

# Evaluation of normalization technique on classification with deep learning features

André D. Freitas<sup>a</sup>, Adriano B. Silva<sup>a</sup>, Alessandro S. Martins<sup>b</sup>, Leandro A. Neves<sup>c</sup>, Thaína A. A. Tosta<sup>d</sup>, Paulo R. de Faria<sup>e</sup>, Marcelo Z. do Nascimento<sup>a</sup>

<sup>a</sup>Computer Science Faculty, Federal University of Uberlândia, Uberlândia, Minas Gerais, Brazil

<sup>b</sup>Federal Institute of Triângulo Mineiro, Ituiutaba, Minas Gerais, Brazil

<sup>c</sup>Department of Computer Science and Statistics, São Paulo State University, S.J.do Rio Preto, São Paulo, Brazil

<sup>d</sup>Science and Technology Institute, Federal University of São Paulo, S.J. dos Campos, São Paulo, Brazil

<sup>e</sup>Department of Histology and Morphology, Institute of Biomedical Science, Federal University of Uberlândia, Uberlândia, Minas Gerais, Brazil

{andredias343}@gmail.com

**Resumo**—Cancer is one of the diseases with the highest mortality rate in the world. Dysplasia is a difficult-to-diagnose precancerous lesion, which may not have a good Hematoxylin and Eosin (H&E) stain ratio, making it difficult for the histology specialist to diagnose. In this work, a method for normalizing H&E stains in histological images was investigated. This method uses a generative neural network based on a U-net for image generation and a PatchGAN architecture for information discrimination. Then, the normalized histological images were employed in classification algorithms to investigate the detection of the level of dysplasia present in the histological tissue of the oral cavity. The CNN models as well as hybrid models based on learning features and machine learning algorithms were evaluated. The employment of the ResNet-50 architecture and the Random Forest algorithm provided results with an accuracy rate around 97% for the images normalized with the investigated method.

**Index Terms**—Normalization, H&E Stain, Classification, Dysplasia, Convolutional Neural Network.

## I. INTRODUÇÃO

O câncer é uma doença genética que surge a partir de uma mutação no DNA, podendo ser causada por mutação hereditária ou desenvolvida ao longo da vida pela exposição a agentes carcinógenos, como a radiação ou o cigarro. O processo de carcinogênese é um processo lento e pode ser dividido em três principais estágios. O primeiro estágio é o de iniciação, em que os genes sofrem ação de algum agente cancerígeno, modificando esses genes. Nesse estágio ainda não é possível identificar clinicamente a presença de lesões como o tumor. O segundo estágio faz com que a célula se transforme em uma célula maligna. Por fim ocorre o estágio de progressão, onde acontece a multiplicação descontrolada e irreversível das células alteradas já instaladas no corpo, podendo evoluir até o surgimento das primeiras manifestações clínicas da doença [1]. Desse modo, se torna importante não apenas desenvolver métodos para o tratamento, mas também para o diagnóstico precoce dessa doença [2]. O câncer da cavidade oral é a sexta maior causa de morte por câncer no mundo, apresentando 50% de chances de sobrevivência em 5 anos após diagnóstico [1].

Uma forma de contribuir para diminuir o número de casos com as lesões do tipo carcinoma na cavidade bucal é analisar e diagnosticar as lesões pré-cancerosas. A displasia, uma lesão pré-câncer, se caracteriza como a ocorrência de anomalias relacionadas ao desenvolvimento de um órgão ou tecido, intimamente relacionadas a alterações genéticas e agressão continuada, alterando o tamanho, cor ou o formato da célula de forma não natural [3], [4]. É um tipo de anormalidade que pode evoluir rapidamente e se tornar um tumor maligno. A análise e identificação de células displásicas é importante para auxiliar o especialista no processo de um diagnóstico precoce. Isso permite que o paciente possa receber o tratamento correto antes que a lesão se torne mais grave ou possa evoluir para um câncer, interrompendo o desenvolvimento da doença antes mesmo do estágio de iniciação [5].

O diagnóstico dos diversos estágios das células displásicas geralmente é uma tarefa muito difícil. Nas últimas décadas, sistemas de apoio ao diagnóstico (do inglês, *computer-aided diagnosis* - CAD) vêm contribuindo ao fornecer uma segunda opinião ao especialista durante a análise do estágio dessa lesão. Um sistema CAD possui diversas etapas para análise de uma imagem histológica. As principais etapas são pré-processamento, segmentação, extração de características e classificação [6], [7]. A etapa de pré-processamento busca melhorar as informações presentes na imagem minimizando ruídos que porventura existam no processo de aquisição ou digitalização da imagem. A segmentação tem como objetivo identificar as diferentes estruturas celulares das amostras, o que é um grande obstáculo em análises computacionais, visto que essa tarefa depende diretamente da qualidade da imagem original e da remoção de ruídos realizada na etapa de pré-processamento. A extração de características permite a representação quantitativa das regiões das imagens histológicas. Por fim, a etapa de classificação utiliza os resultados das etapas anteriores para identificar as classes do problema especificado, como por exemplo, a severidade das lesões da displasia [8].

As redes neurais convolucionais (do inglês, *convolutional neural network* - CNN) são definidas como um tipo de redes

neurais artificiais que apresentam um conjunto de camadas que exploram o aprendizado de modelos mais complexos permitindo explorar etapas de segmentação e classificação de imagens. Há diversas vantagens de uma CNN comparada a métodos tradicionais. Entre suas principais características estão os benefícios de automaticamente aprender características úteis de alto nível diretamente sem ter que extrair as características *handcraft* [9]. Diversos modelos de CNNs vêm sendo empregados na literatura para análise de imagens histológicas em patologia digital. Esses modelos permitem extrair características semânticas de alto nível das imagens histológicas.

A normalização de corantes de Hematoxilina e Eosina (H&E) é uma etapa muito utilizada no pré-processamento feita de forma a manter um bom contraste com a preservação de todas as informações de origem na imagem processada conforme descrito por [10]. Diferentes estratégias de normalização de coloração têm sido propostas para reduzir a inconsistência dos tecidos corados em sistemas automatizados [11]. Na literatura, esses métodos de normalização são classificados em correspondência de histogramas, transferência de cores, correspondência espectral e aprendizagem generativa [12], [13]. Grande parte das técnicas de normalização limitam-se a serem aplicadas com base em uma imagem de referência, que podem resultar em erros de normalização quando não são representativas. A abordagem baseada em aprendizagem trata o problema da normalização como um problema de transferência de estilo, usando redes neurais adversárias generativas (do inglês, *generative adversarial networks* - GANs). Investigar as contribuições de métodos de normalização em uma classificação de tecidos histológicos para lesões e níveis de displasia torna-se relevante para a área em sistemas CAD.

Esse trabalho investiga um método de normalização baseado em GAN para pré-processamento de imagens histológicas de displasia. O método emprega uma rede neural generativa baseada numa arquitetura *U-net* para a geração das imagens e uma arquitetura *PatchGAN* para discriminação das informações. Essas imagens são avaliadas por modelos CNNs para investigar as contribuições em relação ao melhoramento nas fases de treinamento e teste para classificação das lesões. Foram implementados modelos CNNs baseados nas arquiteturas ResNet e AlexNet, assim como modelos híbridos que empregam as *learning features* das camadas de *flattening* das redes com os algoritmos de aprendizado de máquina *Random Forest* (RF) e *Support Vector Machine* (SVM). Os resultados foram analisados com métricas de melhoramento da qualidade das imagens, assim como o desempenho da classificação com as medidas de acurácia, sensibilidade e especificidade.

## II. MATERIAIS E MÉTODOS

### A. Banco de Displasia

O banco de imagens utilizado nesse trabalho foi desenvolvido no estudo proposto em [5]. As imagens são amostras de tecidos de língua de camundongos contendo lesões induzidas experimentalmente utilizando o carcinógeno 4NQO e aprovadas pelo Comitê de Ética na Utilização de Animais. As imagens histológicas digitais foram obtidas com o uso do

microscópio óptico Leica DM500, em magnificação de 400x, e armazenadas em arquivos em formato TIFF, utilizando o modelo de cores RGB e resolução original de 2048×1536 pixels. Essas imagens foram recortadas em regiões de interesse com resolução de 450×250 pixels para áreas de presença da displasia. O banco de imagens possui 296 imagens, divididas igualmente nas classes severa, moderada, leve e saudável.

### B. Método para a Classificação do Nível da Displasia

O sistema proposto é composto de três fases: normalização de corantes H&E, extração de características e classificação. Para a etapa de normalização foi implementado o método de normalização baseado em um modelo GAN. Na extração e classificação foi desenvolvido um modelo híbrido que combina características dos modelos CNN e algoritmos de aprendizado de máquina para a classificação das lesões. Na Figura 1 são exemplificadas as etapas do sistema proposto.

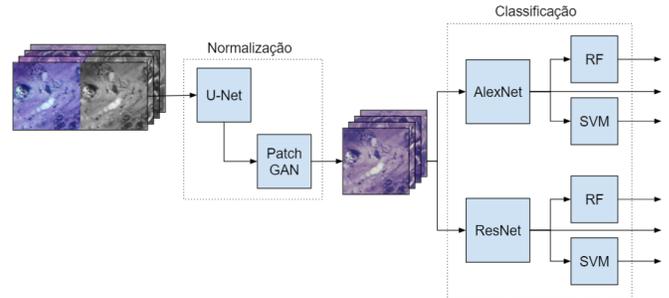


Figura 1: Modelo desenvolvido neste trabalho para classificação das imagens histológicas de displasia e para as características da normalização dos corantes. Utilizando a arquitetura STST para normalização e modelos híbridos para classificação.

O sistema foi desenvolvido utilizando a linguagem Python, e as bibliotecas Keras e Scikit-learn. Os experimentos foram conduzidos em um computador com processador Intel i3-6006U, 4GB de memória RAM e sistema operacional Windows 10.

**Normalização dos corantes H&E:** para realizar a normalização das imagens foi utilizado o modelo denominado *Stain-to-Stain normalization* (STST) proposto por [11]. Na Figura 2 é ilustrada a arquitetura do modelo STST em que são representados os dados como entrada, as imagens reais e suas respectivas versões em níveis de cinza.

Essa abordagem emprega um processo de representação imagem por imagem para a normalização do H&E em imagens histológicas. Para essa tarefa nos modelos tradicionais de normalização, uma imagem de referência deve ser usada para a normalização dos corantes. No modelo, essa etapa ocorre por meio de GANs condicionais (cGANs) definida pelos autores em [14].

A cGAN aprende a mapear por meio de observação em uma imagem  $x$  e um vetor de ruídos randômicos  $z$  para uma

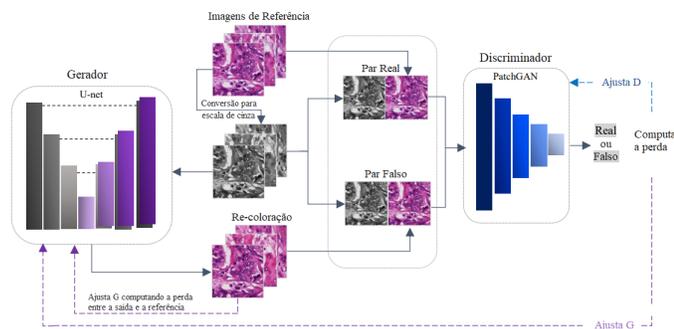


Figura 2: Modelo da GAN para a normalização *Stain-to-Stain* [11]

saída  $y$ ,  $G:z \rightarrow y$ . O gerador  $G$  é treinado para produzir saídas que não podem ser distinguidas das imagens reais por um discriminador  $D$ , o qual é treinado para fazer a detecção das imagens falsas. Essa abordagem emprega uma rede  $U-net$  para a geração das imagens e uma arquitetura  $PatchGAN$  para discriminação das informações. No gerador, as camadas de codificadores e decodificadores são diretamente conectadas por *Skip Connection* (SC) (“Conexões de Salto”). Uma vez que a conexão por SC pode transportar as informações de baixo nível ao longo das camadas da rede codificadora e decodificadora, isso contribui no desempenho da representação dos corantes H&E. No modelo discriminador,  $PatchGAN$ , cada imagem é dividida em segmentos  $n \times n$ , para fazer a predição de cada parte em verdadeiro ou falso. Finalmente, a média de todas as respostas baseadas nos segmentos é calculada para a classificação.

Nesse modelo, pares de imagens devem ser usadas no treinamento, as quais representam a imagem original e a correspondente imagem ajustada. Como não há esse conjunto, as imagens em escala de cinza e as imagens em modelo de cores RGB são repassadas como entrada. Essa abordagem segue as recomendações sugeridas por [11].

Após várias iterações e gerações das redes, a acurácia é melhorada e as imagens resultantes proporcionam um padrão normalizado de cores para os corantes H&E. Nesse trabalho, o treinamento foi realizado a partir de 20 imagens selecionadas por um especialista com boa distribuição das cores que representam os corantes. Dessas imagens foram gerados 481 *patches* de tamanho 256 x 256 pixels para o treinamento e 592 *patches* de 256 x 256 pixels para os testes. Foi utilizada uma quantidade de imagens maior para a fase de testes porque essa etapa resulta nas imagens normalizadas, que foram utilizadas para comparação dos resultados dos métodos utilizados, assim gerando uma quantidade de dados maior para realizar essa comparação de forma mais relevante. Foram usadas 30 épocas para treinamento desse modelo.

**Classificação das Lesões com CNN:** Após essa etapa, as imagens histológicas foram classificadas com a AlexNet e ResNet para uma avaliação em relação ao processo de normalização assim como analisar o desempenho das abordagens sobre os diversos níveis de displasia. Com a base

de imagens normalizada, regiões foram recortadas em *patches* de tamanho 32 x 32 pixels resultando em 11.840 imagens, sem sobreposição dos *patches*. Esse recorte ocorreu devido a capacidade de processamento dos dados e necessidade de memória para processamento. Em seguida, um aumento de dados (do inglês, *data augmentation* - DA) foi empregado produzindo imagens com modificações em operações de rotação e espelhamento. As imagens foram geradas em diferentes direções, totalizaram-se 94.720 imagens divididas nas quatro classes de imagens histológicas. Nessa etapa, a divisão das imagens foi realizada da seguinte maneira: 20% foi utilizada na fase de teste. Dos 80% restantes, 20% foram selecionados para validação, e o restante foi utilizado para treinamento. Os modelos adotados foram treinados por 50 épocas.

A rede AlexNet foi proposta em 2012 por Alex Krizhevsky, Geoffrey Hinton e Ilya Sutskever. Esse modelo tem uma arquitetura que emprega cinco camadas convolucionais com função de ativação ReLU, camadas de *pooling* após a primeira, segunda e quinta camadas de convolução, e por fim duas camadas densas. Essa rede foi criada com o objetivo de vencer o desafio “*ImageNet Large Scale Visual Recognition Challenge*” e alcançou a acurácia em reconhecimento de imagens maior do que modelos anteriores [15], [16]. Na implementação da rede AlexNet foram feitas algumas adaptações na arquitetura com objetivo de melhorar o desempenho na classificação. Essas modificações foram a inserção de duas camadas de *dropout*, sendo uma no início e uma no final da arquitetura, e a utilização de regularizadores de *kernel* e *bias*. Essas adaptações são ofertadas pela biblioteca do Keras. A camada de *dropout* é responsável por desativar aleatoriamente alguns neurônios da camada anterior, e os regularizadores são capazes de aplicar penalidades nas funções de ativação do modelo, aumentando a otimização. O principal motivo de realizar essas alterações foi evitar o *overfitting* do modelo original, que ocorreu nos primeiros treinamentos da rede.

O segundo modelo empregado foi a ResNet (*Residual Network*) em que são empregadas as SC, que são capazes de pular uma ou mais camadas para transmitir o valor residual da entrada de uma camada para a saída de outra, reduzindo assim a degradação de informações ao longo da rede. Dessa maneira, são mantidas intactas as informações de entrada, produzindo um erro de treinamento menor do que os modelos que não utilizam blocos residuais [16]. As camadas da ResNet podem variar, existindo arquiteturas com 50, 101 ou 152 camadas. Neste trabalho foi utilizado o modelo ResNet-50, com 50 camadas, um dos modelos mais utilizados da arquitetura, capaz de produzir resultados relevantes ao contexto de aplicação.

**Classificação das Lesões com Características Aprendidas:** A partir do treinamento dos modelos CNNs, foi realizada a extração de características da última camada antes da classificação com o *softmax*. Essas informações foram extraídas do “*flattening*” e foram empregadas para construir os vetores de características a serem avaliados com outros algoritmos. Para os algoritmos de aprendizado de máquina

foram extraídas 1024 características por imagem e foram avaliadas com os algoritmos RF e SVM. Essa abordagem que faz a combinação das características aprendidas nas camadas convolucionais e os algoritmos de aprendizagem de máquina são denominadas modelos híbridos.

Os algoritmos RF e SVM foram utilizados a partir da biblioteca scikit-learn do Python. O algoritmo RF tem estruturas de dados baseadas em conjuntos árvores de decisão, chamadas florestas, e SVMs são algoritmos de aprendizado supervisionado que possuem o objetivo de fazer a separação dos dados de entrada utilizando o conceito estatístico da separabilidade linear e gradiente descendente [17]. Para a RF foi utilizada uma floresta com 20 árvores e os parâmetros padrões da biblioteca, como por exemplo, máximo de *features* igual a raiz quadrada do número de características, número mínimo de amostras em um nó folha igual a 1, e número mínimo de amostras necessárias para dividir um nó interno igual a 2. Para a técnica SVM, foi utilizado o *kernel* polinomial de grau 3 e os parâmetros padrões da biblioteca, como gama escalar, tolerância para critério de parada igual a  $1e-3$  e tamanho do cache para o processamento igual a 200 MB. Ambos os algoritmos foram treinados sob as características de treino dos modelos e avaliadas sob as características de teste.

### C. Métricas de Avaliação

As avaliações do método de normalização foram realizadas pelas métricas *Feature SIMilarity index* (FSIM), *Peak signal-to-noise ratio* (PSNR) e *Quaternion Structural SIMilarity index* (QSSIM), como utilizado por trabalhos da literatura. A medida FSIM foi utilizada para quantificar a similaridade entre os atributos das imagens originais e normalizadas. Essa métrica deve ser maximizada para indicação de relevantes resultados. A medida PSNR apresenta relação inversa à medida de erro quadrático médio que quantifica as diferenças de intensidade entre os pixels das imagens original e normalizada. Dessa forma, quanto maior o resultado da PSNR, maior é a qualidade da imagem. A medida QSSIM é uma medida de qualidade das degradações na luminância e crominância da imagem normalizada em relação à imagem original. Sua maximização indica melhores resultados da normalização pela pequena degradação da luminância como por borrachas e ruídos.

Para avaliar os resultados da classificação com a CNN, RF e SVM foram utilizadas as métricas dadas por sensibilidade, especificidade, precisão e acurácia, retiradas a partir da matriz de confusão. A matriz de confusão combina as classes reais e as classificações realizadas, como exemplificado na Figura 3.

## III. EXPERIMENTOS E RESULTADOS

### A. Avaliação da Normalização dos Corantes H&E

Nas Figuras 4(b), 4(d), 4(f) e 4(h) são mostrados os resultados da normalização com o método STST para as imagens histológicas. Os resultados mostram visualmente que o método é capaz de fazer uma distribuição mais uniforme dos corantes das imagens que estão mais degradadas em relação a distribuição dos corantes (ver Figuras 4(a), 4(b), 4(g) e 4 (h)). Nota-se visualmente que a intensidade de brilho foi melhorada

		Classificação	
		Sim	Não
Real	Sim	Verdadeiro Positivo (VP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (VN)

Figura 3: Exemplo de uma matriz de confusão simples. A partir dela são retiradas as métricas para a avaliação dos resultados.

para uma uniformização da distribuição dos corantes. Para comprovar isso, na Tabela I são apresentados os valores das métricas de melhoramento dos corantes. Nota-se que, nesses valores, todos os níveis de displasia tiveram um melhoramento semelhante às imagens de tecido saudável.

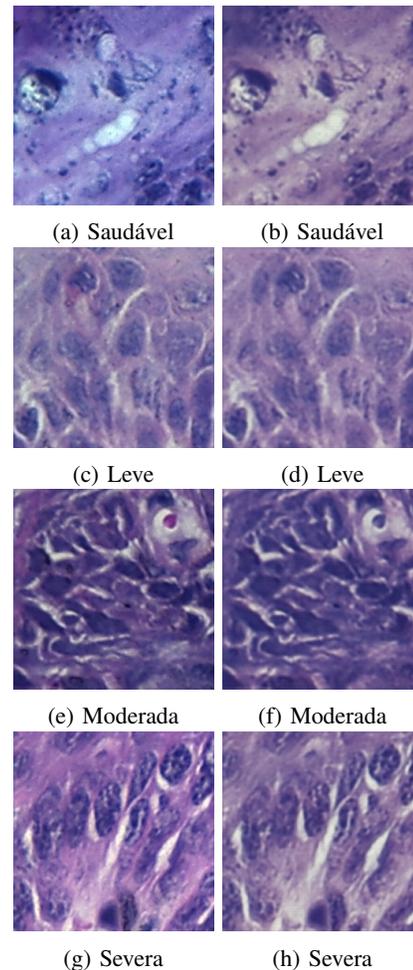


Figura 4: Exemplo de normalizações de cada classe usando o método STST.

### B. Avaliação da Classificação com CNN

Neste experimento foi considerada a avaliação da classificação dos níveis de displasia das imagens originais

Tabela I: Resultados das métricas de melhoramento de imagem.

Métrica	Classe	FSIM	QSSIM	PSNR
Média	Saudável	0,94	0,85	11,53
	Leve	0,90	0,83	14,14
	Moderado	0,90	0,84	14,46
	Severo	0,91	0,84	12,80
	Desvio Padrão	Saudável	0,08	0,08
	Leve	0,09	0,10	1,75
	Moderado	0,08	0,09	2,73
	Severo	0,08	0,08	1,90

e das imagens normalizadas. Até o momento da divulgação deste artigo não foi encontrado outro trabalho que utilizasse o mesmo banco de imagens para classificação da displasia, para que fosse possível realizar uma comparação dos resultados alcançados. Dessa maneira, para realizar essa investigação foram propostos quatro cenários para os modelos de classificação: i) treinamento com as imagens originais e teste com as imagens originais; ii) treinamento com as imagens originais e teste com as imagens normalizadas; iii) treinamento com as imagens normalizadas e teste com as imagens originais e iv) treinamento com as imagens normalizadas e teste com as imagens normalizadas. Os resultados da classificação para o cenário *i* é apresentado na Tabela II. Nessa investigação observou-se que a ResNet proporcionou os melhores resultados em relação a métrica de acurácia para grande parte dos níveis de displasia investigados. Esse comportamento também foi observado nas outras medidas investigadas nesse estudo. Apenas para o grupo de lesões moderadas a rede AlexNet foi superior nas métricas. No entanto, ressalta-se que na média geral entre os modelos a arquitetura ResNet tem desempenho acima (acurácia de 90,60%).

Os resultados do cenário *ii* são mostrados na Tabela III. Nessa avaliação observa-se que os valores das medidas foram inferiores aos resultados da investigação apenas com os dados originais (cenário *i*). Isso mostra que os modelos são sensíveis as características das cores das imagens, ou seja, durante a fase de treinamento os modelos aprenderam as informações de intensidades de brilho sem a normalização, ou seja, sem melhoramento das intensidades dos corantes H&E. Isso afetou a fase de teste do modelo com as imagens normalizadas. Para o cenário *iii*) os resultados são apresentados na Tabela IV. Essa investigação também mostra que a normalização provoca influência nos resultados das redes. As alterações nas características do contraste das imagens, fase treinamento, influenciaram o desempenho dos modelos.

O cenário *iv* é apresentado na Tabela V. Nesse experimento é possível observar que o modelo ResNet proporcionou os melhores resultados para as classes investigadas (acurácia média de 97,01), mostrando que o uso da normalização foi relevante para o desempenho das redes em relação ao modelo sem nenhum tratamento (cenário *i*) tendo as imagens normalizadas avaliadas nas fases de treinamento e teste. No entanto, para a AlexNet esse comportamento foi diferente, ou seja, sem a normalização o modelo fornece resultados mais relevantes numa classificação, mostrando que o método de normalização pode influenciar de forma negativa o modelo.

Tabela II: Resultados (%) da classificação realizada com as imagens originais, sem investigação da normalização.

Modelo	Métrica	Saudável	Leve	Moderada	Severa	Média
AlexNet	Sensib.	91,29	78,12	80,08	68,97	79,60
	Especif.	90,82	91,19	91,95	98,88	93,20
	Precisão	76,86	74,21	77,28	95,35	80,90
	Acurácia	90,94	87,99	88,93	91,40	89,80
	ResNet	Sensib.	65,65	82,78	79,17	97,63
	Especif.	99,65	93,20	91,64	90,54	93,80
	Precisão	98,45	79,80	76,40	77,47	83,00
	Acurácia	91,14	90,65	88,46	92,31	90,60

Tabela III: Resultados (%) da classificação com as imagens normalizadas empregadas apenas na fase de teste.

Modelo	Métrica	Saudável	Leve	Moderado	Severo	Média
AlexNet	Sensib.	7,15	61,28	29,78	39,09	34,33
	Especif.	93,32	58,86	80,39	79,79	78,09
	Precisão	26,34	32,58	34,17	39,19	33,07
	Acurácia	71,75	59,45	67,50	69,62	67,08
	ResNet	Sensib.	0,11	81,90	27,65	9,48
	Especif.	99,77	33,70	73,45	99,35	76,57
	Precisão	13,51	28,61	26,25	82,99	37,84
	Acurácia	74,82	45,51	61,79	76,89	64,75

Tabela IV: Classificação (%) com as imagens normalizadas empregadas na fase de treinamento e as imagens originais avaliadas na fase de teste.

Modelo	Métrica	Saudável	Leve	Moderado	Severo	Média
AlexNet	Sensib.	1,03	61,06	15,27	30,50	26,97
	Especif.	92,56	31,66	97,34	81,18	75,69
	Precisão	4,43	22,48	66,28	35,07	32,06
	Acurácia	69,65	38,86	76,44	68,51	63,37
	ResNet	Sensib.	0,00	11,10	0,46	98,71
	Especif.	99,40	92,64	99,18	12,17	75,85
	Precisão	0,00	32,87	15,94	27,25	19,01
	Acurácia	74,51	72,67	74,03	33,80	63,75

Tabela V: Resultados (%) obtidos com as imagens normalizadas para as fases de treinamento e teste.

Modelo	Métrica	Saudável	Leve	Moderado	Severo	Média
AlexNet	Sensib.	93,27	76,90	31,59	35,54	59,33
	Especif.	64,73	83,93	97,74	99,17	86,39
	Precisão	46,90	60,83	82,69	93,45	70,97
	Acurácia	71,88	82,21	80,89	83,27	79,56
	ResNet	Sensib.	96,61	93,00	90,74	95,80
	Especif.	98,32	98,74	97,57	97,39	98,01
	Precisão	95,06	96,00	92,73	92,44	94,06
	Acurácia	97,89	97,33	95,83	96,99	97,01

### C. Avaliação da Classificação com Características Aprendidas

Os resultados das classificações híbridas com os algoritmos RF e SVM com as características das últimas camadas da CNNs são apresentadas nas Tabelas VI e VII. Nesses experimentos foram considerados apenas os cenários *i* e *iv* devido aos resultados mais relevantes observados nas tabelas. Os dados das imagens sem normalização mostram que essa abordagem híbrida foi promissora para os modelos das características aprendidas pelas CNNs, pois a AlexNet elevou sua média de acurácia de 89,80% para 92,73% e 93,913% para os algoritmos RF e SVM, respectivamente. Com o modelo ResNet, o valor de acurácia de 90,60% com o modelo apenas CNN foi elevado para 97,34%, com o RF, e 97,05%, com o SVM.

Com as imagens normalizadas, os valores médios da

Tabela VI: Valores das métricas para classificação dos modelos híbridos utilizando as imagens originais sem aplicação da normalização.

Modelo	Métrica	Saudável	Leve	Moderado	Severo	Média		
RF	AlexNet	Sensib.	92,62	84,18	80,81	84,31	85,48	
		Especif.	97,02	91,36	95,42	96,85	95,16	
		Precisão	91,22	75,97	85,79	89,91	85,72	
		Acurácia	95,92	89,60	91,70	93,71	92,73	
	ResNet	Sensib.	97,53	93,90	91,32	96,03	94,70	
		Especif.	99,13	97,01	97,95	98,82	98,23	
		Precisão	97,41	91,06	93,84	96,46	94,69	
		Acurácia	98,73	96,25	96,26	98,13	97,34	
	SVM	AlexNet	Sensib.	92,96	91,51	80,81	86,19	87,87
			Especif.	98,08	89,89	97,63	98,20	95,95
			Precisão	94,19	74,60	92,11	94,10	88,75
			Acurácia	96,80	90,29	93,35	95,20	93,91
ResNet		Sensib.	98,08	93,36	90,28	94,70	94,11	
		Especif.	98,39	96,76	97,94	99,04	98,03	
		Precisão	95,33	90,35	93,74	97,04	94,11	
		Acurácia	98,32	95,93	95,99	97,95	97,05	

Tabela VII: Resultados da classificação dos modelos híbridos utilizando as imagens normalizadas.

Modelo	Métrica	Saudável	Leve	Moderado	Severo	Média		
RF	AlexNet	Sensib.	76,77	70,11	61,95	71,74	70,14	
		Especif.	89,93	90,45	87,85	91,90	90,03	
		Precisão	71,80	70,43	63,53	74,69	70,11	
		Acurácia	86,63	85,47	81,25	86,86	85,05	
	ResNet	Sensib.	96,69	94,23	91,36	94,19	94,12	
		Especif.	98,87	98,36	97,00	97,89	98,03	
		Precisão	96,63	94,92	91,24	93,70	94,12	
		Acurácia	98,33	97,35	95,57	96,96	97,05	
	SVM	AlexNet	Sensib.	77,38	75,03	74,38	75,44	75,56
			Especif.	93,44	92,44	87,29	94,20	91,84
			Precisão	79,77	76,31	66,66	81,26	76,00
			Acurácia	89,42	88,18	84,00	89,51	87,78
ResNet		Sensib.	93,30	89,53	92,70	88,91	91,11	
		Especif.	98,80	98,43	92,82	98,09	97,03	
		Precisão	96,30	94,86	81,52	93,93	91,65	
		Acurácia	97,42	96,25	92,79	95,79	95,56	

acurácia foram de 79,56% e 97,01% para os modelos AlexNet e ResNet, respectivamente. No modelo híbrido, a acurácia foi para 85,05% (AlexNet) e 97,05% (ResNet) com o algoritmo RF. Com o algoritmo SVM, esses valores foram modificados passando para 87,78% (AlexNet) e 95,56% (ResNet). Nota-se que os algoritmos RF e SVM permitiram melhorar os valores da acurácia no modelo híbrido com as características aprendidas para a arquitetura AlexNet. No entanto, esses valores foram inferiores aos obtidos apenas com a arquitetura ResNet.

#### IV. CONCLUSÃO

Neste trabalho, foi realizado um estudo de um modelo para classificação de lesões de displasia em um sistema CAD que possibilite auxiliar os profissionais especialistas em histologia a diagnosticar com maior exatidão a lesão pré-cancerígena chamada displasia. Além disso, a displasia é um tipo de lesão de difícil detecção, e por isso se faz importante o desenvolvimento de novas abordagens que possam melhorar a detecção desta doença.

Desse modo, o modelo desenvolvido neste trabalho obteve resultados satisfatórios, pois a normalização foi realizada de forma a manter as características das imagens, e por consequência teve um impacto positivo na classificação para grupos de lesões em alguns dos cenários investigados. Os classificadores RF e SVM obtiveram resultados relevantes para as características obtidas com a CNN. Um valor de acurácia superior à 95% para o caso do modelo da arquitetura

ResNet50. O estudo mostrou também a importância dos modelos híbridos que combinam as arquiteturas CNN para extração de características e algoritmos de aprendizado de máquina para a classificação. Os resultados foram melhores do que o modelo que usa apenas a camada *softmax* para classificação. Em trabalhos futuros serão investigados outros modelos de arquiteturas CNNs para avaliar o comportamento nas fases dos cenários propostos.

#### AGRADECIMENTOS

Esse estudo recebeu suporte financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (Projetos #304848/2018-2, #430965/2018-4 e #313365/2018-0) e da Fundação de Amparo a Pesquisa do Estado de Minas Gerais - FAPEMIG (#APQ-00578-18 e # APQ-01129-21).

#### REFERÊNCIAS

- [1] “Instituto nacional de câncer,” INCA, 2019. [Online]. Available: <https://www.inca.gov.br>
- [2] J. A. A. Jothi and V. M. A. Rajam, “A survey on automated cancer diagnosis from histopathology images,” *Artificial Intelligence Review*, vol. 48, no. 1, pp. 31–81, 2017.
- [3] T. Fonseca-Silva, M. G. Diniz, S. F. Sousa, R. S. Gomez, and C. C. Gomes, “Association between histopathological features of dysplasia in oral leukoplakia and loss of heterozygosity,” *Histopathology*, vol. 68, no. 3, pp. 456–460, 2016.
- [4] A. Ham and D. H. Cormack, “Histologia,” in *Histologia*, 1983, pp. 906–906.
- [5] Anonymous, “Métodos computacionais para análise e classificação de displasias em imagens da cavidade bucal,” Master’s thesis, Universidade Federal de Uberlândia - Faculdade de Computação, 2019.
- [6] Roberto, G. F. Lumini, A. Neves, L. A. do Nascimento, and M. Zanchetta, “Fractal neural network: A new ensemble of fractal geometry and convolutional neural networks for the classification of histology images,” *Expert Systems with Applications*, vol. 166, p. 114103, 2021.
- [7] W. Li, J. Li, K. V. Sarma, K. C. Ho, S. Shen, B. S. Knudsen, A. Gertych, and C. W. Arnold, “Path r-cnn for prostate cancer diagnosis and gleason grading of histological images,” *IEEE transactions on medical imaging*, vol. 38, no. 4, pp. 945–954, 2018.
- [8] R. C. Gonzalez and R. Woods, “Digital image processing,” 2018.
- [9] M. Salvi, U. R. Acharya, F. Molinari, and K. M. Meiburger, “The impact of pre-and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis,” *Computers in Biology and Medicine*, p. 104129, 2020.
- [10] C.-M. Chen, Y.-S. Huang, P.-W. Fang, C.-W. Liang, and R.-F. Chang, “A computer-aided diagnosis system for differentiation and delineation of malignant regions on whole-slide prostate histopathology image using spatial statistics and multidimensional densenet,” *Medical physics*, vol. 47, no. 3, pp. 1021–1033, 2020.
- [11] P. Salehi and A. Chalechale, “Pix2pix-based stain-to-stain translation: A solution for robust stain normalization in histopathology images analysis,” *2020 International Conference on Machine Vision and Image Processing (MVIP)*, 2020.
- [12] T. A. A. Tosta, P. R. de Faria, L. A. Neves, and M. Z. do Nascimento, “Computational normalization of h&e-stained histological images: Progress, challenges and future potential,” *Artificial intelligence in medicine*, vol. 95, pp. 118–132, 2019.
- [13] A. BenTaieb and G. Hamarneh, “Adversarial stain transfer for histopathology image analysis,” *IEEE transactions on medical imaging*, vol. 37, no. 3, pp. 792–802, 2017.
- [14] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” *arXiv preprint arXiv:1411.1784*, 2014.
- [15] C. Y. S. W. P. S. M. S. N. B. C. V. E. A. A. S. A. Md Zahangir Alom, Tarek M. Taha and V. K. Asari, “The history began from alexnet: A comprehensive survey on deep learning approaches,” *arXiv.org*, 2018.
- [16] Q. Z. D. A. M. J. D. Hang Yu, Laurence T. Yang, “Convolutional neural networks for medical image analysis: State-of-the-art, comparisons, improvement and perspectives,” *Elsevier*, 2021.
- [17] V. Vapnik, “An overview of statistical learning theory,” *Neural Networks, IEEE Transactions on*, vol. 10, no. 5, pp. 988–999, 1999.