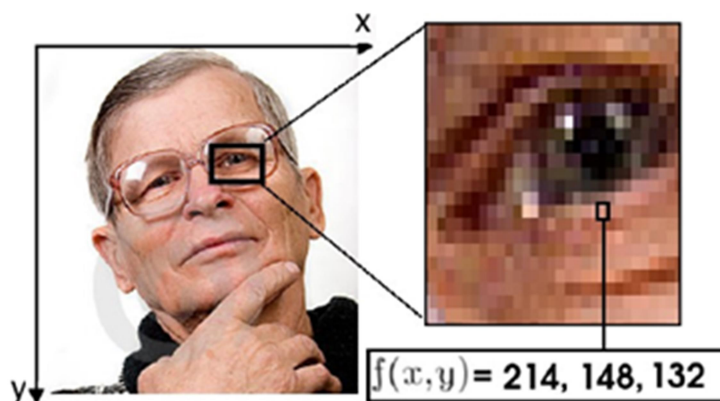
## 2 ESTADO DA ARTE

Ao longo deste Capítulo serão apresentados conceitos fundamentais para o entendimento deste trabalho. Inicialmente será discutido acerca da Imagem Digital e a manipulação da informação nela contida por meio do Processamento Digital de Imagens e a obtenção de dados pertinentes através da *segmentação*. Por fim, será expressa a definição de Redes Neurais Artificiais e sua aplicação na classificação de padrões e seu processo de aprendizagem por meio de exemplos (conjunto de treinamento).

### 2.1 Imagem Digital

A imagem digital pode ser representada como um plano bidimensional com finitas coordenadas  $x$  e  $y$  em que cada ponto da coordenada  $f(x,y)$  possui um valor relacionado à forma em que a imagem é descrita (GONZALEZ; WOODS, 2001). Toda imagem é composta por um número finito de elementos  $f(x,y)$  que possuem localização e valores específicos. Estes elementos recebem a denominação pixel. Quando exibida em uma tela onde muitos pixels estão dispostos em uma área, dá-se a impressão de continuidade, sendo praticamente impossível identificar as fronteiras entre cada ponto. Na Figura 5 a região selecionada é expandida de forma que seja possível identificar a fronteira entre pixels.

Figura 5 - Região expandida para exibição das fronteiras entre os pixels e  $f(x,y)$  com os valores RGB do pixel selecionado.



Fonte: Elaboração do próprio autor a partir de imagem da internet

As formas de representação variam em atributos (OLIVEIRA et al., 2009) e complexidade (BACKES, 2010). Existem diversos meios de representar uma imagem digital de acordo com os parâmetros discriminados. Definindo as cores como parâmetros, as imagens são classificadas como *binárias*, *níveis de cinza* ou *coloridas*.

Na representação *binária* os pixels assumem os valores  $f(x,y)=0$  para a cor preta e  $f(x,y)=1$  para a cor branca (Figura 6[a]). Imagens nesta representação requerem pouco espaço para armazenamento e demandam menos processamento. As imagens representadas em *níveis de cinza* (Figura 6[b]), também conhecidas como *monocromáticas*, possuem pixels que assumem valores entre  $f(x,y)=0$  (preto) e  $f(x,y)=255$  (branco). Demais valores dentro deste intervalo correspondem a tons de cinza. Já a representação de uma imagem *colorida* está diretamente relacionada aos atributos do espaço de cor (*color space*) utilizados. No espaço de cor RGB (Figura 6[c]), a imagem digital é descrita como uma matriz de duas dimensões e três planos. Cada plano é responsável por armazenar o valor de uma tonalidade: *red* (vermelha), *green* (verde) e *blue* (azul). Assim como na escala de cinza, a tonalidade das cores está compreendida entre os valores 0 (inexistência da cor) e 255 (em sua totalidade). Neste sistema aditivo a soma dos três planos resulta na cor final do pixel (GONZALEZ; WOODS, 2001).

Figura 6 - Representação de uma imagem em três formas: em [a] a imagem binária utiliza somente as cores preta e branca. A Imagem em [b] é representada utilizando escalas de cinza. Em [c] a imagem é representada por cores.



Fonte: Dorsey (2015)

A exemplo do RGB, existem outras formas de representar uma imagem colorida. As formas de representação diferenciam-se pelos parâmetros que as compõem. O espaço de cor HSV descreve a cor pela sua matiz (*Hue*), saturação (*Saturation*) e iluminação (*Value*). Já o espaço de cor YCbCr descreve utilizando a combinação dos parâmetros de luminância (*Y*),

crominância azul ( $C_b$ ) e crominância vermelha ( $C_r$ ) (PREMA; MANIMEGALAI, 2012).

A representatividade dos espaços de cores e suas principais características serão discutidas no Capítulo 3.

Além da descrição por cores, uma imagem também pode ser descrita por meios que estão além da limitação existente no campo visual humano. A visão computacional, utilizando dos métodos de processamento de imagens, manipula imagens geradas por fontes das quais a visão humana não associa ao campo visual. Essas fontes incluem ultrassom, microscopia eletrônica e imagens geradas por computador (GONZALEZ; WOODS, 2001), (TAGLIAFERRO, 2007).

## 2.2 Processamento Digital de Imagens

Para que uma aplicação seja capaz de classificar a informação contida em uma imagem, inicialmente é preciso definir como a informação será obtida. Processar uma imagem consiste em transformá-la sucessivamente com o objetivo de extrair com maior facilidade a informação nela presente (ALBUQUERQUE, 2000). Os métodos de manipulação da informação fundamentam-se na compreensão de imagens, análise em multirresolução e em multifrequência, análise estatística, codificação etc. Pinto (2009) caracteriza o Processamento Digital de Imagens (PDI) como uma subárea do estudo de Processamento de Sinais Digitais. Os sinais, assim como as imagens, são um suporte físico que carregam consigo uma informação específica. Esta pode estar associada a uma medida (relacionado a um fenômeno físico) ou a um nível cognitivo (representação do conhecimento). Dentre as técnicas de processamento dessa informação, Gonzalez e Woods (2001) as definem em três níveis:

1. Processos de níveis baixos: compreendem operações primitivas de pré-processamento, como a redução do ruído, o realce de contraste e a intensificação de imagens. Processos deste grupo recebem uma imagem como entrada e geram em sua saída outra imagem modificada.

2. Processos de nível médio: inicialmente utilizam técnicas de segmentação que dividem a imagem em regiões ou objetos a partir de um determinado atributo ou característica. O próximo passo consiste na descrição desses objetos para classificá-los (reconhecimento) dentro de um determinado grupo. Os processos de nível intermediário



também recebem imagens como entrada e em sua saída são expressos os atributos (bordas e contornos, identificação dos objetos) contidos nessas imagens.

3. O processo de nível alto busca realizar, através da análise de imagens, funções cognitivas ao “dar sentido” a um conjunto de objetos reconhecidos. Sua aplicação emula a cognição associada à visão humana.

Seja qual for o nível utilizado, o PDI utiliza funções matemáticas para manipulação de dados numéricos contidos em uma imagem digital. Sua utilização tem crescido ao passo dos avanços nas tecnologias de processamento de dados, pois suas aplicações requerem reconhecimento e análise de grandes volumes de dados no menor espaço de tempo possível (SANDOVAL, 2014).

Um exemplo de sua aplicação é o algoritmo que identifica as placas de trânsito a partir da imagem obtida pelas câmeras de monitoramento do tráfego (ERPEN, 2004). Inicialmente a imagem da placa do automóvel é pré-processada na etapa de binarização que tem o objetivo de reduzir a quantidade de informação irrelevante separando os objetos de interesse que neste caso são as letras e os dígitos. Aplica-se na imagem resultante um algoritmo que realiza o isolamento dos caracteres e identifica se cada região encontrada pertence a um dos sete caracteres da placa. A Figura 7 ilustra o resultado do processo de obtenção dos caracteres da placa de dois veículos.

Figura 7- Segmentação dos caracteres das placas de veículos utilizando métodos de nível baixo.



Fonte: Erpen (2004)

Aplicado em áreas como a robótica, imagens de satélite, diagnósticos médicos,

exploração marítima, restauração de obras de arte e outras diversas, técnicas que utilizam PDI manipulam grande quantidade de dados e possuem a capacidade de extrair informações imperceptíveis aos olhos humanos.

### 2.3 Redes Neurais Artificiais

O método proposto neste trabalho utiliza um classificador dinâmico inspirado na estrutura neural de organismos inteligentes. Uma Rede Neural Artificial (RNA) é um processador paralelamente distribuído constituído de unidades de processamentos simples (neurônios), que têm a propensão natural para armazenar conhecimento experimental (através do treinamento) e torná-lo disponível para uso, (HAYKIN, 1999).

Haykin (1999) destaca a semelhança das RNAs com o cérebro em dois principais aspectos:

- O conhecimento é adquirido pela rede através de seu ambiente por intermédio de um processo de aprendizagem;
- Possui forças de conexão entre os neurônios conhecidas como pesos sinápticos, que são utilizados para armazenar o conhecimento adquirido.

A Figura 8 mostra o modelo de um neurônio biológico (esquerda) e um neurônio artificial (direita). Através de suas conexões (*sinapses*) o neurônio recebe as informações  $X_n$  modificadas pelos pesos  $w_n$ .

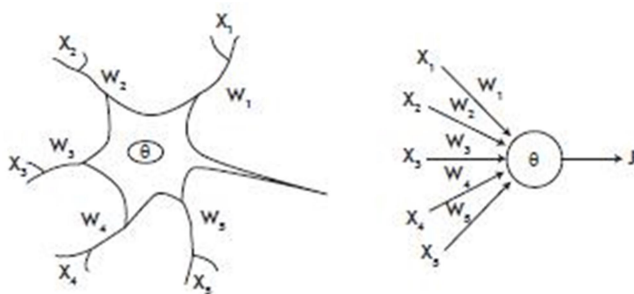


Figura 8 - Estrutura de um neurônio biológico (esquerda) e neurônio artificial (direita).

Fonte: Haykin (1999)

A rede neural extrai seu poder computacional através de sua habilidade de aprender e portanto de generalizar. Trata-se por *generalização* o fato de a rede neural produzir saídas adequadas para as entradas que não estavam presentes durante o treinamento (aprendizagem). Esta capacidade de processamento de informação torna possível para a RNA resolver problemas complexos de grande escala que são atualmente difíceis de tratar. Contudo, na prática, as RNAs não fornecem uma solução trabalhando individualmente. Nesse caso, um problema complexo de interesse é decomposto em um número de tarefas relativamente simples, e atribui-se a redes neurais um subconjunto de tarefas que coincidem com as suas capacidades (HAYKIN, 1999), (WIDROW; LEHR, 1990).

Beale, Hagan e Demuth (2016) apontam que, do desenvolvimento à utilização de uma RNA, as seguintes etapas devem ser observadas:

1. Coleta dos dados;
2. Criação da rede;
3. Treinamento;
4. Teste e análise;
5. Utilização da rede.

Inicialmente os dados devem ser coletados e preparados dentro do contexto na qual a rede será utilizada. Um dos maiores obstáculos está no intuito de incorporar conhecimento prévio à rede, contudo a RNA será tão precisa quanto os dados utilizados para treiná-la.

Durante a criação de uma RNA é definida sua estrutura neural (número de camadas, número de neurônios por camada, tipo de não-linearidade). A maneira pela qual os neurônios estão estruturados está intimamente ligada com o algoritmo de aprendizagem utilizado para treinar a rede.

O treinamento se dá via ajuste de pesos, seguido pela etapa de análise, onde é convalidado se o conhecimento foi adquirido pela rede, ou seja, se a rede está apta para solucionar o problema.

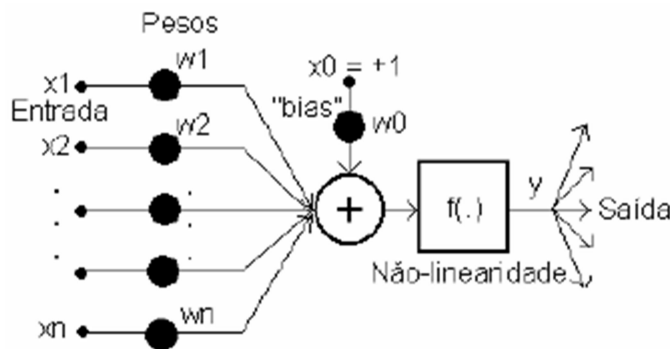
### 2.3.1 Modelo de um neurônio

O neurônio é uma unidade de processamento de informações fundamental para a operação de uma rede neural. Na figura 9 destacam-se os três elementos básicos de um neurônio (HAYKIN, 1999), (WIDROW; LEHR, 1990):

- Conjunto de sinapses: As sinapses são elos de conexão entre os neurônios, caracterizadas por um peso  $w$  ou força própria.
- Combinador linear (somador): Para somar os sinais de entrada ponderados pelas sinapses do neurônio;
- Função não linear (função de ativação): Para restringir a amplitude da saída que será enviada aos demais neurônios. Também conhecida como função restritiva, ela limita o intervalo da amplitude do sinal de saída a um valor finito.

O neurônio pode conter também um peso bias  $w_0$  alimentado por uma constante  $X_0 = +1$  que desempenha o controle do nível de saída do neurônio.

Figura 9 - Modelo de um neurônio.



Fonte: Haykin (1999)

### 2.3.2 Topologia de Redes Neurais Artificiais

A definição da topologia e do algoritmo de treinamento (ajuste dos pesos) possui grande importância para obtenção do resultado desejado (BEALE; HAGAN; DEMUTH, 2016). A

forma pela qual os elementos de processamento (neurônios) e suas conexões (sinapses) são organizadas possui relação direta ao tipo de problema a ser aplicado.

Uma rede neural composta em camadas, na sua forma mais simples, há uma camada de entrada de *nós* que se projeta sobre uma camada de saída de neurônios. Ela é ilustrada na figura 10[a]. Esta rede é chamada de rede de camada única, sendo que tal designação se refere somente à camada dos *nós* computacionais (neurônios). A camada de entrada não é mencionada pois nela não é realizada qualquer cálculo (HAYKIN, 1999).

A rede com múltiplas camadas diferencia-se pela presença de uma ou mais camadas ocultas, cujos *nós* computacionais são chamados de neurônios ocultos. Adicionando-se uma ou mais camadas ocultas, a rede adquire uma perspectiva global apesar de sua conectividade local, devido ao conjunto extra de conexões sinápticas e da dimensão extra de interações neurais (CHURCHLAND; SEJNOWSKI, 1994).

Os nós da camada de entrada da rede alimentam a primeira camada de neurônios (primeira camada oculta) cujos sinais processados são enviados aos neurônios da segunda camada. Os sinais de saída da segunda camada são utilizados como entradas para a terceira camada, e assim por diante por todo resto da rede. O conjunto de sinais de saída dos neurônios da camada de saída (final) da rede constitui a resposta global para o padrão fornecido pelos nós de fonte da camada de entrada.

Na figura 10[b] é ilustrada uma rede neural de múltiplas camadas alimentada para frente com única camada oculta referida como uma rede 10-4-2 pois sua estrutura é constituída por 10 nós na camada de entrada, 4 neurônios ocultos e 2 neurônios de saída.

### **2.3.3 Processo de aprendizagem**

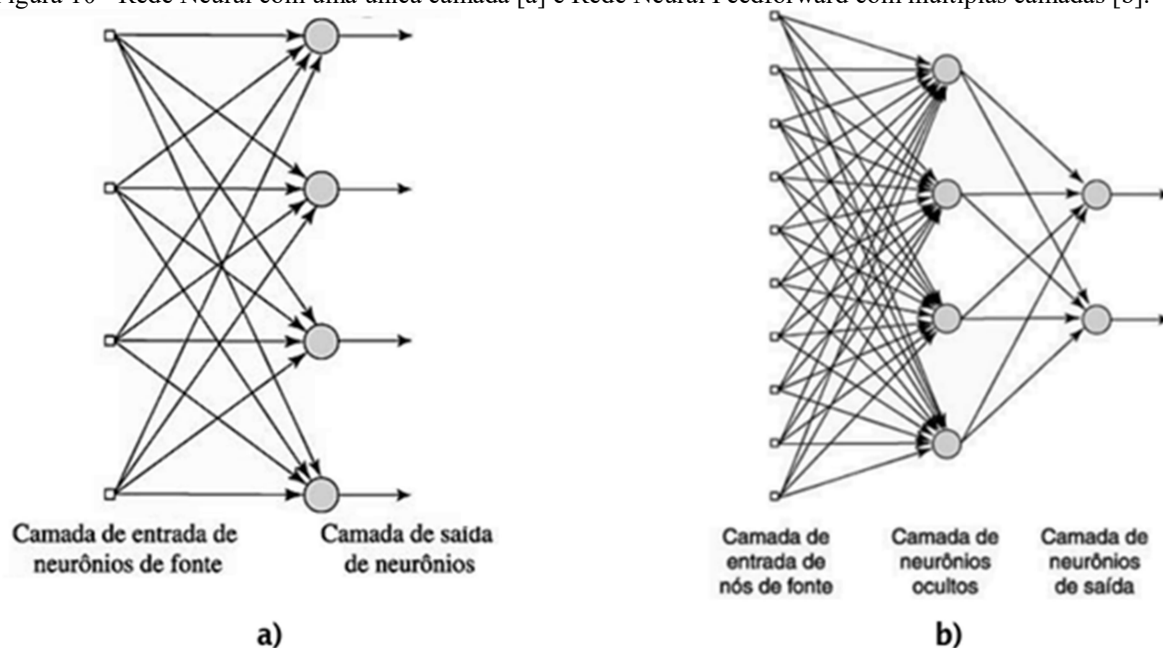
Haykin (1999) define a aprendizagem, no contexto de Redes Neurais como: “Um processo pelo qual os parâmetros livres de uma rede neural são adaptados através de um processo de estimulação pelo ambiente no qual a rede está inserida. O tipo de aprendizagem é determinado pela maneira pela qual a modificação dos parâmetros ocorre”.

Uma RNA aprende acerca de seu ambiente através de um processo iterativo de ajustes aplicados a seus parâmetros livres (*pesos sinápticos e bias*).

No processo de aprendizagem supervisionada, pode-se considerar como *professor* o conhecimento sobre o ambiente. Ele é representado pelo conjunto de exemplos de *entrada e*

sua respectiva *saída desejada*. Define-se *saída desejada* como a ação ótima a ser realizada pela rede.

Figura 10 - Rede Neural com uma única camada [a] e Rede Neural Feedforward com múltiplas camadas [b].



Fonte: Haykin (1999)

Os parâmetros livres da rede são ajustados sob influência combinada do vetor de treinamento e do sinal de erro que é obtido na diferença entre a resposta desejada e a resposta obtida (real) da rede. O ajuste é feito passo a passo, iterativamente. O conhecimento ambiente do professor é então transferido para a rede através do treinamento. Quando a condição é alcançada, ou seja, quanto mais próxima for a saída da rede com a saída desejada, podemos então dispensar o professor e deixar a RNA lidar com o ambiente por si mesma.

### 2.3.4 RNA usada no reconhecimento de padrões

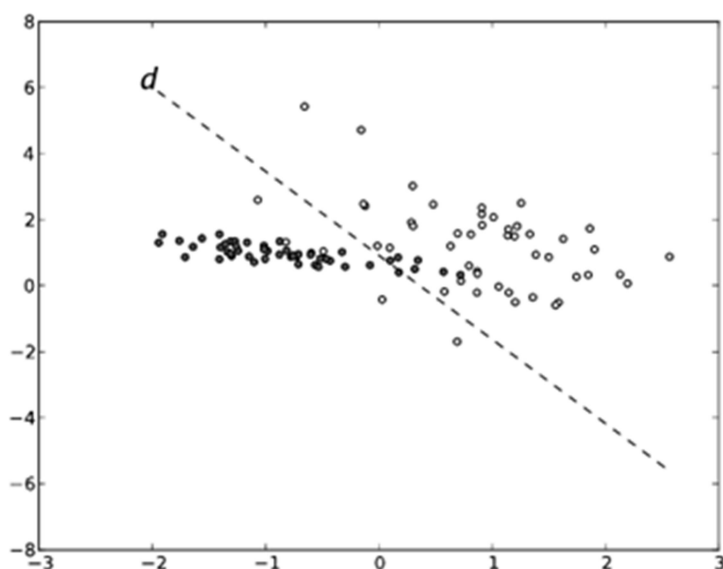
A RNA realiza o reconhecimento de padrões passando inicialmente por uma fase de treinamento, onde é apresentado repetidamente à rede um conjunto de padrões de entrada junto com a classe à qual cada padrão pertence (HAYKIN, 1999), (TAGLIAFERRO, 2007), (BRITO BISNETO, 2011).

Aplicada em um problema de duas classes (Figura 11), ao passo que os exemplos são apresentados, a rede neural ajusta o espaço de decisão  $d$  que representa a separabilidade entre

as classes. O treinamento segue até que a divisão entre as classes resulte na menor taxa de erro possível.

Na utilização da RNA, apresenta-se à rede um novo padrão que não foi visto antes, mas que pertence a uma das classes definidas durante o treinamento da rede. A partir das características do novo padrão apresentado, a rede neural atribui esse padrão à classe correspondente. A capacidade de atribuir corretamente um padrão novo à sua classe correspondente é conhecida como *generalização*.

Figura 11 - Ajuste da divisão  $d$  entre duas classes realizado a partir dos padrões apresentados durante o treinamento da rede neural artificial.



Fonte: Haykin (1999)

## 2.4 Resumo e discussão

O processo de identificação de pixels de pele proposto emula a visão e cognição humana através do processamento digital de imagem e das redes neurais artificiais, respectivamente.

No PDI, a segmentação realiza a extração da região de interesse a fim de minimizar a quantidade de informação a ser processada. Por sua vez a RNA, ao receber os dados pré-processados, realiza a classificação com base nas características apresentadas. Para tornar-se

capaz de realizar tal tarefa, a RNA é primeiramente submetida a um processo de aprendizado, no qual o conhecimento é adquirido a partir de exemplos e seus pesos sinápticos são ajustados (representação do conhecimento).

Neste método, as imagens são descritas pelas cores de cada pixel. Essa representação é feita numericamente e varia de acordo com o espaço de cor utilizado. No próximo Capítulo serão apresentados os espaços de cor que foram objetos de estudo neste trabalho e suas características que favorecem ou prejudicam o processo de classificação de pele.



### 3 ESPAÇOS DE COR

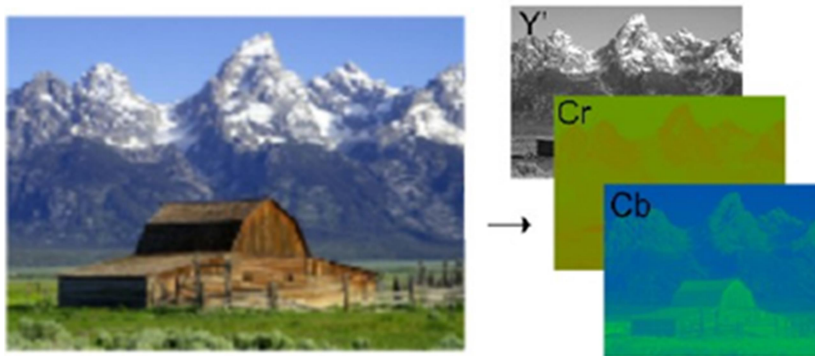
A cor é uma percepção incitada no momento em que diferentes comprimentos de onda alcançam a retina do olho (GONZALEZ; WOODS, 2001). Este estímulo varia de acordo com a quantidade de luz absorvida/refletida na superfície de um objeto e dessa variação são geradas as tonalidades. Se todos os comprimentos de onda passam simultaneamente pela retina, o resultado é a percepção da cor branca, mas caso não haja nenhuma, tem-se a percepção da cor preta (METLEK, 2009). Pela reflexão da luz podem ser obtidas noções de cores e/ou textura de uma superfície. A pele humana possui uma cor característica que é facilmente reconhecida pela percepção humana.

Na visão computacional, a cor da pele é um forte indício da presença humana em mídias, fato que torna o processo de identificação e extração possíveis por meio da identificação de pixels com cor pele. O desenvolvimento de um método que utiliza a cor de pele como critério para classificação depara-se em três questões principais (VEZHNEVETS; SAZONOV; ANDREEVA, 2003): Qual espaço de cor utilizar? Como a cor da pele está distribuída no espaço de cor escolhido? Qual será a natureza do classificador empregado?

Espaços de cores são estratégias pelas quais é possível criar, especificar e visualizar a cor em uma imagem digital (SANTOS, 2012). Uma imagem em níveis de cinzas é descrita por uma única função bidimensional  $f(x, y)$  enquanto a imagem colorida é descrita por duas ou três funções  $f(x, y)$ . Cada uma dessas funções corresponde a um componente característico do espaço de cor definido. A imagem colorida na Figura 12 é descrita no espaço de cor YCbCr. Segundo Sandoval (2014), os espaços de cor possuem diferentes propriedades e características que podem viabilizar ou dificultar o processo de identificação de pixels que contém pele.

Em um espaço amostral constituído por todas as tonalidades descritas por um espaço de cor, as tonalidades de cor de pele preenchem um subespaço específico chamado de *skin cluster* (Figura 13). Através destes clusteres observa-se que os tons de pele não estão inseridos aleatoriamente no espaço de cor, mas sim clusterizados em uma determinada região que varia em tamanho e localização de acordo com o espaço de cor utilizado (ELGAMMAL; MUANG; HU, 2009). Os clusteres de pele no espaço de cor CbCr (Figura 13 [d], [e] e [f]) preenchem uma região concêntrica enquanto no espaço RG (Figura 13 [a], [b] e [c]) os clusteres estão distribuídos ao longo do espaço de cor de maneira disforme.

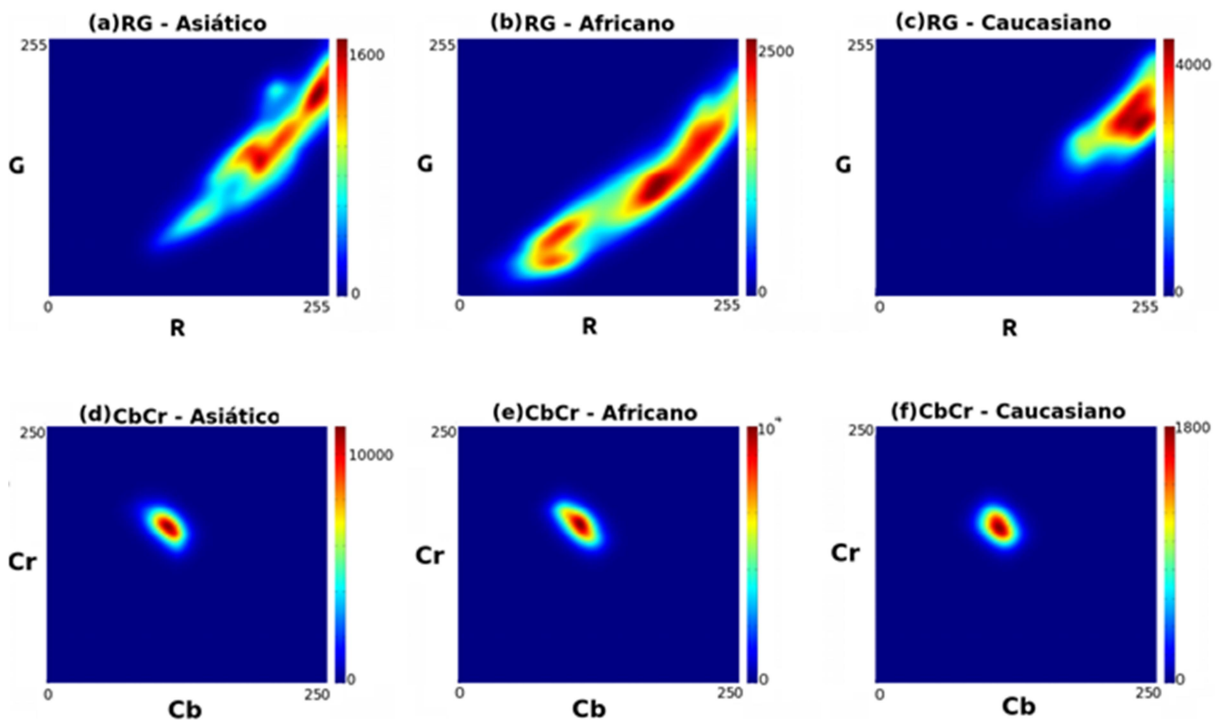
Figura 12 - Imagem descrita pelos três parâmetros no espaço de cor YCbCr.



Fonte: Mehta (2014)

Figura 13 - Clusteres (regiões) de pele de diferentes etnias nos espaços de cor RG e CbCr.

### Densidade de amostras de pele



Fonte: Elgammal, Muang e Hu (2009)

Definir se a cor de um pixel está ou não contida neste cluster é um problema de classificação de duas classes (VEZHNEVETS; SAZONOV; ANDREEVA, 2003), (ELGAMMAL; MUANG; HU, 2009). O classificador empregado (neste trabalho, a RNA) deve ser capaz de definir os limites do cluster de pele e rotular se a tonalidade de um pixel

apresentado está contida neste cluster. Um método de classificação eficaz apresenta baixos índices de falsos positivos (pixels *não pele* identificados como *pele*) e falsos negativos (pixels *pele* identificados como *não pele*).

A escolha do espaço de cor define o modelo do *cluster de pele* e este possui relação direta com a eficácia do processo de detecção (ELGAMMAL; MUANG; HU, 2009). Neste capítulo serão abordados os espaços de cor utilizados neste trabalho e as características dos parâmetros que as compõem.

### 3.1 Espaço de Cor RGB (Red Green Blue)

Sua popularização originou-se nos monitores CRT dada sua conveniência em descrever as tonalidades por meio da combinação das três cores primárias: vermelho (R), verde (G), azul (B) (Figura 14). A intensidade de cada parâmetro é expressa no intervalo entre 0 (inexistência da cor) e 255 (em sua total tonalidade). Dessa forma as cores vermelha [255, 0, 0], verde [0, 255, 0] e azul [0, 0, 255] compõem o sistema aditivo que combinadas formam outras cores como o amarelo [255, 255, 0], o preto [0, 0, 0] e o branco [255, 255, 255]. A alta correlação entre os três canais de cores resulta na mistura entre elementos de luminância (luz) e cromaticidade (cor) tornando este espaço de cor de modo perceptual disforme fazendo com que o valor RGB da pele de um mesmo indivíduo sofra grandes mudanças se capturadas em condições de iluminação diferentes (ELGAMMAL; MUANG; HU, 2009). Por esta razão o espaço de cor RGB apresenta o *cluster de pele* disforme que não está concentrado em uma específica área se comparado ao *cluster* de outros espaços de cores (visto na Figura 13). O *cluster de pele* estende-se de tal forma para refletir as diferentes intensidades de iluminação nas amostras de pele.

Precursor nas pesquisas de classificação de cor de pele, o espaço RGB foi objeto de abordagens estáticas (por exemplo, classificadores explícitos) ao longo da década de 90 (FLECK; FORSYTH; BREGLER, 1996), (JONES; REHG, 2002), (KOVAC; PEER; SOLINA, 2003) e atualmente o maior poder de processamento dos computadores tem possibilitado o desenvolvimento de métodos dinâmicos (BHATT et al., 2009), (BHOYAR; KAKDE, 2010).

Figura 14 - [a] Imagem original e a distribuição em três canais de cores, [b] vermelho, [c] verde e [d] azul.



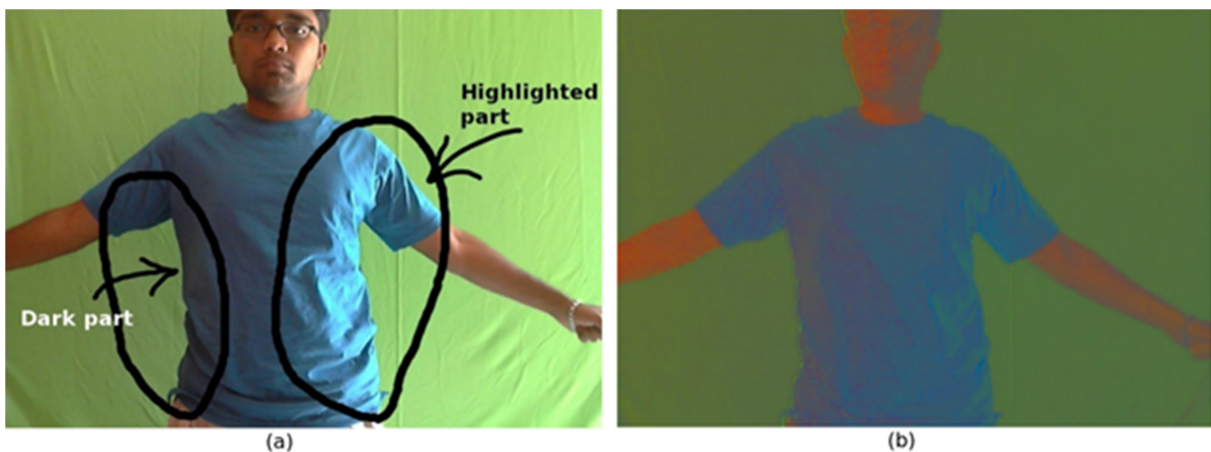
Fonte: Elaboração do próprio autor a partir de uma imagem de domínio público

Apesar da sua limitação característica, o RGB é amplamente utilizado por se tratar de um espaço de cor padrão em grande parte dos aparelhos de captura e processamento e dada a simplicidade de sua aplicação.

### 3.1.1 Espaço de cor RGB normalizado

A normalização do espaço de cor RGB tem como finalidade reduzir os efeitos da iluminação em uma imagem digital (Figura 15).

Figura 15 – Imagem original [a] e redução dos efeitos da iluminação obtidas com a normalização [b].



Fonte: Shende (2013)

O espaço de cor RGB é normalizado através das equações:

$$r = \frac{R}{R + G + B} \quad (8)$$

$$g = \frac{G}{R + G + B} \quad (9)$$

$$b = \frac{B}{R + G + B} \quad (10)$$

A soma dos componentes normalizados corresponde a uma unidade ( $r+g+b=1$ ). Portanto a normalização permite a representação de cores utilizando somente os parâmetros  $r$  e  $g$  sem perda de informações, uma vez que o parâmetro  $b$  pode ser obtido por  $r$  e  $g$  ( $b=1-r-g$ ). A representação de cores utilizando apenas dois parâmetros simplifica o processamento ao utilizar uma coordenada bidimensional ao invés da coordenada tridimensional fornecida pelo RGB. Os parâmetros  $r$  e  $g$  (que formam o espaço  $Nrg$ ) são também conhecidos como *tons puros* por diminuir a dependência da iluminação em certas circunstâncias (SKARBEEK; KOSCHAN, 1994).

### 3.2 Espaço de Cor YCbCr

Gerado a partir da codificação não linear do espaço de cor RGB pelas equações:

$$Y = 0,299R + 0,587G + 0,114B \quad (11)$$

$$C_r = R - Y \quad (12)$$

$$C_b = B - Y \quad (13)$$

O espaço YCbCr descreve a cor através de parâmetros independentes (Figura 16). A luminância é representada exclusivamente pelo componente  $Y$  e pode ser expressa entre os intervalos de 0 a 1 ou 0 a 255. As diversas tonalidades de cores são representadas pela combinação entre as cores  $C_b$  (azul) e  $C_r$  (vermelha) com valores dentro do intervalo de 0 a

255.

Figura 16 - [a] Imagem original e a distribuição em três canais. [b] Iluminação (Y), [c] crominância Azul (Cb) e [d] crominância vermelha (Cr)



Fonte: Elaboração do próprio autor a partir de uma imagem de domínio público

Em razão do isolamento da componente de iluminação  $Y$  o *cluster de pele* é localizado de forma concentrada e compacta dentro do espaço amostral das cores  $CbCr$ . Outro aspecto favorável está na proximidade da localização dos clusters de pele de diferentes etnias, característica que viabiliza maior eficácia na delimitação de fronteiras entre as classes *pele* e *não pele*. Shin e Tsap (2002) afirmam que os parâmetros  $Cb$  e  $Cr$  compõem o melhor classificador de pele de duas dimensões ao descartar o parâmetro  $Y$ .

### 3.3 Espaço de Cor HSV (Hue Saturation Value)

Classificado como um espaço de cor perceptual, o HSV originou-se na necessidade de especificar numericamente as propriedades intuitivas de uma cor (Figura 17). O parâmetro *Hue* (matiz) é a medida do comprimento de onda da luz refletida pelo objeto, ou seja, a cor. É usualmente representado em graus (0 a 360) ou em valores de 0 a 100%. A intensidade dessa cor é definida pela *Saturation* (saturação), também conhecida como “pureza” da cor. Valores baixos expressam uma tonalidade fraca enquanto valores elevados indicam maior presença da cor e compreendem entre 0 e 100%. *Value* (valor) compõe a representação da luminosidade. Também é expresso entre os valores 0 e 100%.



Figura 17 - [a] Imagem original e a distribuição em três canais, [b] Matiz (H), [c] Saturação (S) e [d] Iluminação (V)



Fonte: Elaboração do próprio autor a partir de uma imagem de domínio público

A discriminação entre as propriedades de luminância e crominância nos espaços de cor perceptuais tem se tornado o foco de pesquisas que trabalham na segmentação de pixels de pele. Tratado de forma isolada, o parâmetro *Hue* possui propriedades invariantes à fontes de luz branca e superfícies foscas (SKARBEEK; KOSCHAN, 1994). Zarit, Super e Quek (1999) concluíram que o HSV obteve melhor eficácia na detecção de pixels com cor de pele para um conjunto de 200 amostras com taxa de 91% de verdadeiros positivos enquanto os espaços de cores YIQ (SURAMPALLI; DAYANAND; DHANANJAY, 2012) e YUV (AL-TAIRI et al., 2014), também pertencentes ao grupo de espaços de cor perceptuais, apresentaram elevadas taxas de erro de classificação.

### 3.4 Resumo e discussão

Trabalhos direcionados à segmentação de pele sucedem da exclusão do parâmetro relacionado à iluminação (ELGAMMAL; MUANG; HU, 2009), (SANDOVAL, 2014). Em suas pesquisas, os parâmetros Y e V podem ser descartados sem depreciar a informação contida, seguindo a idéia de que a tonalidade de pele é invariante à iluminação. A normalização do espaço RGB (Nrg), resultando na transformação de um espaço tridimensional em bidimensional também é apontada como solução para minimizar a dependência do espaço RGB à diferentes fontes de iluminação.

A fim de investigar tal afirmação, na fase de criação e teste serão desenvolvidas duas Redes Neurais Artificiais para cada espaço de cor: RGB e Nrg, YCbCr e CbCr, HSV e HS. Cada espaço de cor terá uma RNA que recebe todos os parâmetros e outra em que o parâmetro correspondente à iluminação (normalizado, no RGB) é descartado. Dessa forma, com base nas

saídas produzidas, será possível comprovar se a exclusão da variável da iluminação traz melhoras no processo de identificação de pele dentro do modelo proposto.



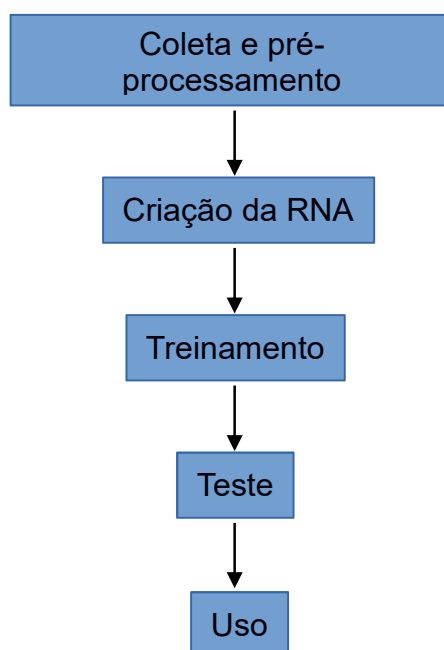
## 4 METODOLOGIA PROPOSTA E RESULTADOS OBTIDOS

Neste capítulo estão registradas as etapas do desenvolvimento, treinamento e utilização dos classificadores de pixels de pele e a análise com base nos resultados obtidos.

### 4.1 Desenvolvimento do Classificador Neural

Em observância às melhores práticas para obtenção do resultado almejado, o desenvolvimento da Rede Neural Artificial seguiu as etapas (Figura 18) definidas por Beale et. al (2015):

Figura 18 – Organograma com as etapas do desenvolvimento ao uso das RNAs para cada espaço de cor.



Fonte: Beale, Hagan e Demuth (2016)

#### 4.1.1 Coleta De Dados

Os dados utilizados para treinamento e teste do classificador neural compõem o *Skin Segmentation Dataset* disponibilizado pela *UCI Machine Learning Repository*

<<http://archive.ics.uci.edu/ml/datasets/Skin+Segmentation>>. O *dataset* contém 245.057 exemplos de tonalidades, das quais 50.859 são amostras de *pele* e 194.198 *não pele*. Os exemplos de pele representam diferentes grupos etários (jovens, adultos e idosos) de diversificada etnia e gênero (BHATT; DHALL, 2016).

No *dataset*, cada tonalidade é descrita por uma linha composta por quatro colunas. As três primeiras colunas informam os valores R, G e B da tonalidade. A quarta coluna define a classe, atribuindo o valor 1 (um) para *pele* e valor 0 (zero) para *não-pele*.

A partir dos exemplos do *dataset* (Tabela 1), os valores do espaço RGB foram normalizados (Tabela 2) e convertidos para os espaços de cor HSV (Tabela 3) e YCbCr (Tabela 4).

**Tabela 1: Valores da tonalidade no espaço de cor RGB e classificação *pele* e *não pele* encontrados no *Skin Segmentation Dataset*.**

<b>R</b>	<b>G</b>	<b>B</b>	<b>y</b>
178	175	130	1
178	175	131	0
178	175	132	0
176	173	128	1

Fonte: Bhatt e Dhall (2016)

**Tabela 2: Valores da tonalidade no espaço de cor Nrg obtidas através da normalização do RGB e classificação *pele* e *não pele*.**

<b>r</b>	<b>g</b>	<b>y</b>
0,3685	0,3623	1
0,3677	0,3615	0
0,3670	0,3608	0
0,3689	0,3626	1

Fonte: Elaboração do próprio autor a partir de Bhatt e Dhall (2016)

**Tabela 3: Valores da tonalidade no espaço de cor YCbCr obtidas através da conversão do RGB e classificação *pele e não pele*.**

<b>Y</b>	<b>Cr</b>	<b>Cb</b>	<b>y</b>
130	133	108	1
163	132	108	0
163	132	109	0
161	133	108	1

Fonte: Elaboração do próprio autor a partir de Bhatt e Dhall (2016)

**Tabela 4: Valores da tonalidade no espaço de cor HSV obtidas através da conversão do RGB e classificação *pele e não pele*.**

<b>H</b>	<b>S</b>	<b>V</b>	<b>y</b>
56	27	69.8	1
56	26.4	69.8	0
56	25.8	69.8	0
56	27.3	69	1

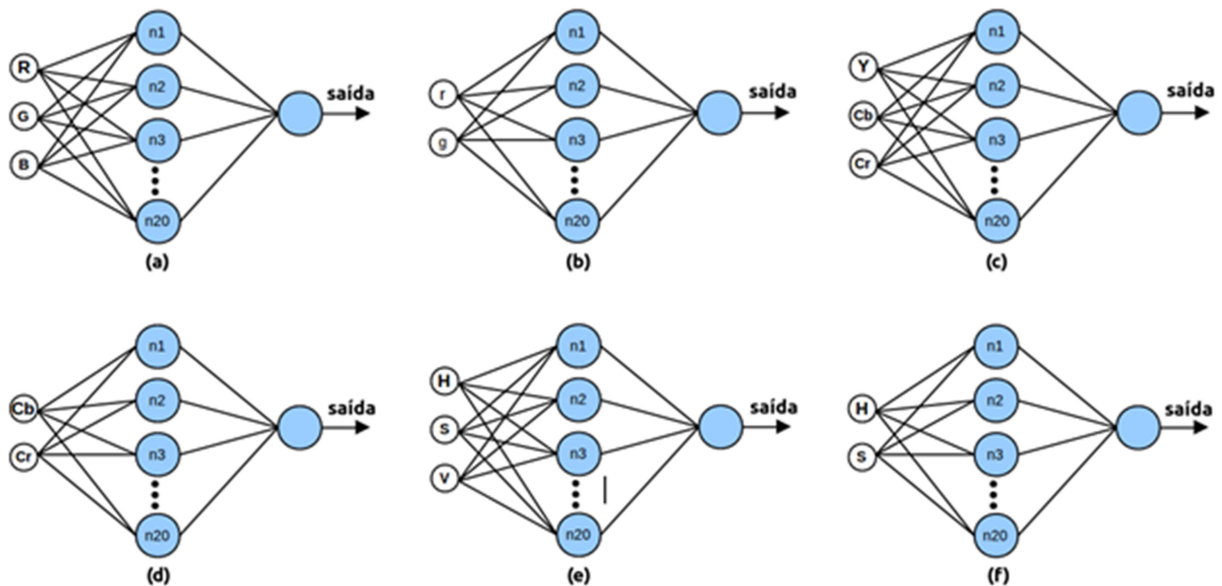
Fonte: Elaboração do próprio autor a partir de Bhatt e Dhall (2016)

#### **4.1.2 Desenvolvimento das Redes Neurais Artificiais**

Os espaços de cor são compostos por parâmetros de naturezas distintas (Capítulo 3), portanto surge a necessidade do desenvolvimento de uma RNA para cada espaço de cor (Figura 19), uma vez que a Rede Neural Artificial adquire o conhecimento exclusivo do ambiente em que é treinada (HAYKIN, 1999) (BEALE; HAGAN; DEMUTH, 2016). As etapas de desenvolvimento, teste e uso foram realizadas no ambiente MATLAB através da ferramenta *nn toolbox* (MATLAB Neural Network Toolbox versão 2016b). As RNAs possuem a arquitetura multicamadas feedforward. Buscando melhores resultados utilizando a estrutura

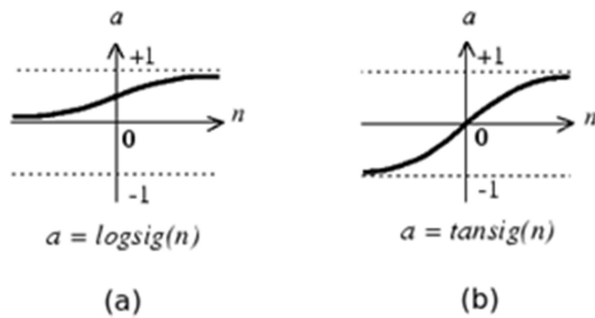
mais simples, foram desenvolvidas e testadas RNAs com 3 e 4 camadas contendo 5, 10, 15 e 20 neurônios em cada camada. Após inúmeros testes, optou-se por trabalhar com as RNAs constituídas por três camadas contendo vinte neurônios na camada oculta sem perda de desempenho. A camada de entrada de cada RNA possui o número de nós receptores correspondente a quantidade de parâmetros de cada espaço de cor: 3 nós para RGB, HSV e YCbCr; 2 nós para Nrg, HS e CbCr. A camada intermediária (oculta) possui 20 neurônios (Figura 19). O sinal de saída de cada neurônio é processado pela função de ativação LOGSIG (Figura 20[a]) que limita o alcance de sua saída dentre os valores 0 (*não pele*) e 1 (*pele*). A camada de saída contém um único neurônio e sua função de ativação é a TANSIG (Figura 20[b]), comumente utilizada em aplicações para classificação de padrões (BEALE; HAGAN; DEMUTH, 2016). Nela será produzida a saída resultante de todos os cálculos realizados pelos neurônios ao longo da rede.

Figura 19 - Modelo das RNAs para os espaços de cor (a)RGB, (b)Nrg, (c)YCbCr, (d)CbCr, (e)HSV e (f)HS.



Fonte: Elaboração do próprio autor

Figura 20 – Função de ativação nos neurônios da camada oculta [a] e da camada de saída [b]



Fonte: Beale, Hagan e Demuth (2016)

#### 4.1.3 Treinamento das Redes Neurais Artificiais

O desenvolvimento de uma Rede Neural Artificial consiste na criação de uma função matemática que recebe como entrada vetores numéricos e gera saídas numéricas. O valor gerado na saída da RNA é influenciado pelo *vetor de entrada*, pelos *neurônios* através de suas *funções de ativação* e pelo conjunto de *pesos* e *bias*. Durante o treinamento da RNA são feitos ajustes no valor desses *parâmetros livres* de forma que, dado o conjunto de treinamento (vetor de entrada e saída desejada) a saída computada pela rede seja a mais próxima da saída desejada.

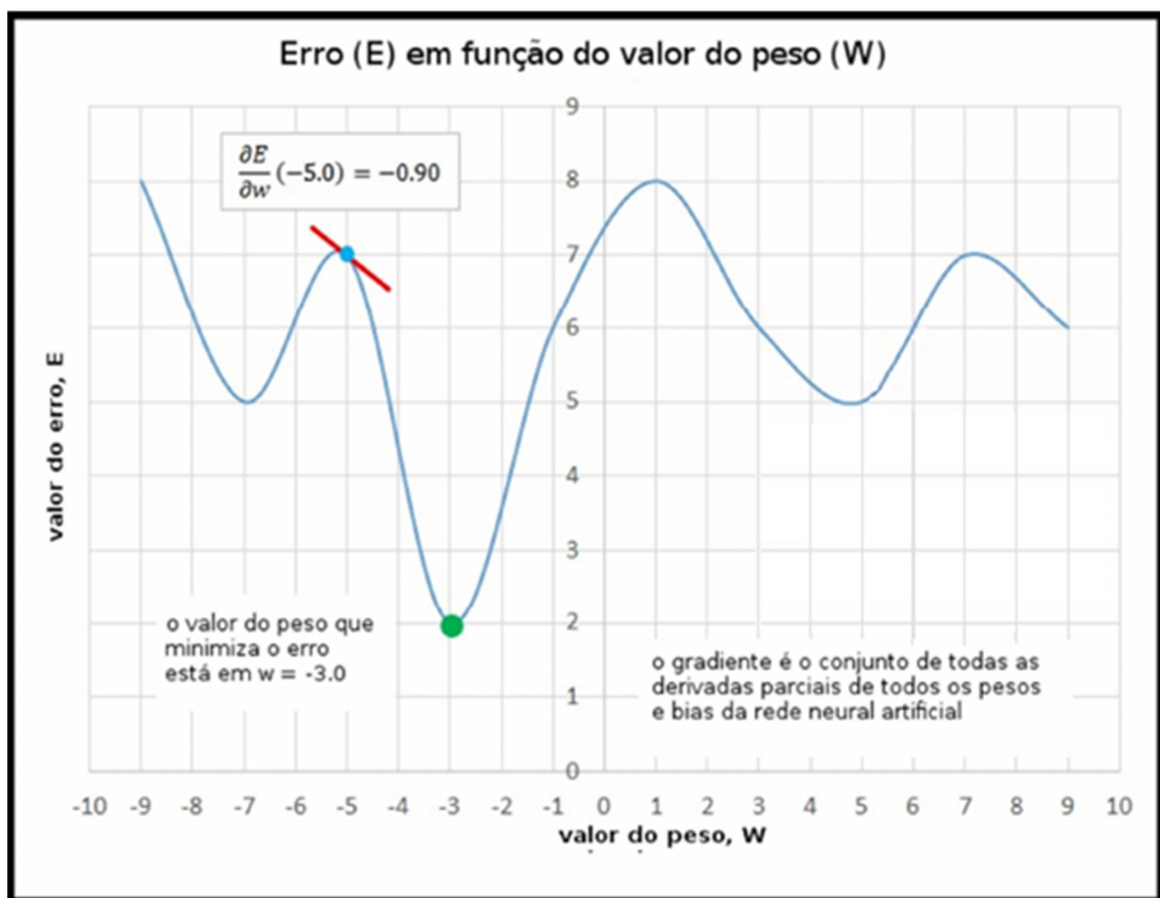
O método de ajustes dos *parâmetros livres* é determinado pelo *algoritmo de treinamento*, neste trabalho, o *Resilient Backpropagation*. Trata-se de um algoritmo similar ao *Backpropagation* convencional, porém, com duas modificações: tende a convergir para o *erro mínimo* mais rapidamente e não recorre à *taxa de aprendizagem constante* (MCCAFFREY, 2015).

##### 4.1.3.1 Resilient Backpropagation

Os algoritmos de aprendizagem baseiam-se no conceito matemático chamado *gradiente descendente*. Esse conceito é ilustrado no gráfico (Figura 21) que mostra a variação do *erro quadrático* produzido pela RNA em relação à mudança do valor do *peso*  $w$  (assumindo que o

valor dos outros *pesos* e *bias* sejam mantidos na mesma forma). A *derivada parcial*  $\frac{\partial E}{\partial w}$  pode ser entendida como o *declive* do *erro* para o valor de  $w$  ou seja, qual direção o ajuste de  $w$  deve seguir para minimizar o *erro quadrático* da rede. O conjunto de todas as *derivadas parciais* de pesos e bias é chamado *gradiente* (BEALE; HAGAN; DEMUTH, 2016), (MCCAFFREY, 2015).

Figura 21 - Variação do *erro* em função do valor do peso  $w$



Fonte: Adaptado de Mccaffrey (2015)

O exemplo no gráfico aponta que, supondo que a *derivada parcial* para  $w = -5,0$  seja  $\frac{\partial E}{\partial w} = -0,9$ . O  *sinal* da *derivada parcial* indica qual direção seguir e a *inclinação* (conhecida como *magnitude*) indica o quanto deve se mover para obter o menor *erro* na saída da RNA.

Ao longo do treinamento, o *Backpropagation* convencional utiliza a *magnitude* da

*derivada parcial* para determinar o quanto o valor do peso  $w$  deve ser ajustado (HAYKIN, 1999). Supõe-se que, baseado na magnitude da *derivada parcial* associada a *taxa de aprendizagem*  $\tau$  o *Backpropagation* convencional defina o tamanho do ajuste de peso  $\Delta w = +7,0$  fazendo com que  $w$  ultrapasse o valor ótimo que está situado em 3,0. Na próxima iteração  $\Delta w$  pode ser muito grande e extrapolar novamente para outra direção. Esta oscilação pode perdurar por várias iterações e ainda assim o *erro mínimo* não ser encontrado.

Para evitar situações como esta utiliza-se a *taxa de aprendizagem* com valores ínfimos para minimizar o impacto da *magnitude*. Desta forma o valor de  $w$  otimizado não é ultrapassado, porém o treinamento torna-se mais lento ao passo que se aproxima do erro mínimo. Em suma, elevado  $\tau$  faz com que o treinamento oscile para ambos os lados “saltando” a solução ótima local e  $\tau$  com valores pequenos tornam o treinamento mais lento ao realizar pequenos ajustes nos pesos em cada iteração.

O *Resilient Backpropagation* traz duas significantes mudanças: Primeiro, desconsidera a *magnitude*, utilizando somente o  *sinal do gradiente*. Segundo, ao invés de empregar uma *taxa de aprendizagem*  $\tau$  única para todos os parâmetros livres, ela mantém valores de ajustes separados para cada *peso* e *bias*, que são adaptados ao longo do treinamento (RIEDMILLER; BRAUN, 1993), (MCCAFFREY, 2015).

Para cada peso  $w$  é atribuído um valor de atualização  $\Delta$  individual. Este valor é adaptativo e evolui durante o processo de atualização com base no sinal da derivada parcial, de acordo com a regra de aprendizagem:

$$\Delta(t) = \left\{ \begin{array}{l} \eta^+ \Delta(t-1), \text{ se } \frac{\partial E(t-1)}{\partial w} \frac{\partial E(t)}{\partial w} > 0 \\ \eta^- \Delta(t-1), \text{ se } \frac{\partial E(t-1)}{\partial w} \frac{\partial E(t)}{\partial w} < 0 \\ \Delta(t-1), \text{ se } \frac{\partial E}{\partial w} = 0 \end{array} \right. \quad (14)$$

em que  $0 < \eta^- < 1 < \eta^+$  e  $\frac{\partial E}{\partial w}$  é a derivada parcial (BRAUN; RIEDMILLER, 1993 citados por SCHMITT, 2009).

Verbalizando o processo de adaptação: Se a derivada parcial anterior e atual mantém o mesmo sinal indicando que o valor de  $w$  para o *erro mínimo* não foi ultrapassado e que o último ajuste reduziu o *erro*, deve-se manter na mesma direção. Inclusive mais rápido. Para tanto o valor de atualização  $\Delta$  é acrescido pelo fator  $\eta^+$  para acelerar a convergência. Por outro lado, quando a *derivada parcial* anterior e atual mudam de sinal indicando que o último ajuste foi grande demais e o valor do peso  $w$  para o *erro mínimo* foi ultrapassado, o valor de atualização  $\Delta$  é penalizado por  $\eta^-$  para não se mover tão longe (MCCAFFREY, 2015).

Uma vez que o *valor de atualização*  $\Delta$  para cada peso é adaptado, o tamanho do ajuste do peso  $\Delta$  é conduzido pela seguinte regra: se a *derivada parcial* é positiva (erro aumentando),  $w$  é decrescido pelo valor de atualização  $\Delta$ . Se a derivada é negativa, soma-se o valor de atualização:

$$\Delta w(t) = \left\{ \begin{array}{l} -\Delta(t), \text{ se } \frac{\partial E}{\partial w}(t) > 0 \\ +\Delta(t), \text{ se } \frac{\partial E}{\partial w}(t) < 0 \\ 0, \text{ se } \frac{\partial E}{\partial w}(t) = 0 \end{array} \right\} \quad (15)$$

$$w^{t+1} = w^t + \Delta w^t$$

O processo de atualização dos pesos ocorre após a apresentação de todos os padrões de treinamento, ou seja, realizado em um treinamento em lote (modo batch) (RIEDMILLER; BRAUN, 1993).

Cada RNA foi treinada por três vezes consecutivas. Os resultados após cada treinamento apresentaram pouca variação, evidenciando que o processo de aprendizagem observou as boas práticas recomendadas por Beale, Hagan e Demuth (2016).

A Tabela 5 lista informações pertinentes do treinamento das RNAs. Os espaços de cores Nrg e HSV tiveram o treinamento de maior duração com elevado número de iterações. Por outro lado o espaço de cor YCbCr convergiu rapidamente para a solução otimizada, seja em 3 ou em 2 dimensões (com a exclusão do parâmetro de iluminação). Os espaços de cores RGB e



YCbCr retornaram melhores resultados na *performance* (valor 0 indica solução ótima). Suas *performances* foram as melhores em comparação com os espaços de cor Nrg e CbCr, citados como alternativa para os problemas com iluminação em classificação de pele (SANDOVAL, 2014).

**Tabela 5: Dados da etapa de treinamento das RNAs para cada Espaço de cor.**

<b>Espaço de cor</b>	<b>Tempo</b>	<b>Iterações</b>	<b>Performance</b>	<b>Erros de validação</b>
<b>RGB</b>	00min56s	167	0,000886	6
<b>Nrg</b>	03min13s	566	0,00255	6
<b>YCbCr</b>	00min47s	139	0,000959	6
<b>CbCr</b>	00min41s	117	0,00135	6
<b>HSV</b>	04min14s	737	0,00192	6
<b>HS</b>	02min56s	518	0,00487	6

Fonte: Dados da pesquisa do autor

#### 4.1.3.2 Critérios de parada do treino

O algoritmo de aprendizagem das RNA realiza ajustes nos parâmetros livres até que uma das condições preestabelecidas seja alcançada. Estas condições são conhecidas como *critérios de parada* (BEALE; HAGAN; DEMUTH, 2016). Por meio destes critérios são definidos o *número máximo de iterações* (nesse caso 1000), a *performance* da RNA (0, indicando que a solução ótima foi encontrada) e o número máximo de *erros de Validação* (definido 6 como padrão na *nntool*). Este último critério foi determinante para o término do treinamento para as RNAs em todos os espaços de cor.

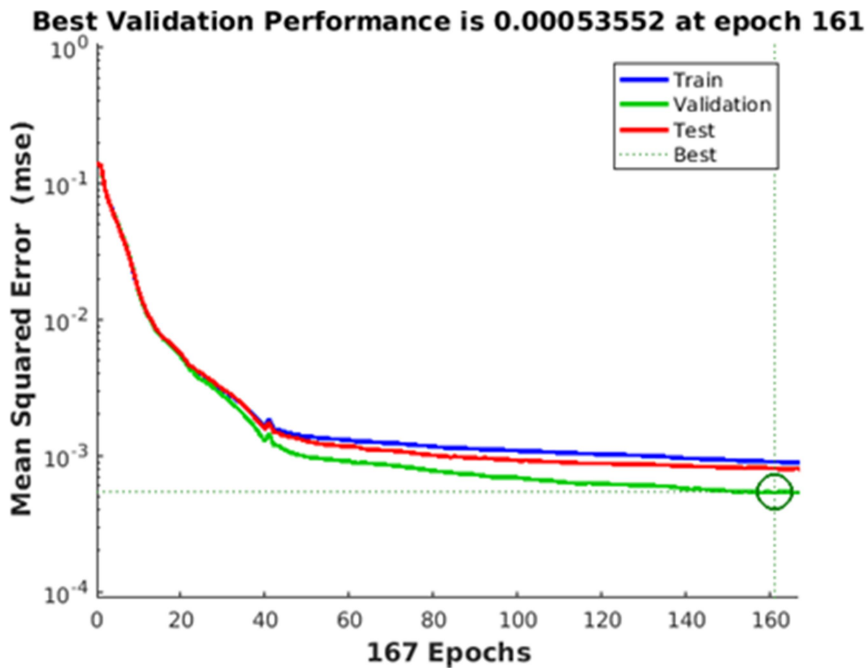
Durante o processo de aprendizagem, a validação dos resultados gerados pela RNA tem por objetivo evitar que ela “decore” os dados utilizados para treinamento e torne-se incapaz de *generalizar*. Entende-se por *generalização* a habilidade de classificar corretamente uma nova informação que pertença a um dos subgrupos e não tenha sido apresentada na etapa de treinamento (HAYKIN, 1999). Para assegurar a capacidade de generalização de uma RNA, o

*dataset* é subdividido em três grupos: 70% para *treino*, 15% para *validação* e 15% para *teste* (BEALE; HAGAN; DEMUTH, 2016).

Após o algoritmo de aprendizagem apresentar todos os dados do grupo de *treino* (uma iteração completada) e realizar o ajuste nos pesos convergindo para a solução ótima, são apresentados à RNA os dados do grupo de *validação*. Caso a saída da RNA para estes novos dados também seja otimizada, o treinamento continua. Por outro lado, se a RNA convergir para os dados de *treino* e não convergir para os dados de *validação*, significa que o último ajuste no peso diminuiu a capacidade de generalizar (este caso é conhecido como *overfitting*) (MCCAFFREY, 2015), (POOLE; MACKWORTH, 2017). Caso essa condição continue após seis iterações contínuas, os pesos retornam para seus valores antes do início dos erros de validações e o treinamento é finalizado. Por fim apresenta-se à rede o terceiro grupo, de *teste*, para verificar se a RNA é capaz de classificar dados em modo *offline* quando o ajuste dos pesos é finalizado.

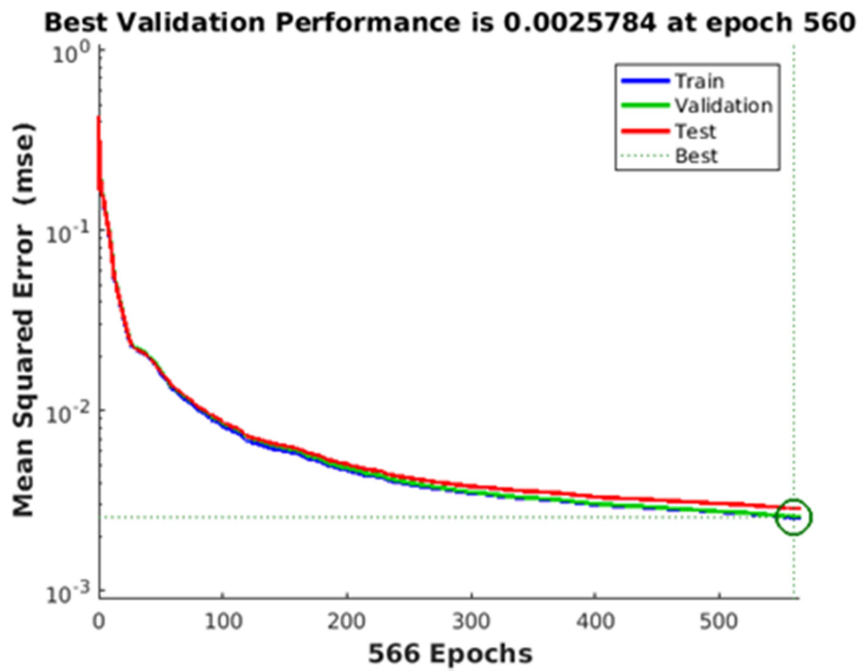
Os gráficos nas Figuras 22 a 27 mostram a medida do *erro médio* (*mean square error*) ao longo das iterações (*epoch*) durante o treino da RNA. É possível identificar que o *erro médio* dos dados de *treino*, *validação* e *teste* convergem na mesma proporção, indicando que o conhecimento é transmitido gradativamente às RNAs ao longo do treinamento. São exibidas também informações da variação do valor do *gradiente* e dos *erros de validação* ao longo do treinamento (Figuras 28 a 33 ). Ao passo que as RNAs adquirem o conhecimento, valor do *gradiente* diminui, indicando que menores ajustes serão necessários na próxima iteração. Os espaços de cor RGB, YCbCr e HS apresentaram menores variações no pico entre iterações contínuas. Os *erros de validação* (*val fail*) apontaram que o espaço HSV teve elevada taxa de erros de validação mesmo com a diminuição para o espaço bidimensional em HS. Por outro lado, o espaço de cor YCbCr apresentou menor número de *erros de validação*.

Figura 22 – Performance da RNA ao longo das iterações no espaço de cor RGB.



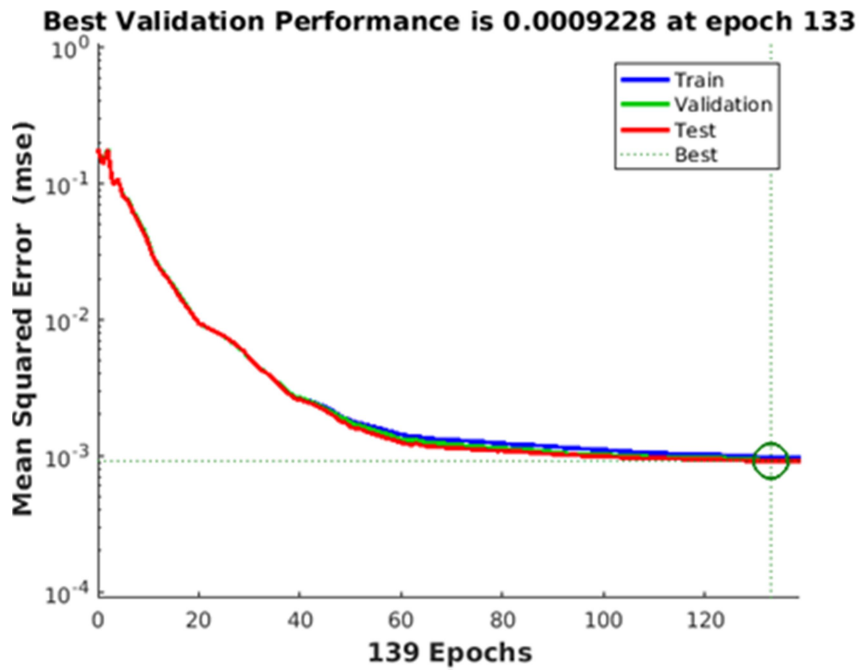
Fonte: Dados da pesquisa do autor

Figura 23 – Performance da RNA ao longo das iterações no espaço de cor Nrg.



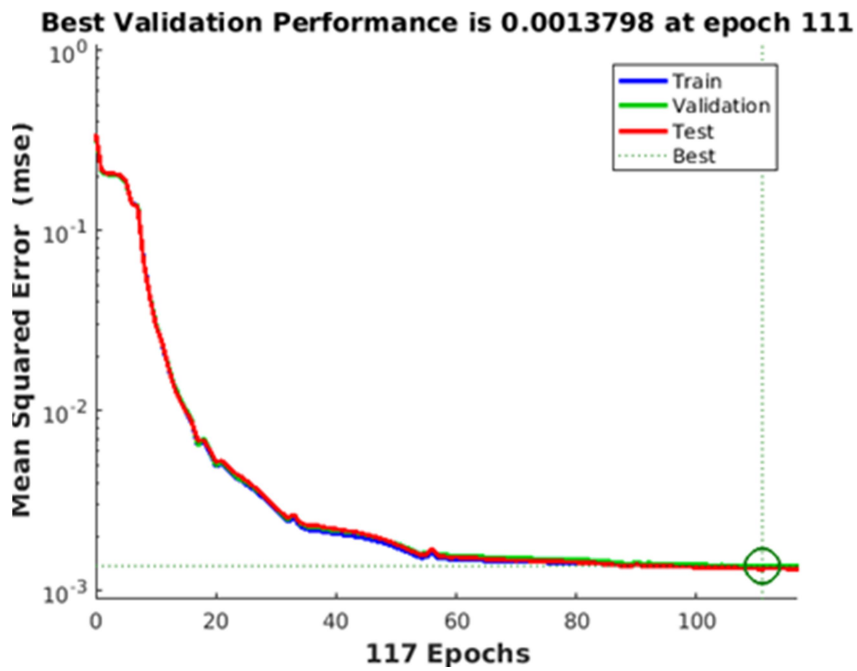
Fonte: Dados da pesquisa do autor

Figura 24 – Performance da RNA ao longo das iterações no espaço de cor YCbCr.



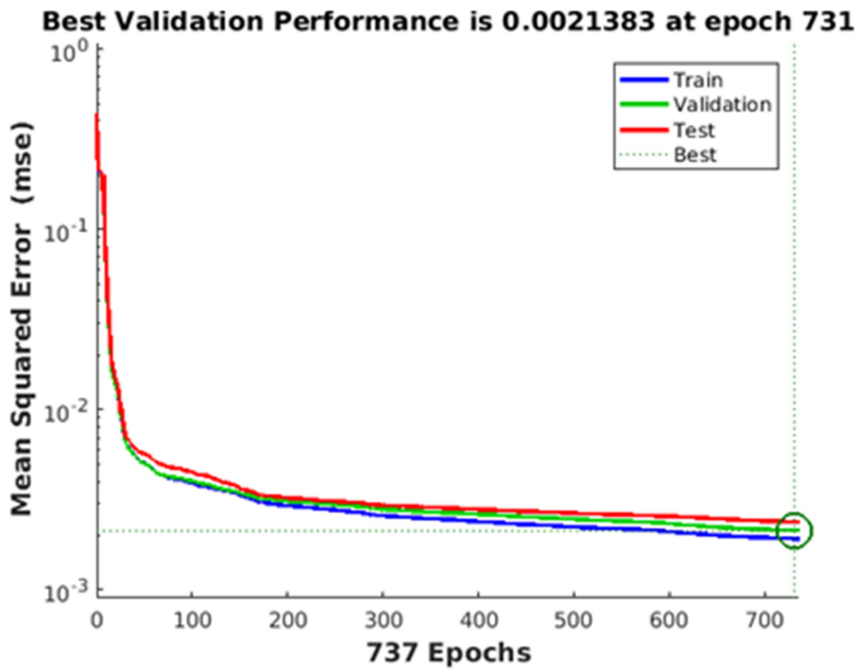
Fonte: Dados da pesquisa do autor

Figura 25 – Performance da RNA ao longo das iterações no espaço de cor CbCr.



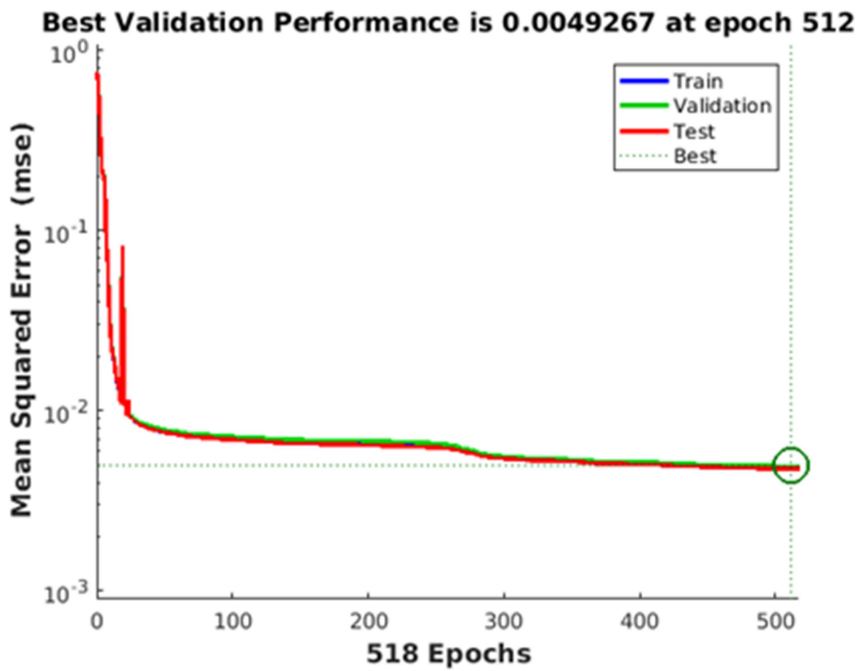
Fonte: Dados da pesquisa do autor

Figura 26 – Performance da RNA ao longo das iterações no espaço de cor HSV.



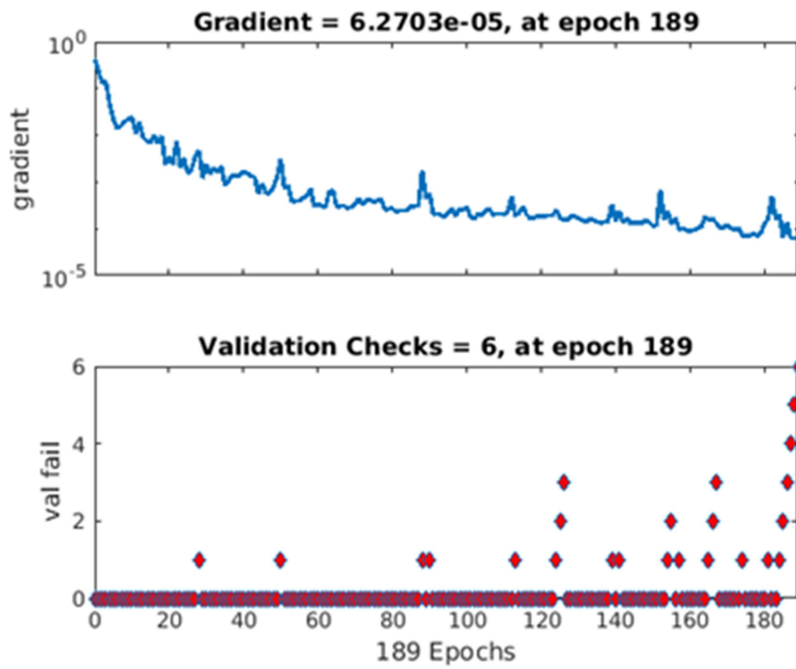
Fonte: Dados da pesquisa do autor

Figura 27 – Performance da RNA ao longo das iterações no espaço de cor CbCr.



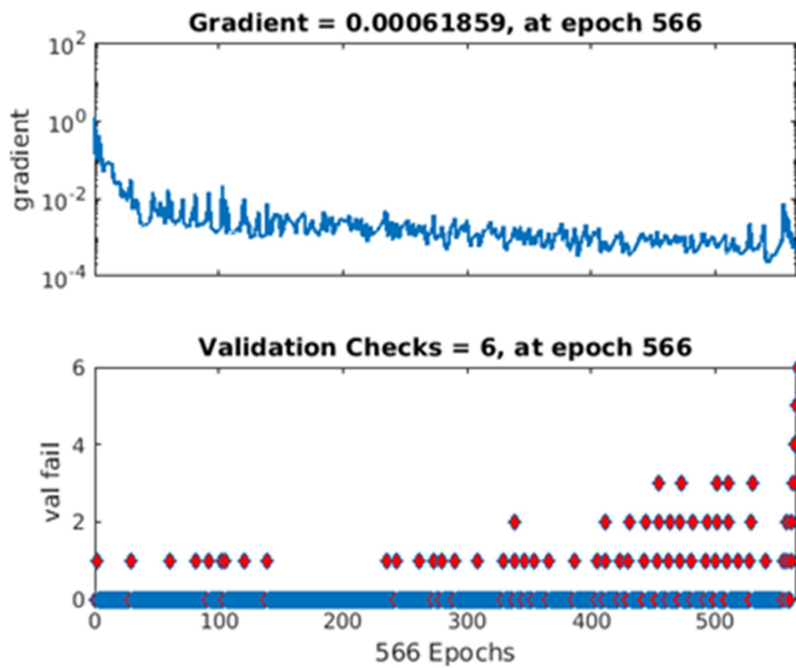
Fonte: Dados da pesquisa do autor

Figura 28 - Gradiente e validação de erro ao longo do treinamento da RNA no espaço de cor RGB.



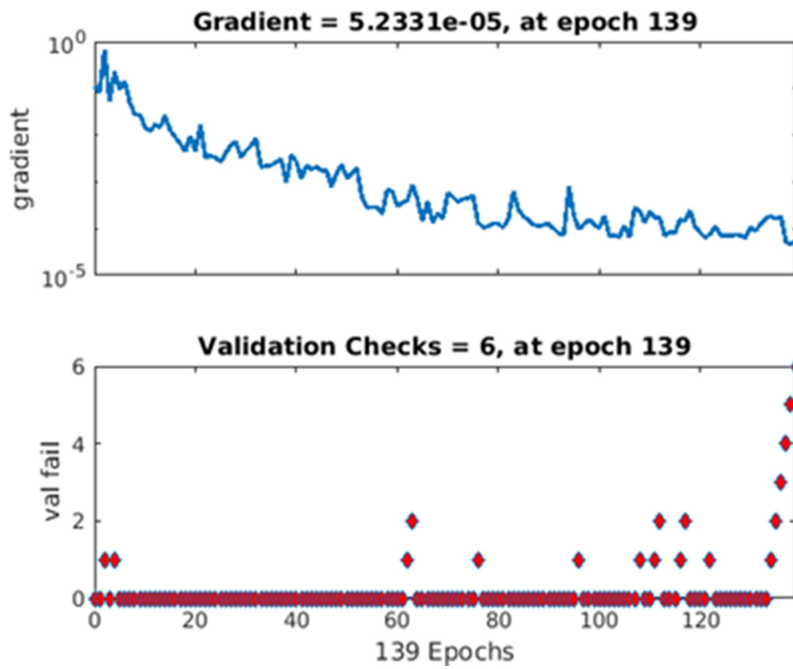
Fonte: Dados da pesquisa do autor

Figura 29 - Gradiente e validação de erro ao longo do treinamento da RNA no espaço de cor Nrg.



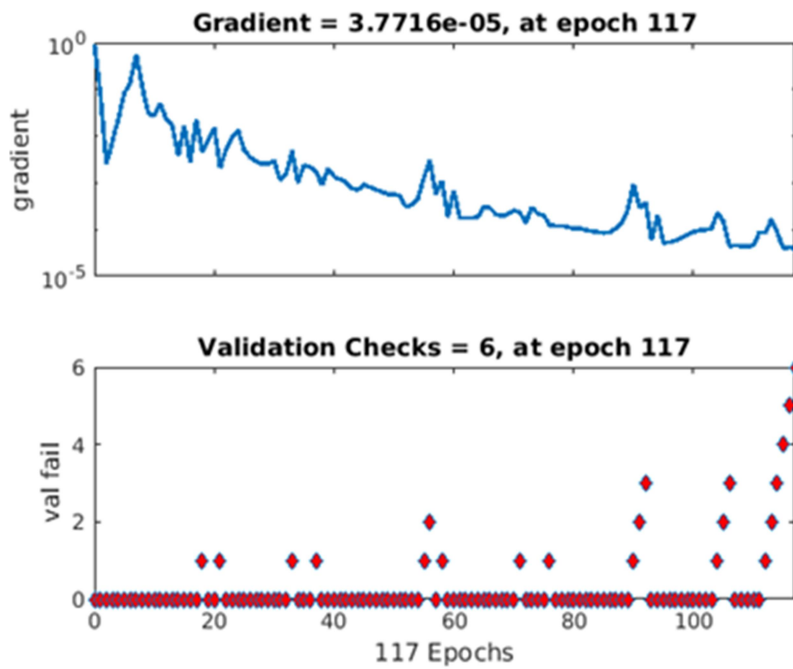
Fonte: Dados da pesquisa do autor

Figura 30 - Gradiente e validação de erro ao longo do treinamento da RNA no espaço de cor YCbCr.



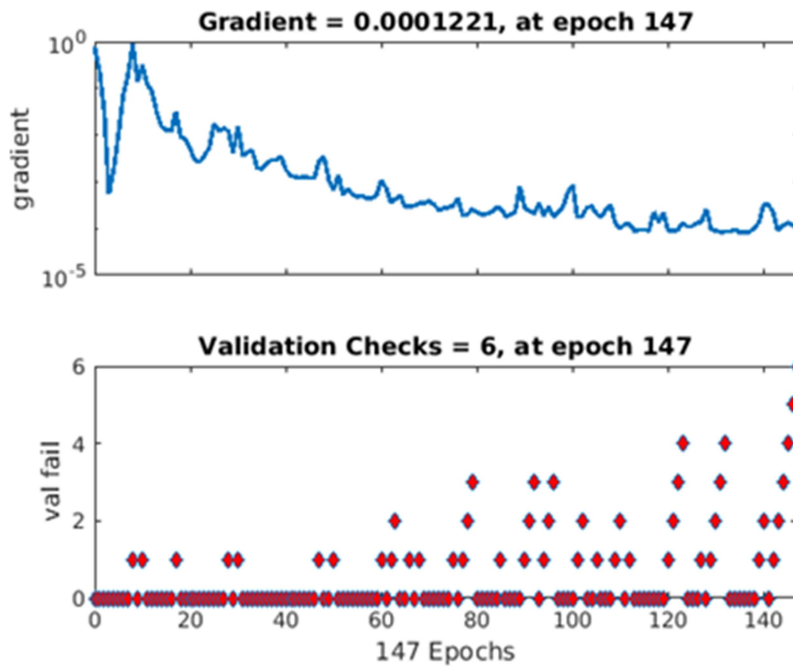
Fonte: Dados da pesquisa do autor

Figura 31 - Gradiente e validação de erro ao longo do treinamento da RNA no espaço de cor RGB.



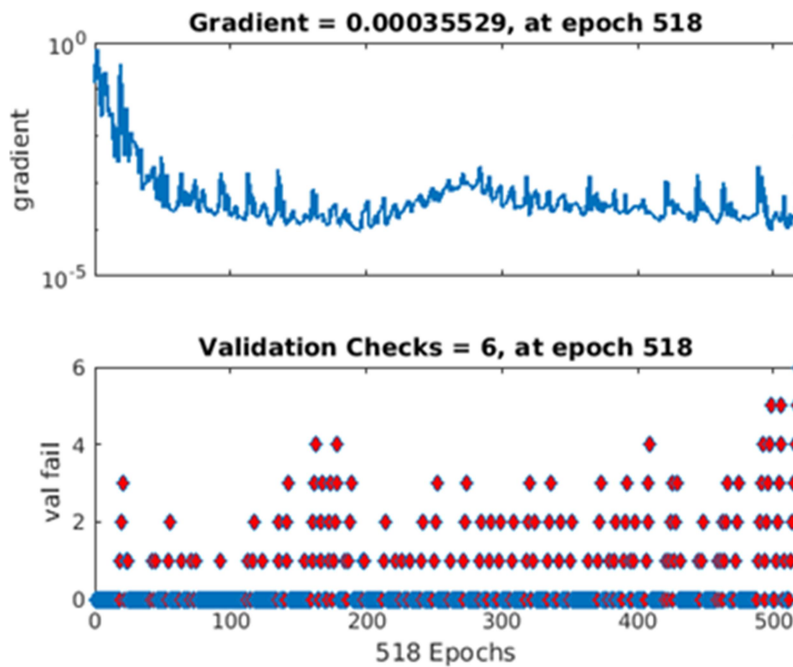
Fonte: Dados da pesquisa do autor

Figura 32 - Gradiente e validação de erro ao longo do treinamento da RNA no espaço de cor HSV.



Fonte: Dados da pesquisa do autor

Figura 33 - Gradiente e validação de erro ao longo do treinamento da RNA no espaço de cor HS.



Fonte: Dados da pesquisa do autor

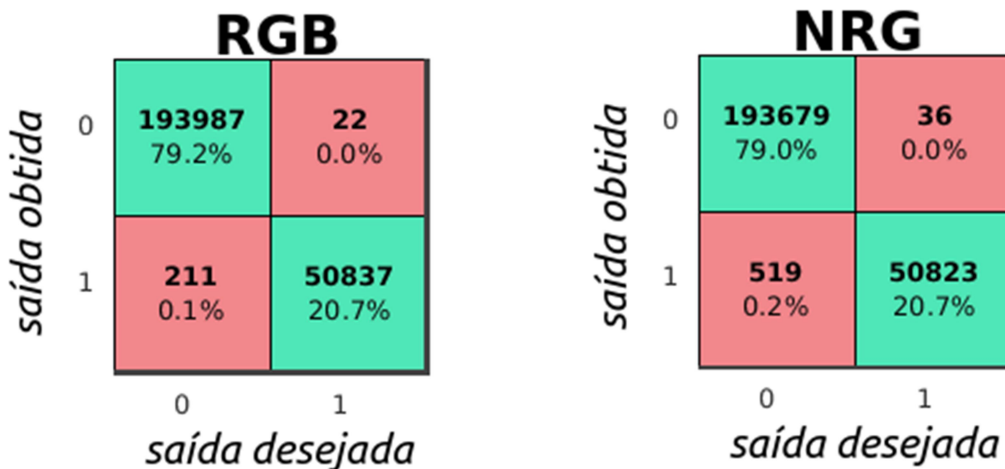


#### 4.1.4 Uso da Rede Neural Artificial

Para acerrar se as RNAs treinadas são capazes de classificar tonalidades de pele, inicialmente foi utilizado o próprio *Skin dataset*, apresentando somente os valores de entrada. A classificação realizada pelas RNAs foi confrontada com a saída desejada e expressas em uma matriz de confusão (Figuras 34 a 36). As matrizes são compostas por linhas que correspondem a *classificação realizada pela RNA (output class)* e colunas que correspondem a *saída desejada (target class)*. As células na diagonal verde mostram a quantidade (e porcentagem) dos exemplos corretamente classificados, ou seja, casos *verdadeiros negativos* e *verdadeiros positivos* respectivamente. A diagonal vermelha mostra os erros de classificação para cada classe (BEALE; HAGAN; DEMUTH, 2016).

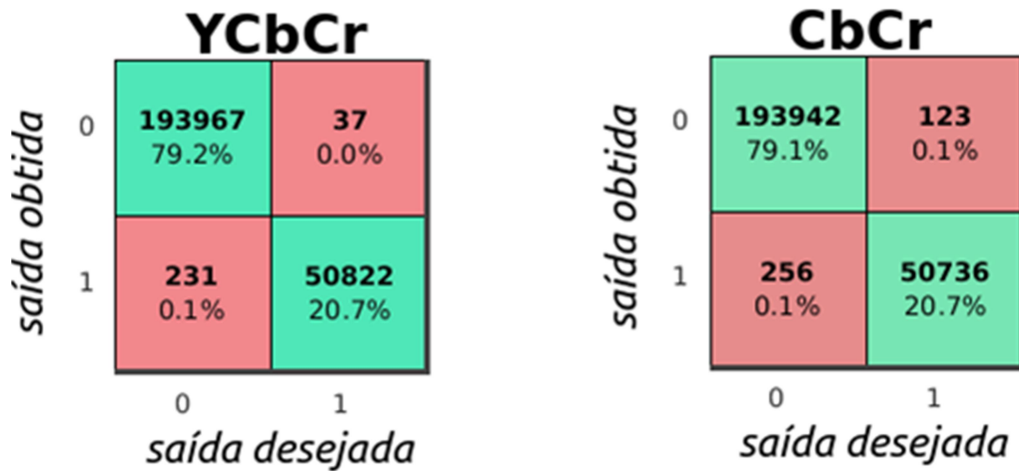
Embora a diferença da precisão geral entre as RNAs seja mínima frente ao grande número de exemplos que compõem o *dataset* (245.057), o espaço de cor RGB apresentou menor erro de classificação com 211 casos *falsos positivos* (classificar como *pele* um pixel *não pele*) e 22 casos *falsos negativos* (classificar como *não pele* um pixel de *pele*). A normalização do espaço RGB para Nrg resultou em um significativo aumento nos casos de falsos positivos, subindo para o total de 519.

Figura 34 - Matriz de confusão *saída desejada x saída obtida* dos espaços de cor RGB (esquerda) e Nrg (direita).



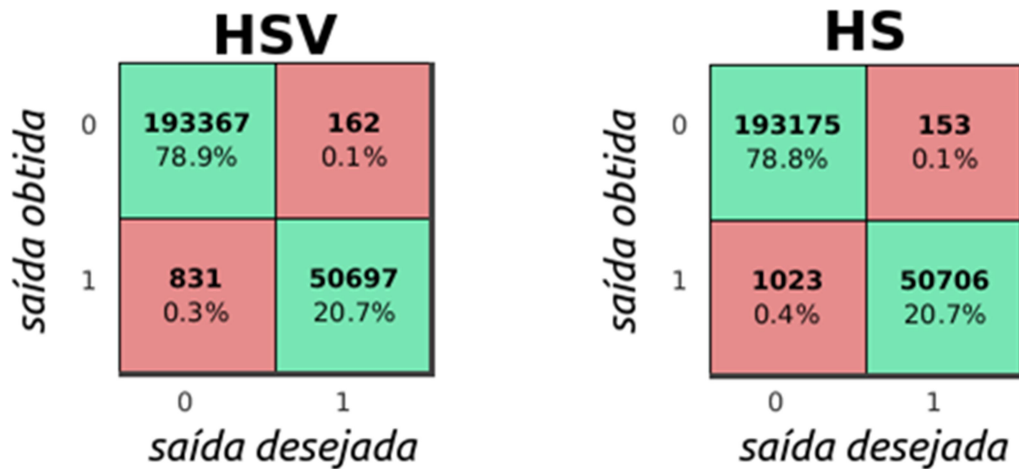
Fonte: Dados da pesquisa do autor

Figura 35 - Matriz de confusão *saída desejada x saída obtida* dos espaços de cor YCbCr (esquerda) e CbCr (direita).



Fonte: Dados da pesquisa do autor

Figura 36 - Matriz de confusão *saída desejada x saída obtida* dos espaços de cor HSV (esquerda) e HS (direita).



Fonte: Dados da pesquisa do autor

O espaço de cor YCbCr apresentou resultados semelhante ao RGB com 231 casos de falsos positivos e 37 falsos negativos. A exclusão da iluminação com o espaço CbCr fez com que o número de classificações errôneas aumentasse consideravelmente, porém com menos efeito se comparado aos espaços RGB e Nrg.

A matriz de confusão do espaço de cor HSV atesta o elevado número de *erros de*

*validação* durante a fase de treino. Os casos de falsos positivos e falsos negativos totalizam 993 no espaço tridimensional (HSV) e 1176 com a exclusão da iluminação (HS).

Diante da expressiva quantidade de dados fornecidos pelo *dataset*, nota-se que o conhecimento foi transmitido às RNAs, que assimilaram os exemplos apresentados durante a aprendizagem ao expressar precisão entre 99,5% a 99,9% de acertos de classificação.

#### 4.1.4.1 Máscara Binária Produzidas Pelas RNAs

Nesta etapa foram utilizadas imagens como vetores de entrada para as RNAs. Elas são apresentadas e classificadas pixel a pixel. A imagem gerada pela saída da RNA possui somente duas cores: preto para os pixels que foram classificados como *não pele* (valor 0) e branco para os pixels classificados como *pele* (valor 1). Esta imagem é denominada máscara binária.

Cada imagem do COMPAQ *dataset* (JONES; REHG, 2002) possui uma máscara binária resultante da segmentação dos pixels de pele (Figura 37). Estas servirão como referência neste estudo apontando a *saída desejada* da classificação a ser realizada pelas RNAs.

Figura 37 - Imagem original e sua máscara binária.



Fonte: Jones e Rehg (2002)

Das imagens selecionadas foram extraídos vetores com os valores RGB de cada pixel.

Estes foram normalizados para o espaço Nrg e convertidos para o espaço HSV e YCbCr e então foram apresentados como vetor de entrada (*input*) para a RNA do espaço de cor correspondente. Com base no conhecimento adquirido na etapa de treinamento, as Redes Neurais Artificiais classificaram se o pixel contém cor de *pele* (retornando o valor 1 em sua saída) ou *não pele* (valor 0 em sua saída).

Na Figura 38 são apresentadas as máscaras binárias geradas pelas RNAs. A segmentação nos espaços de cor RGB e YCbCr apresentou resultados semelhantes. Constatase que a normalização do RGB e a exclusão do componente de luminância em HSV não proporcionou melhorias no processo de segmentação. O espaço de cor CbCr apresentou elevada taxa de falsos positivos no cabelo. As estampas da camisa, dentes, sombreamento na região do pescoço e da face resultaram em classificações errôneas (falsos positivos e falsos negativos) em todos os espaços de cor. Dentre as saídas obtidas, o espaço de cor HS teve menor precisão.

Figura 38 - Máscara binária gerada pelas RNAs de cada espaço de cor.

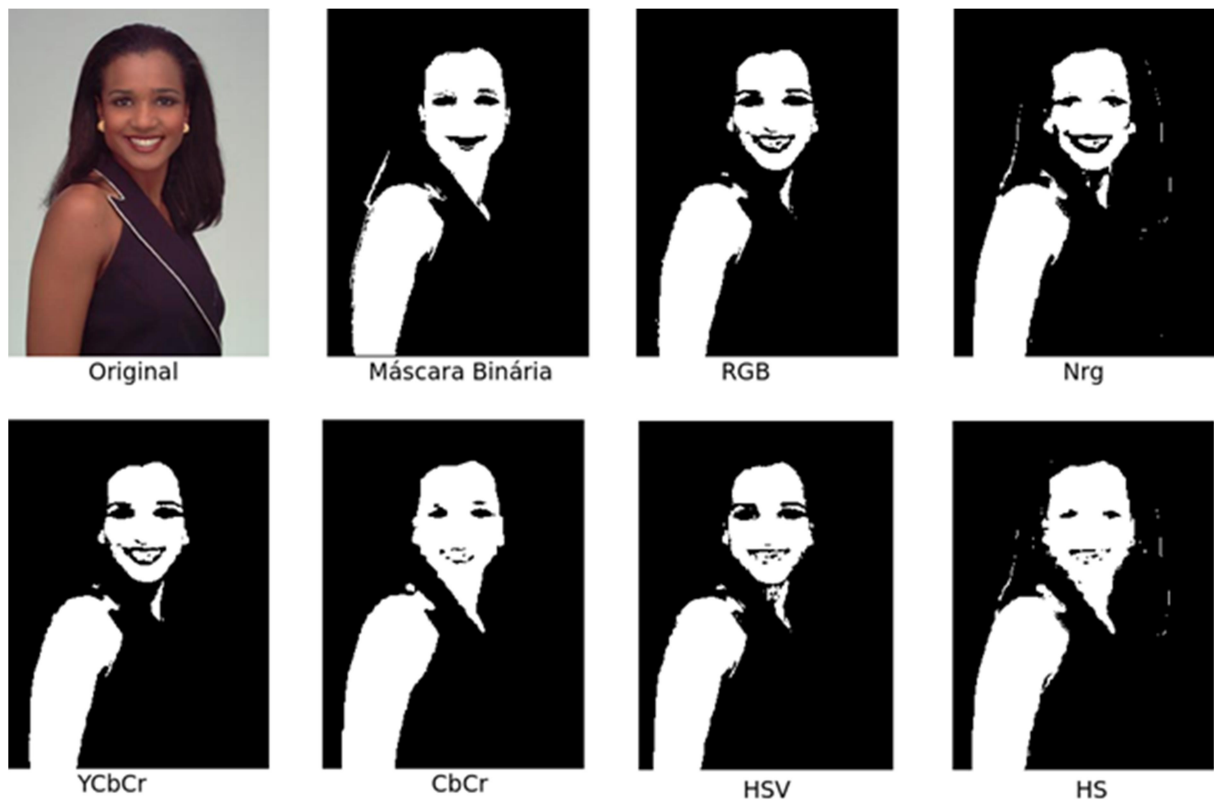


Fonte: Dados da pesquisa do autor

No próximo exemplo (Figura 39) as RNAs tiveram resultados bem próximos à máscara

binária do COMPAQ. Os espaços de cor CbCr e HS, livres da dependência do parâmetro correspondente à luminância, segmentaram corretamente regiões sombreadas do pescoço que foram erroneamente classificadas como *não pele* nos espaços RGB, YCbCr, Nrg e HSV.

Figura 39 - Imagem original, máscara binária (Jones e Rehg, 1998) e saídas geradas pelas RNAs de cada espaço de cor.



Fonte: Dados da pesquisa do autor

As imagens do COMPAQ *dataset* são datadas há mais de 20 anos, portanto optou-se também utilizar fotos recentes coletadas na internet. Foram selecionadas fotos capturadas em áreas externas e em ambientes não controlados, com presença de objetos que possuem a tonalidade visualmente semelhante à pele como madeira e areia.

Inicialmente foi realizado um teste em uma imagem que contém características necessárias para ser classificada como imprópria por conter pessoas nuas (Figura 40). A foto possui diferentes tonalidades de pele e grande quantidade de areia.

Figura 40 - Foto com diferentes tonalidades de pele e a presença de areia.



Fonte: <http://twitter.com/AtarkistAq> (2014).

A classificação dos pixels de pele nos seis cenários foi satisfatória apresentando uma visível segmentação entre as regiões de *pele* e *não pele* (Figuras 41 a 46). A alta detecção de pixels de pele no espaço CbCr (Figura 44) aponta a melhor segmentação dentre os espaços de cor selecionados. Sua segmentação mostra com exatidão as fronteiras entre os pixels de pele e não pele, resultando em uma segmentação “nítida”. Independente da iluminação, obteve-se êxito na identificação de regiões sombreadas como a área dos olhos, pescoço e pernas. O espaço de cor HS (Figura 46) apresentou problemas na identificação entre as fronteiras, resultando em “borrões” na segmentação. RGB (Figura 41), YcbCr (Figura 43) e HSV (Figura 45) realizaram uma segmentação eficiente porém a limitação na identificação de áreas com pouca luminosidade (região dos olhos, pescoço e pernas) acarreta elevado índice de falso negativo, tendo impacto negativo na eficácia na classificação. O espaço Nrg (Figura 42) apresentou instabilidade, com resultados variados para diferentes tonalidades de pele, tornando-o menos recomendado para aplicação.



Figura 41 - Segmentação de pixels de pele no espaço de cor RGB.



Fonte: Dados da pesquisa do autor

Figura 42 - Segmentação de pixels de pele no espaço de cor Nrg.



Fonte: Dados da pesquisa do autor.

Figura 43 - Segmentação de pixels de pele no espaço de cor YCbCr.



Fonte: Dados da pesquisa do autor

Figura 44 - Segmentação de pixels de pele no espaço de cor CbCr.



Fonte: Dados da pesquisa do autor



Figura 45 - Segmentação de pixels de pele no espaço de cor HSV.



Fonte: Dados da pesquisa do autor

Figura 46 - Segmentação de pixels de pele no espaço de cor HS.



Fonte: Dados da pesquisa do autor

No próximo exemplo (Figura 47), as RNAs apresentaram falsos positivos nos objetos de

madeira em todos os cenários (Figuras 48 a 53). A exclusão do parâmetro de luminosidade em Nrg, CbCr e HS não diminuiu os erros de classificação. Embora a segmentação de diferentes tons de pele e regiões sombreadas tenha sido eficaz, o processo é penalizado pela alta presença de falsos positivos ocasionados pelo piso de madeira.

Figura 47 – Imagem original.



Fonte: Ymca of Dayton (2017)

Figura 48 - Segmentação de pixels de pele no espaço de cor RGB.



Fonte: Dados da pesquisa do autor

Figura 49 - Segmentação de pixels de pele no espaço de cor Nrg.



Fonte: Dados da pesquisa do autor

Figura 50 - Segmentação de pixels de pele no espaço de cor YCbCr.



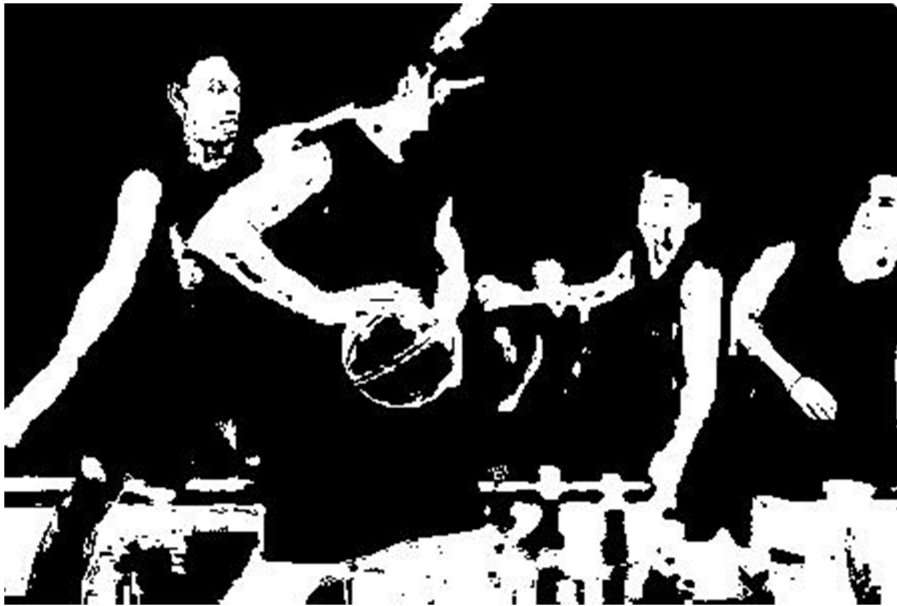
Fonte: Dados da pesquisa do autor

Figura 51 - Segmentação de pixels de pele no espaço de cor CbCr.



Fonte: Dados da pesquisa do autor

Figura 52 - Segmentação de pixels de pele no espaço de cor HSV.



Fonte: Dados da pesquisa do autor

Figura 53 - Segmentação de pixels de pele no espaço de cor HS.



Fonte: Dados da pesquisa do autor

Foram realizados testes buscando avaliar a resposta das RNAs diante de exemplos com diferentes etnias (Figura 54). O espaço de cor CbCr continua apresentando baixo índice de falsos positivos ao segmentar corretamente regiões de pele sombreadas.

Evidenciou-se a precisão nula em todos os cenários diante de grupos étnicos africanos como o Anuak (Figura 54[b]). A deficiência de classificação atribui-se a falta de exemplos de pele destes grupos durante a etapa de treinamento.

Figura 54 - Classificação de pixels de pele em grupos de etnias diversificadas.



Fonte: Elaboração do próprio autor a partir de imagens obtidas na internet

#### 4.2 Estudo de Caso: Classificador de nudez para selfies

Atualmente crianças e adolescentes utilizam *smartphones* e *tablets* para recreação. Estes dispositivos possuem recursos de captura e armazenamento de mídias digitais como vídeos e fotos e também possibilitam a interação com pessoas de todo o mundo por meio da Internet. Longe da atenção dos responsáveis, a utilização indiscriminada destes recursos podem acarretar sérias questões psicológicas no desenvolvimento destas crianças <<http://mdemulher.abril.com.br/familia/o-perigo-de-mostrar-demais-na-internet/>>.

Com base neste cenário é proposto um classificador de *selfies* para *tablets* e *smartphones* para crianças e adolescentes: Uma aplicação que, no momento da captura da foto, realiza a análise em tempo real buscando identificar características de nudez. Caso

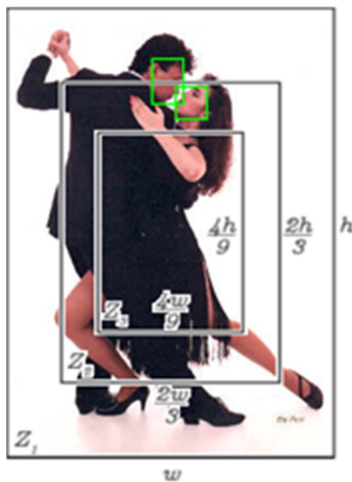


encontre, a foto é imediatamente excluída. Se nenhuma característica de nudez for identificada, a foto estará disponível para visualização e compartilhamento.

Este experimento objetiva *avaliar as características extraídas a partir da máscara binária gerada pela RNA*. Aplica-se o algoritmo de detecção de faces (VIOLA; JONES, 2001) e de zoneamento que divide a imagem em regiões. A divisão da imagem em três regiões concêntricas parte da hipótese de que em imagens com nudez há maior predominância de região da pele no centro da imagem (SANTOS, 2012), (SANDOVAL, 2014).

As características a serem extraídas são (Figura 55): (a) proporção de pixels de pele na imagem, (b) detecção de faces, (c) porcentagem da região de faces em relação ao total da imagem, (d) proporção de pixels na região da face e (e) porcentagem de pixels de pele nas zonas  $Z_1$ ,  $Z_2$  e  $Z_3$  (SANDOVAL, 2014).

Figura 55 - Extração de características do classificador de nudez



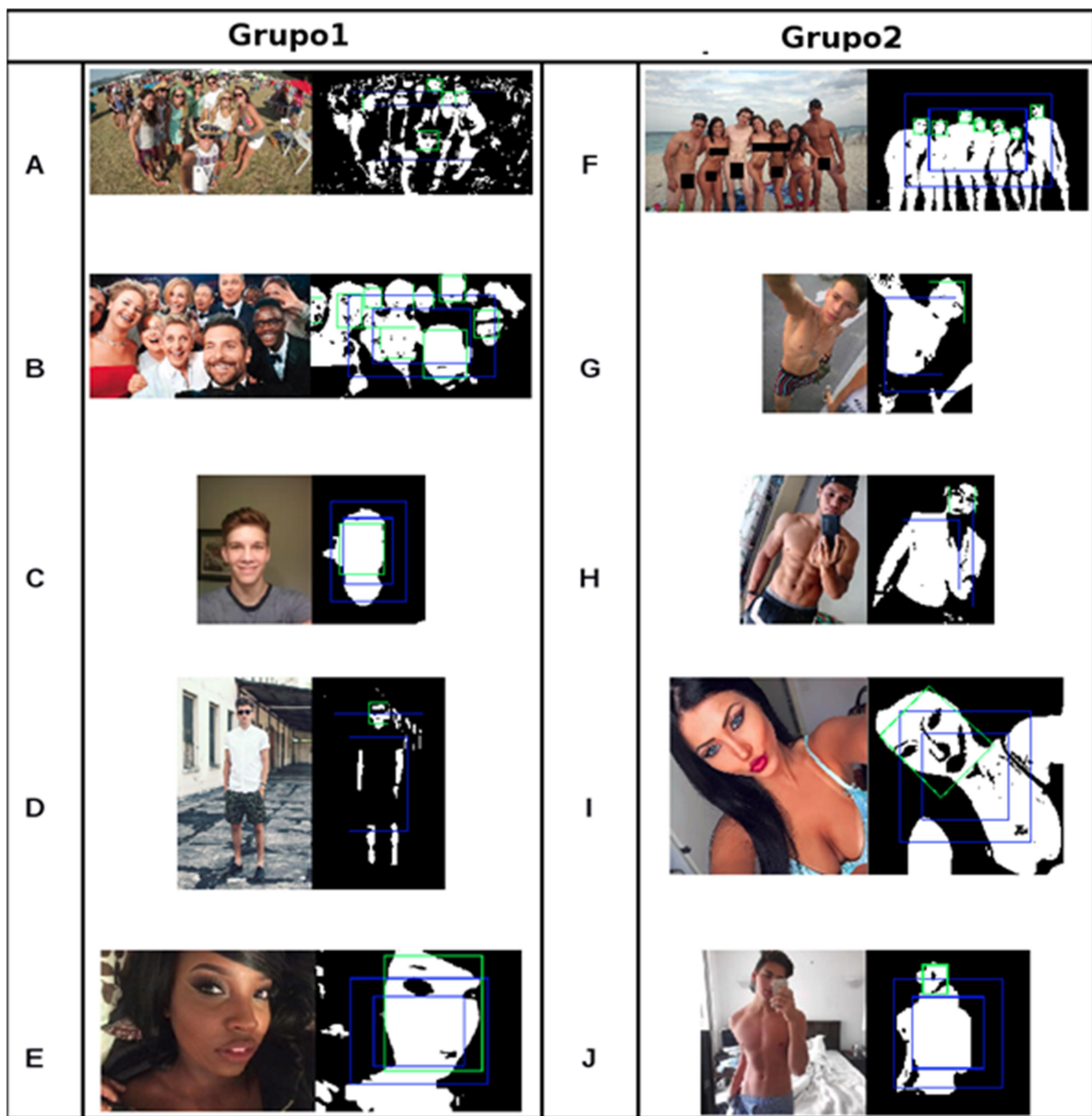
Fonte: Adaptado de Sandoval (2014)

Foram utilizadas dez imagens (Figura 56) divididas em dois grupos: O *grupo1* possui imagens acessíveis às crianças [A] a [E]. As imagens do *grupo2* contém nudez ou características que as tornem proibitivas [F] a [J].

Os resultados obtidos comprovam que a baixa proporção de pixels de *pele de faces* em relação à *totalidade dos pixels de pele* na imagem pode ser um forte indicio de nudez. As fotos de perfil do *grupo1* possuem elevada porcentagem de pele na região da face, resultando

em pouca concentração de pixels de pele nas regiões do corpo. A diminuição da proporção de *pixels de pele e faces* indica maior quantidade de pixels situados na região do corpo, tornando um forte indício de nudez [F], [G], [H] [I] e [J]. Com base nesta proporção pode-se desenvolver um classificador com perfil permissivo (que permita maior proporção de pixels de pele fora da região da face) ou mais restritivo (que limite a proporção de pixels de pele fora da região da face).

Figura 56 - Segmentação dos pixels de pele e extração de características em *selfies*.



Fonte: Elaboração do próprio autor a partir de imagens obtidas na internet



A Tabela 6 lista as características extraídas nas imagens 56[A] a 56[J]. Dentre as características, destaca-se que imagens do *grupo1* possuem grande concentração de pixels de pele em faces e baixíssima concentração de pele em  $Z_1$  e  $Z_3$ . A principal característica das imagens do *grupo2* é a expressiva concentração de pele em  $Z_3$  (que não está contida em faces) acompanhada de pequenas regiões de faces distribuídas entre  $Z_1$  e  $Z_2$ .

**Tabela 6: Características extraídas a partir da máscara binária das imagens 56[A] a 56[J]**

Figura	% Total pixels de pele	Faces	% Área da Face	% Pixels de pele em Faces	% Pixels de pele em $Z_1$	% Pixels de pele em $Z_2$	% Pixels de pele em $Z_3$ (não faces)
<b>A</b>	22,49 %	9	10,72%	20,60%	13%	34,76%	35,81%
<b>B</b>	44,20%	10	22,93%	51,87%	33%	49%	10,25%
<b>C</b>	22%	1	11,81%	53%	1,5%	25,02%	0%
<b>D</b>	4,4%	1	0,71%	16%	1,2%	8,4%	0%
<b>E</b>	39,85%	1	27,43%	68,84%	26,69%	36,86%	0%
<b>F</b>	36,97%	7	2,31%	6,26%	21,20%	51,52%	67,51%
<b>G</b>	40,59%	1	8,04%	16,82%	24,98%	44,81%	81,79%
<b>H</b>	34,86%	1	2,78%	7,9%	10,47%	47,77%	89,87%
<b>I</b>	45,86%	1	11,41%	24,87%	38,46%	55,31%	57,14%
<b>J</b>	24,21%	1	2,74%	11,33%	3,81%	29,91%	78,02%

Fonte: Dados da pesquisa do autor

Os dados obtidos compõem o vetor de entrada e saída (conjunto de treinamento) de um classificador que, a partir das características apresentadas, criará a fronteira de decisão entre os dois grupos de imagens (*nudez e não nudez*).

Como todo filtro em tempo real de baixo custo de processamento, trata-se de um método primário que possui algumas limitações. Fotos que possuam pixels de pele mas nenhuma face identificada seriam consideradas como impróprias (dada a possibilidade de se tratar de uma foto em uma parte específica do corpo), assim como fotos tiradas em trajes de banho, por possuir elevada proporção de pixels de pele que não estão na face, resultando em elevados casos de falsos positivos. Considerando que o objetivo do filtro é cercar a

possibilidade de que uma criança ou pré-adolescente registre e compartilhe fotos que de alguma forma possam ter alguma conotação inapropriada torna-se uma solução aplicável.

## 5 CONCLUSÃO

Neste trabalho foi proposto um classificador de pixels com *cor de pele* utilizando Redes Neurais Artificiais treinadas com exemplos do *Skin Segmentation Dataset* utilizando o algoritmo *Resilient Backpropagation* para treinamento. Na etapa de desenvolvimento foram selecionados três espaços de cor em seis cenários: RGB, Nrg, YCbCr, CbCr, HSV e HS.

Durante a etapa de treinamento (aprendizagem) foram apresentados exemplos fornecidos pelo *Skin Segmentation Dataset*. Constata-se que o conhecimento representado pelos padrões foi transmitido às RNAs, tornando-as capazes de realizar a segmentação dos pixels de pele com baixo custo de tempo e processamento (Tabela 7).

Quanto à máscara binária, espaço RGB e YCbCr apresentaram saídas similares ao produzi-las com nitidez, entretanto demonstra dificuldades em segmentar áreas sombreadas como a região dos olhos e pescoço.

A normalização do espaço RGB não resultou em melhoria na produção da máscara binária se comparada às produzidas pelo RGB.

O espaço de cor HSV registrou menor eficácia dentro do modelo proposto. Similar às informações obtidas no gráfico de performance e matriz de confusão durante a etapa de treinamento, o uso deste espaço de cor resultou em classificações imprecisas produzindo máscaras binárias fragmentadas. Embora a exclusão do parâmetro de luminância resulte em melhor eficácia na segmentação em áreas sombreadas, a elevada taxa de falsos positivos em objetos do cenário torna este espaço de cor inviável para aplicação.

As melhores segmentações foram produzidas pela RNA no espaço de cor CbCr. A exclusão do parâmetro de luminância Y proporcionou maior precisão na segmentação de áreas sombreadas. Suas máscaras binárias delimitam distintamente a fronteira entre a pele dos outros elementos da imagem, tornando este espaço de cor aplicável para o classificador.

Por fim, no estudo de caso, a máscara binária produzida pela RNA no espaço de cor CbCr viabilizou a extração de características para classificação de nudez em fotos.

**Tabela 7: Tempo decorrido para criação da máscara binária nas imagens utilizadas na etapa de teste (Capítulo 4) em um computador Intel Core 2 duo 3.06GHz 4GB memória RAM.**

Figura	Quantidade de Pixels	Tempo para classificação (em segundos)					
		RGB	Nrg	YCbCr	CbCr	HSV	HS
<b>37</b>	85.974	0,325	0,277	0,286	0,258	0,460	0,383
<b>39</b>	62.208	0,298	0,365	0,287	0,226	0,288	0,228
<b>40</b>	159.000	0,275	0,328	0,205	0,333	0,317	0,310
<b>47</b>	135.000	0,370	0,362	0,347	0,355	0,358	0,373
<b>54[a]</b>	120.000	0,397	0,255	0,241	0,226	0,231	0,234
<b>54[b]</b>	60.000	0,187	0,153	0,144	0,143	0,147	0,158
<b>54[c]</b>	422.136	0,773	0,733	0,672	0,637	0,672	0,675
<b>54[d]</b>	216.600	0,515	0,450	0,407	0,358	0,386	0,347
<b>54[e]</b>	119.784	0,355	0,292	0,275	0,227	0,270	0,249
<b>56[A]</b>	921.600	1,457	1,483	1,367	1,355	1,425	1,334
<b>56[B]</b>	159.900	0,270	0,335	0,300	0,304	0,312	0,294
<b>56[C]</b>	30.000	0,116	0,159	0,129	0,121	0,107	0,129
<b>56[D]</b>	146.705	0,261	0,307	0,305	0,256	0,270	0,263
<b>56[E]</b>	67.500	0,155	0,191	0,222	0,214	0,244	0,205
<b>56[F]</b>	159.000	0,275	0,328	0,205	0,333	0,317	0,310
<b>56[G]</b>	307.200	0,509	0,564	0,536	0,486	0,503	0,498
<b>56[H]</b>	112.500	0,234	0,246	0,239	0,275	0,233	0,234
<b>56[I]</b>	384.400	0,614	0,689	0,637	0,610	0,621	0,672
<b>56[J]</b>	120.000	0,101	0,165	0,116	0,127	0,111	0,138

Fonte: Dados da pesquisa do autor

## 5.1 Observações e tópicos para próximas pesquisas

Dentre os elementos que induziram a erros de classificação destacam-se:

- Dentes.
- Cabelo de tonalidade clara.
- Tons de pele de grupos étnicos africanos como Zulu, Xhosa e Anuak.
- Madeira

O *Skin Segmentation Dataset*, utilizado como modelo no processo de aprendizagem das RNAs foi gerado a partir das imagens do banco de dados FERET. Estas foram registradas entre os anos 1993 e 1996, ou seja, há mais de 20 anos. Desde então os equipamentos e recursos utilizados para captura de imagens tiveram grande avanço, assegurando melhor qualidade e maior quantidade de informações em uma imagem (Figura 38). Partindo deste pressuposto, para obtenção de melhores resultados, é recomendável a utilização de um *dataset* composto por imagens obtidas recentemente por dispositivos de captura atuais.

Figura 57 - Evolução da qualidade e nível de detalhamento entre uma fotografia registrada em 1998 (esquerda) e nos dias atuais (direita).



Fonte: Jones e Rehg (2002), Shutterstock (2017)

O classificador de nudez em *selfies* (capítulo 4.3.3) será capaz de produzir saídas

otimizadas se o *dataset* de treinamento for composto por imagens capturadas com dispositivos que possuam *hardware* com especificações semelhantes aos do *tablet/smartphone*. Em outras palavras, quanto mais próxima forem as condições de captura com os exemplos apresentados durante o treinamento da RNA, maior será a precisão no processo de classificação de pixels *pele e não pele*.

Para controlar a taxa de falsos positivos, torna-se viável inclusão de técnicas de processamento de imagens digitais como descritores de forma para reconhecimento de padrões do corpo humano (identificação de membros isolados ou em conjunto) aumentando o custo de tempo e processamento.

A utilização do classificador de pixels de pele não se limita exclusivamente a identificadores de nudez. A máscara binária pode ser aplicada em sistemas de vigilância em tempo real, tratamento de fotos artísticas, identificação automatizada de pornografia infantil, dentre outras aplicações que utilizem como recurso o reconhecimento de pele humana.

O desdobramento deste trabalho apontou a necessidade do desenvolvimento de um *dataset* com amostras de pixels de *pele e não pele* composto por imagens registradas com equipamentos atuais, dado que os *datasets* disponibilizados (e de grande importância na área de segmentação de pele) são datadas entre 1993 (FERET *dataset*) e 2002 (COMPAQ *dataset*). A incapacidade de segmentar pixels com tons de pele de determinados grupos étnicos aponta a escassez de exemplos resultando em não generalização da RNA nessas situações. Portanto, o novo modelo de *dataset* deve também conter amostras de grupos étnicos ainda mais diversificados.

## REFERÊNCIAS

- ABDULLAH-AL-WADUD, M.; CHAE, O. Skin segmentation using color distance map and water-flow property. In: INTERNATIONAL CONFERENCE ON INFORMATION ASSURANCE AND SECURITY, 4, 2008, Naples. **Proceedings of the...** Naples: [S.n.], 2008. p. 83-88.
- ALBUQUERQUE, M. P. de. **Processamento de imagens: métodos e análises**. 2000. Disponível em: <<http://www.cbpf.br/~mpa/curso2.htm>>. Acesso em: 25 ago. 2016.
- ALMOHAIR, H. K. et al. Skin detection in luminance images using threshold technique. **International Journal of the Computer, The Internet And Management**, Bangkok, v. 15, n. 1, p. 25-32, 2007.
- AL-MOHAIR, H. K.; MOHAMAD-SALEH, J.; SUANDI, S. A. Human skin color detection: A Review on Neural Network perspective. **International Journal Of Innovative Computing, Information And Control**, Fukuoka, v. 8, n. 12, p.8115-8131, 2012.
- AL-TAIRI, Zaher Hamid et al. Skin Segmentation Using YUV and RGB Color Spaces. **Journal Of Information Processing Systems**, Seoul, v. 10, n. 2, p. 283-299, 2014.
- ANWAR, N.; RAHMAN, A.; CHONG, K.; JOHN, S. **RGB-H-CbCr skin color model for human face detection**. 2006. Disponível em: <[http://pesona.mmu.edu.my/~johnsee/research/papers/files/rgbhcbcr\\_m2usic06.pdf](http://pesona.mmu.edu.my/~johnsee/research/papers/files/rgbhcbcr_m2usic06.pdf)>. Acesso em: 12 out. 2017.
- BACKES, A. R. **Estudos de métodos de análise de complexidade em imagens**. 2010. 181 f. Tese (Doutorado) - Curso de Ciências de Computação e Matemática Computacional, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2010.
- BASTOS, I. L. O.; ANGELO, M. F. Reconhecimento de pele em imagens digitais utilizando redes neurais artificial. In: WORKSHOP DE VISÃO COMPUTACIONAL, 10, 2014, Uberlândia. **Anais...** Campo Grande: UCDB, 2014. p. 125 – 130.
- BEALE, M. H.; HAGAN, M. T.; DEMUTH, H. B. **Neural network toolbox user's guide**. Disponível em: <<https://www.mathworks.com/help/nnet/>>. Acesso em: 15 mar. 2016.
- BELÉM, R. J. S. **Deteção automática de conteúdo ofensivo na Web**. 2006. 52 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Instituto de Computação, Universidade Federal do Amazonas, Manaus, 2006.
- BHATT, R. B. et al. Efficient skin region segmentation using low complexity fuzzy Decision Tree Model. **India Conference (indicon)**. Ahmedabad, p. 1-4. dez. 2009.
- BHOYAR, K.; KAKDE, O. Skin color detection model using neural networks and its performance evaluation. **Journal Of Computer Science**, Patiala, v. 6, n. 9, p. 963-968, 2010.
- BRITO BISNETO, C. R. de. **Reconhecimento de objetos utilizando redes neurais artificiais e geometria fractal**. 2011. 92 f. Dissertação (Mestrado) - Curso de Modelagem

Computacional e tecnologia Industrial, Senai Cimatec, Salvador, 2011.

BRAND, J.; MASON, J. S. A comparative assessment of three approaches to pixel-level human skin-detection. In: INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION. ICPR-2000, 15, 2001, Washington. **Proceedings of the...** Washington: IEEE, 2001. v. 15, n. 1, p. 1056-1059.

CHURCHLAND, P. S.; SEJNOWSKI, T. J. **The computational brain**. Cambridge: Mit Press Cambridge, 1994.

DORSEY, E. **1999 rochester institute of technology**: introduction to digital imaging. 2015. Disponível em: <<http://slideplayer.com/slide/7598271/>>. Acesso em: 05 nov. 2016.

ELGAMMAL, A.; MUANG, C.; HU, D. Skin detection: s short tutorial. **Encyclopedia Of Biometrics**. [S.l.: S.n.], 2009. p. 1218-1224.

ERPEN, L. R. C. **Reconhecimento de padrões em imagens por descritores de forma**. 2004. 113 f. Dissertação (Mestrado) - Curso de Pós-graduação em Computação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2004.

FLECK, M. M.; FORSYTH, D. A.; BREGLER, C. Finding naked people. **Lecture Notes In Computer Science**, Cambridge, v. 1065, n. 1, p.593-602, 1996. Disponível em: <[http://dx.doi.org/10.1007/3-540-61123-1\\_173](http://dx.doi.org/10.1007/3-540-61123-1_173)>. Acesso em: 12 set. 2017.

GONZALEZ, R. C.; WOODS, R. **Digital image processing**. 2. ed. Boston: Addison-wesley Longman Publishing, 2001.

GOUVEIA, W. da R. **Detecção de faces humanas em imagens coloridas utilizando redes neurais artificiais**. 2010. 112 f. Dissertação (Mestrado) - Curso de Engenharia Elétrica, Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2010.

HAYKIN, S. **Neural networks**: a comprehensive foundation. New York: Cambridge University Press, 1999.

JEONG, Chi-yoon; KIM, Jong-sung; HONG, Ki-sang. Appearance-based nude image detection. In: INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION, ICPR2004, 17, 2004, Cambridge. **Proceedings...** Cambridge: IEEE, 2004. v. 4, n. 4, p. 467-470. Disponível em: <<http://dx.doi.org/10.1109/icpr.2004.1333803>>. Acesso em: 07 out. 2017.

JONES, M. J.; REHG, J. M. Statistical color models with application to skin detection. In: 1999 IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (cat. no Pr00149), 2000, Fort Collins. **Proceedings of the...** Fort Collins: IEEE, v. 46, n. 1, p. 81-96, 2002. Disponível em: <<http://dx.doi.org/10.1109/cvpr.1999.786951>>. Acesso em: 15 ago. 2017.

KOVAC, J.; PEER, P.; SOLINA, F. Human skin color clustering for face detection. In: THE IEEE REGION 8 EUROCON 2003: COMPUTER AS A TOOL, 2003, Ljubljana. **Proceedings of the...** Ljubljana: IEEE, 2003. p. 144-148.



MCCAFFREY, J. **How to use resilient back propagation to train neural networks**. 2015. Disponível em: <<https://visualstudiomagazine.com/articles/2015/03/01/resilient-back-propagation.aspx>>. Acesso em: 20 maio 2016.

MCCAFFREY, J. **Neural network train-validate-test stopping**. 2015. Disponível em: <<https://visualstudiomagazine.com/articles/2015/05/01/train-validate-test-stopping.aspx>>. Acesso em: 4 set. 2015.

MEHTA, H. **Communication medium in modern tv**. 2014. Disponível em: <<https://www.slideshare.net/HARRY-MEHTA/communication-medium-in-modern-tv>>. Acesso em: 03 fev. 2017.

METLEK, S. **Production line of the colored cylinder parts machine defining the vision system and classification**. 2009. Dissertation (Master) - Suleyman Demirel University, Isparta, 2009.

OLIVEIRA, J. de A. et al. Extração de atributos de forma para classificação de imagens de alta resolução do satélite HRC/CBERS-2B. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 14, 2009, São José dos Campos. **Anais...** São José dos Campos: [S.n.], 2009. p. 7015-7022.

PINTO, J. C. L. **Sistema reconhecedor de placas de veículos usando técnicas de processamento digital de imagens**. 2009. 43 f. Trabalho de Conclusão de Curso (Graduação) - Curso de Pesquisa e Desenvolvimento de Software, Instituto Municipal de Ensino Superior de Assis, Assis, 2009.

POOLE, D. L.; MACKWORTH, A. K. **Artificial intelligence: foundations of computational agents**. 2017. Disponível em: <<http://artint.info/2e/html/ArtInt2e.html>>. Acesso em: 03 fev. 2017.

PREMA, C.; MANIMEGALAI, D. Survey on skin tone detection u sing color spaces. **International Journal of Applied Information Systems**, New York, v. 2, n. 2, p. 18-26, 2012.

RIEDMILLER, M.; BRAUN, H. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. In: IEEE INTERNATIONAL CONFERENCE ON NEURAL NETWORKS, 1993, San Francisco. **Proceedings of the...** San Francisco: IEEE, 1993. v. 1, n. 1, p. 586-591. Disponível em: <<http://dx.doi.org/10.1109/icnn.1993.298623>>.

SANDOVAL, D. A. M. **Método para detecção automática de nudez em imagens digitais utilizando redes neurais artificiais**. 2014. 73 f. Trabalho de Conclusão de Curso (Graduação) - Faculdade de Tecnologia, Universidade de Brasília, Brasília, DF, 2014.

SANTOS, C. A. M. dos. **Análise de características para detecção de nudez em imagens**. 2012. 87 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Universidade Federal do Amazonas, Manaus, 2012.

SCHMITT, P. **Redes neurais artificiais aplicadas na classificação litológica das formações Palermo e Rio Bonito na jazida do Leão - RS, com base em perfis geofísicos**. 2009. 94 f. Dissertação (Mestrado) - Curso de Geologia, Universidade do Vale do Rio dos Sinos, São

Leopoldo, 2009.

SHENDE, A. **RGB normalization**. 2013. Disponível em: <<http://akash0x53.github.io/blog/2013/04/29/RGB-Normalization/>>. Acesso em: 10 out. 2017.

SHIN, M. C.; TSAP, L. V. Does colorspace transformation make any difference on skin detection? In: WORKSHOP ON APPLICATIONS OF COMPUTER VISION, 6, 2002. Orlando. **Proceedings of the...** Orlando: IEEE, 2002. p. 275-279.

SKARBEK, W.; KOSCHAN, A. **Color image segmentation: a survey**. Berlin: Institute for Technical Informatics, Technical University of Berlin, 1994. Technical report.

SENA JÚNIOR, D. G. de. **Sistema de visão artificial para identificação de plantas de milho atacadas pela lagarta do cartucho (spodoptera frugiperda)**. 2002. 81 f. Tese (Doutorado) - Curso de Engenharia Agrícola, Universidade Federal de Viçosa, Viçosa, 2002.

SURAMPALLI, G. P.; DAYANAND, J.; DHANANJAY, M. An analysis of skin pixel detection using different skin color extraction techniques. **International Journal of Computer Applications**, New York, v. 54, n. 17, p.1-5, 2012. Disponível em: <<http://dx.doi.org/10.5120/8655-2385>>. Acesso em: 02 set. 2017.

TAGLIAFERRO, F. A. **Técnicas de reconhecimento automatizado de padrões baseados em imagens digitais**. 2007. 44 f. Trabalho de Conclusão de Curso (Graduação) - Curso de Ciência da Computação, Faculdade de Jaguariúna, Jaguariúna, 2007.

TERRILLON, J. C. et al. Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. In: PROCEEDINGS FOURTH IEEE INTERNATIONAL CONFERENCE ON AUTOMATIC FACE AND GESTURE RECOGNITION, 4, 2000, Grenoble. **Proceedings of the...** Grenoble: IEEE, 2000. p. 54-61.

THAKUR, Sayantan et al. Face detection using skin tone segmentation. In: WORLD CONGRESS ON INFORMATION AND COMMUNICATION TECHNOLOGIES, 2011, Mumbai. **Proceedings of the...** Mumbai: IEEE, 2012. p. 53-60.

TWITTER. **Naked among clothed**. 2014. Disponível em: <<https://twitter.com/StarkistAq/status/537968398009765888>>. Acesso em: 10 out. 2016.

VEZHNEVETS, V.; SAZONOV, V.; ANDREEVA, A. A survey on pixel-based skin color detection techniques. INTERNATIONAL CONFERENCE GRAPHICON, 2003, Moskow. **Proceedings of the...** Moskow: IEEE, 2003. p. 85-92. Disponível em: <<http://graphicon.ru/oldgr/en/publications/text/gc2003vsa.pdf>>. Acesso em: 25 jun. 2017.

VIOLA, P.; JONES, M. J. Robust real-time face detection. **International Journal of Computer Vision**, Vancouver, v. 57, n. 2, p. 137-154, 2004.

YANG, G. et al. Research on a skin color detection algorithm based on self-adaptive skin color model. INTERNATIONAL CONFERENCE ON COMMUNICATIONS AND INTELLIGENCE INFORMATION SECURITY, 2010, Nanning. **Proceedings of the...**

Nanning: IEEE, 2010. p. 266-270.

YANG, J.; LU, W.; WAIBEL, A. Skin-color modeling and adaptation. 1998. Disponível em: <[https://www.ri.cmu.edu/pub\\_files/pub1/yang\\_jie\\_1998\\_1/yang\\_jie\\_1998\\_1.pdf](https://www.ri.cmu.edu/pub_files/pub1/yang_jie_1998_1/yang_jie_1998_1.pdf)>. Acesso em: 20 ago. 2017.

YMCA OF DAYTON. **Basketball**. Disponível em: <<https://www.ymcaonline.org/programs/basketball>>. Acesso em: 17 abr. 2017.

ZARIT, B. D.; SUPER, B. J.; QUEK, F. K. H. Comparison of five color models in skin pixel classification. In: WORKSHOP ON RECOGNITION, ANALYSIS AND TRACKING OF FACES AND GESTURES IN REAL-TIME SYSTEMS, 1999, Corfu. **Proceedings of the...** Corfu: IEEE, 1999. p. 58-63.