

Estudo Comparativo entre MobileNet V3 Large e SqueezeNet para Classificação de Células NIH-3T3 em Esferóides

Danilo França
Centro de Engenharia, Modelagem e
Ciências Sociais Aplicadas
Universidade Federal do ABC
São Bernardo do Campo, Brazil
ORCID: 0000-0002-0234-5040

Anderson G. Santiago
Centro de Engenharia, Modelagem e
Ciências Sociais Aplicadas
Universidade Federal do ABC
São Bernardo do Campo, Brazil
ORCID: 0000-0001-7581-1148

Camila C. Santos
Centro de Engenharia, Modelagem e
Ciências Sociais Aplicadas
Universidade Federal do ABC
São Bernardo do Campo, Brazil
ORCID: 0000-0003-0954-3338

Juliana K. M. B. Daguano
Laboratório de Biotecnologia
Centro de Tecnologia da Informação
Renato Archer
Campinas, Brazil
ORCID: 0000-0001-6098-6826

Andrea C.D. Rodas
Centro de Engenharia, Modelagem e
Ciências Sociais Aplicadas
Universidade Federal do ABC
São Bernardo do Campo, Brazil
ORCID: 0000-0001-9920-6882

Resumo — *As redes neurais convolucionais (CNNs) foram desenvolvidas para reconhecimento de padrões em imagens. Seu uso inicial foi para imagens médicas, onde é possível encontrar uma vasta literatura. Com o desenvolvimento de tecnologias mais avançadas para estudo de processos biológicos, as CNNs têm chamado a atenção para a avaliação de imagens biológicas. O objetivo deste trabalho é a exploração de CNNs que consigam identificar estruturas com poucas diferenças de imagens de esferóides sem coloração de diferenciação, que são métodos destrutivos de análise. Esse trabalho implementa duas diferentes arquiteturas de redes neurais convolucionais, MobileNet V3 Large e SqueezeNet, em um estudo comparativo da classificação de esferóides celulares de fibroblastos NIH-3T3 utilizando métricas tais como gráficos de acurácia de treinamento e teste, função perda e escore F1 para determinar qual das arquiteturas de CNN apresenta o melhor desempenho. A implementação foi feita através da variação de alguns hiperparâmetros das arquiteturas de CNN, como número de épocas, taxa de aprendizagem, ângulo de rotação da imagem e batch size, identificando a SqueezeNet como a arquitetura com melhor desempenho para análise dessas imagens. Os hiperparâmetros da implementação que resultou no melhor desempenho são: número de épocas 70, learning rate de 1.0^{-4} , rotação de 20° e batch size de 10.*

Palavras-chave — *Esferóides celulares, Fibroblastos NIH-3T3, Arquiteturas de CNNs, Redes Neurais.*

I. INTRODUÇÃO

Técnicas de cultura celular têm sido empregadas desde a década de 1940 e são utilizadas desde então para examinar a biologia das células e mecanismos moleculares. Apesar de modelos bidimensionais de culturas de células em monocamadas aderentes serem os mais utilizados, não há nesses modelos uma representação fiel do ambiente natural em que as células se encontram, impedindo interações célula-célula e célula-matriz extracelular. A partir dessa necessidade de criação de uma representação mais precisa de

organização celular em tecidos, desenvolveu-se o método de produção de esferóides celulares, agregados de células auto-organizáveis inseridas em ambientes que impedem a fixação em superfícies planas [1]. Esse modelo *in vitro* de células tridimensionais possibilita essas interações intercelulares, além de serem de fácil manuseio e mais apropriados para obtenção de imagens de alta qualidade [2].

Há uma série de inferências possíveis a partir da observação de uma imagem biológica — mais especificamente uma imagem de um grupo de células — tais como compreender o comportamento celular, atentar-se a padrões, classificar as células, modelar a imagem e fazer previsões, dentre outros.

Nos últimos anos, sistemas de reconhecimento de imagens baseados em redes neurais convolucionais (“Convolutional Neural Networks” - CNN), apresentaram uma melhora expressiva, de forma que a capacidade de análise de dados por uma CNN excedeu há muito a capacidade humana em relação ao tempo de análise de imagens, encontrando aplicações de destaque na análise de imagens biológicas [3]. Ao combinar essas imagens com as técnicas de CNNs, com atenção à escolha da arquitetura de CNN mais adequada para a imagem em questão, o resultado é uma otimização de todo esse processo, uma vez que há não só um aumento na acurácia de identificação e classificação, como uma redução do tempo que leva esse procedimento.

Em um trabalho anterior realizado por Santiago e colaboradores [4], imagens de esferóides foram utilizadas para treinamento das CNNs AlexNet, GoogLeNet, ResNet 18 e VGG 16 BN. Entre as CNNs estudadas, a AlexNet foi a que apresentou resultados satisfatórios no reconhecimento e classificação das estruturas presentes nas imagens de esferóides, porém o modelo de AlexNet possui uma quantidade de parâmetros na ordem de 57×10^6 a serem calculados, tornando-se computacionalmente custoso quando aplicada em um conjunto com uma quantidade expressiva de imagens. Assim, é desejável buscar arquiteturas de CNN que possuam menor complexidade que a AlexNet. Desta forma, o artigo apresenta uma comparação entre duas

diferentes arquiteturas de Redes Neurais Convolucionais (“Convolutional Neural Networks” - CNN) para classificação de estruturas em esferóides celulares de fibroblastos NIH-3T3. As arquiteturas consideradas são MobileNet V3 Large e SqueezeNet e seus desempenhos são comparados utilizando métricas tais como os gráficos de acurácia de treinamento e teste, função perda e do escore F1. Assim, espera-se que, quando aplicadas ao conjunto de imagens completo, apresentem maior velocidade tanto em relação ao treinamento, teste e validação do modelo.

II. METODOLOGIA

A. Implementação da CNN

A CNN responsável por classificar as estruturas foi implementada em Python 3.7 distribuição Anaconda com a Interface de Programação de Aplicações (“Application Programming Interface” - API) PyTorch. A especificação do hardware é um Core i7 com 16Gb e uma Unidade de Processamento Gráfico (“Graphic Processing Unit” – GPU) Tesla K80 Nvidia, com 24G de memória RAM.

Foram consideradas duas arquiteturas de CNN: SqueezeNet e MobileV3 Large, considerando quatro classes distintas (background, células vivas dentro do esferoide, células vivas fora do esferoide e células mortas fora do esferoide). Para cada arquitetura, foram considerados diferentes conjuntos de hiperparâmetros e para o processo de treinamento, utilizou-se o otimizador Adam [5], além da função custo de Entropia Cruzada (“Cross Entropy” - CE), definida pela Equação 1 na qual y_i expressa o valor previsto para uma amostra.

$$CE = \sum_{i=0}^{N_{class}} y_i \log(y_i - 1) + (y_i - 1) \log(y_i) \quad (1)$$

Para cada arquitetura, a camada de classificação sofreu alterações na camada de classificação, como mostra a Tabela I abaixo.

TABELA I. CAMADA DE CLASSIFICAÇÃO DA ARQUITETURA DAS CNNs ESCOLHIDAS

Arquitetura de CNN	Camada de Classificação
MobileNet V3 Large	Linear Hardswish Dropout(p = 0.2) Linear(in_features=1280, out_features=4)
SqueezeNet	Dropout (p = 0.5) Conv2d Rectified Linear Unit (ReLU) AdaptiveAvgPool2d

Os valores da Tabela II expressam os hiperparâmetros utilizados na fase de treinamento dos modelos de CNNs MobileNet V3 Large e SqueezeNet, que sofreram alterações e posterior análise de desempenho através da interpretação das curvas e matriz de confusão.

TABELA II. F1 SCORE AVALIADO PARA CADA CLASSE E CADA ARQUITETURA DE CNN

	Hiperparâmetros			
	Número de épocas	Learning rate	Ângulo de rotação	Batch size
Original	60	1.0 ⁻⁴	20	10
Faixa de Variação	55 ~ 70	1.0 ⁻³ ~ 1.0 ⁻⁶	10 ~ 25	5 ~ 15

B. Avaliação da CNN

Uma vez que o conjunto de dados de teste é composto por um número desequilibrado de classes, isto é, algumas classes apresentaram maior quantidade de amostras que outras, foi escolhido o escore F1 como métrica de desempenho, definido pela Equação 2. O escore F1 representa a média harmônica entre a precisão (Eq. 3), que exprime a habilidade do modelo de classificar uma amostra como positiva corretamente classificada e a revocação (Eq. 4), que exprime a habilidade do modelo de detectar amostras positivas.

Assim, as arquiteturas de CNNs e conjuntos de hiperparâmetros cujo resultado da pontuação F1 for mais próximo de 1, será considerada a arquitetura mais adequada para a classificação.

$$F1 = \frac{Precisão \times Revocação}{Precisão + Revocação} \quad (2)$$

$$Precision = TP / (TP + FP) \quad (3)$$

$$Recall = TP / (TP + FN) \quad (4)$$

em que TP é verdadeiro positivo (“True Positive”), FP é falso positivo (“False Positive”) e FN é falso negativo (“False Negative”).

III. RESULTADOS

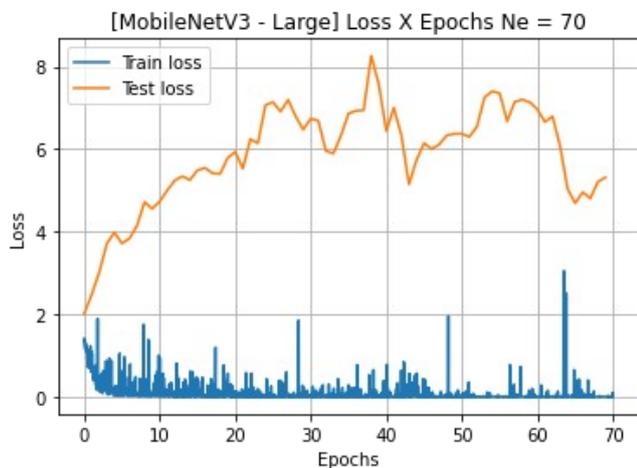
A Tabela III apresenta os valores de escore F1 para ambas as arquiteturas de CNN consideradas, exibindo o escore para cada classe, bem como o valor médio entre eles.

TABELA III. VALORES ORIGINAIS E VARIADOS DOS HIPERPARÂMETROS DAS ARQUITETURAS DE CNN

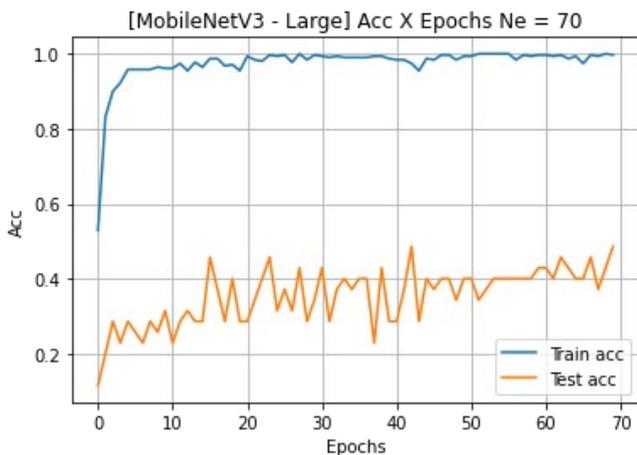
Arquitetura de CNN	Classes				
	0	1	2	3	Valor médio
MobileNet V3 Large	0.95	0.67	0.44	0.54	0.65
SqueezeNet	0.98	0.50	0.50	0.93	0.73

Os gráficos da Fig. 1 referem-se ao conjunto ótimo de hiperparâmetros de acordo com as métricas adotadas para a arquitetura MobileNet V3, apresentando a função custo, acurácia de treino e teste e a matriz de confusão da arquitetura. Esse resultado diz respeito ao seguinte conjunto de hiperparâmetros: Ne = 70, learning rate de 1.0⁻⁴, rotação máxima de 20° e batch size de 15. O tempo de treinamento e teste para a arquitetura foi de 4,74 (min).

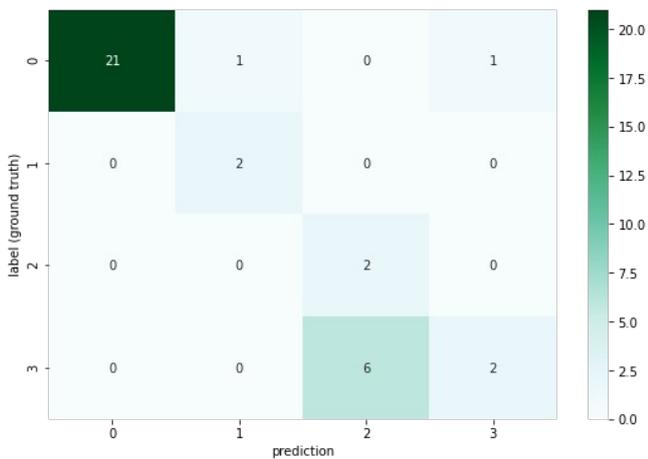
Os gráficos da Fig. 2 referem-se ao conjunto ótimo de hiperparâmetros de acordo com as métricas adotadas para a arquitetura SqueezeNet, apresentando a função custo, acurácia de treino e teste e a matriz de confusão da arquitetura. Esse resultado diz respeito ao seguinte conjunto de hiperparâmetros: $N_e = 60$, *learning rate* de 1.0^{-4} , rotação com variação máxima de 20° e *batch size* de 15. O tempo de treinamento e teste para a arquitetura foi de 5,92 (min).



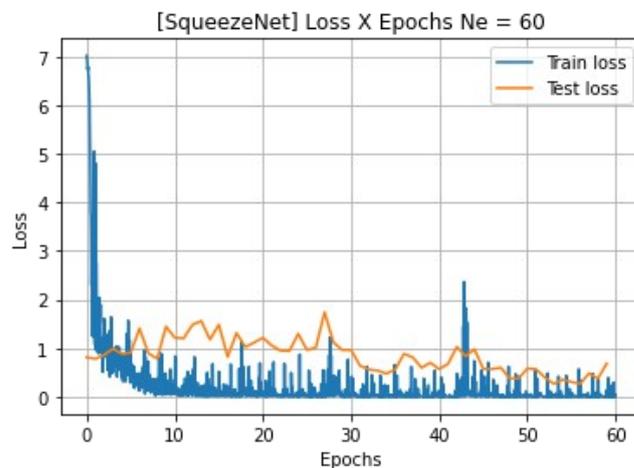
(a)



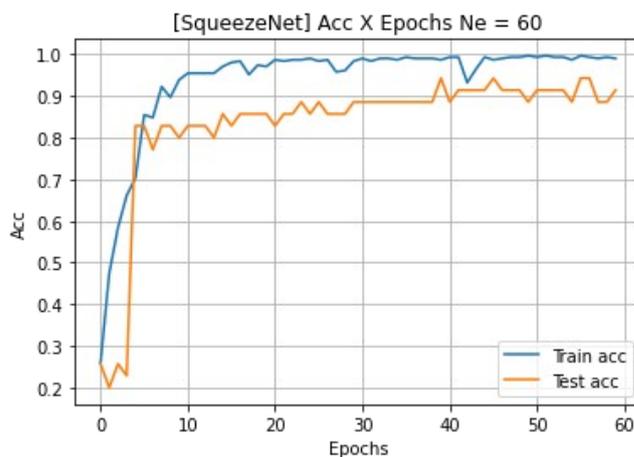
(b)



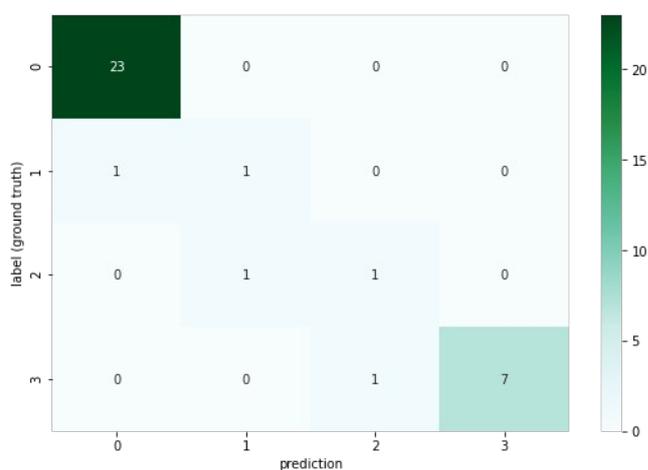
(c)



(a)



(b)



(c)

Fig. 1. Resultados referentes à CNN MobileNet V3 Large (a) Curvas da função custo dos dados de treinamento e de teste (b) Curvas de acurácia dos dados de treinamento e de teste (c) Matriz de confusão para o conjunto de dados de teste: 0 - Background; 1 - Células mortas; 2 - Células vivas fora do esferoide; 3 - Células vivas dentro do esferoide.

Fig. 2: Resultados referentes à CNN SqueezeNet (a) Curvas da função custo dos dados de treinamento e de teste (b) Curvas de acurácia dos dados de treinamento e de teste (c) Matriz de confusão para o conjunto de dados de teste: 0 - Background; 1 - Células mortas; 2 - Células vivas fora do esferoide; 3 - Células vivas dentro do esferoide.

IV. DISCUSSÃO

Através da análise dos resultados apresentados nas Figs. 1 e 2, é possível realizar algumas inferências a respeito do desempenho de cada uma dessas arquiteturas de CNNs na classificação das imagens.

Em uma observação às Figs. 1(a) e 2(a), nota-se que a MobileNet V3 apresenta valores de custo elevados, enquanto a SqueezeNet apresenta valores de custo próximos a zero em praticamente todas as épocas, o que significa que apesar de apresentar um bom desempenho na fase de treinamento, a MobileNet V3 não teve êxito em generalizar essa performance para o conjunto de imagens de teste.

Ao considerar as curvas de acurácias mostradas nas Figs. 1(b) e 2(b), relativas ao conjunto de dados de teste, é evidente que a SqueezeNet teve uma performance superior à MobileNet V3: enquanto a MobileNet V3 apresentou valores de acurácia em torno de 0,4, a SqueezeNet apresentou valores em torno de 0,9, bem próximos à acurácia de treino.

Quando se trata da análise das matrizes de confusão, sendo sua análise quantitativa realizada via escore F1, a SqueezeNet apresentou um valor médio de 0,73, demonstrando ser mais adequada que a MobileNet V3, com um escore F1 médio de 0,65 para classificação das imagens dos fibroblastos NIH-3T3.

V. CONCLUSÃO

O trabalho aborda duas diferentes arquiteturas de CNNs para classificação de imagens de fibroblastos NIH-3T3 e revela a eficácia e importância do uso de métodos não-destrutivos para classificação dessas imagens biológicas. Frente às métricas de classificação estabelecidas, é evidente que a SqueezeNet demonstrou melhor desempenho, sendo considerada mais adequada que a MobileNet V3 Large nesse tipo de classificação. Os

valores das métricas calculadas ratificam, portanto, o uso de CNNs como uma possibilidade de otimizar e automatizar processos de classificação de imagens biológicas, que durante muito tempo foram realizados de forma manual, sem ameaçar a qualidade de classificação.

AGRADECIMENTOS

Os autores agradecem à CAPES – código 001.

REFERÊNCIAS

- [1] K Białkowska, P Komorowski, M. Bryszewska e K. Miłowska. (2020) Spheroids as a Type of Three-Dimensional Cell Cultures-Examples of Methods of Preparation and the Most Important Application. *Int J Mol Sci.* <https://doi.org/10.3390/ijms21176225>.
- [2] Fischer S C (2019) An Introduction to Image-Based Systems Biology of Multicellular Spheroids for Experimentalists and Theoreticians In: Husi H, editor. *Computational Biology* [Internet]. Brisbane (AU): Codon Publications. cap. 1. <https://doi.org/10.15586/computationalbiology.2019.ch1>
- [3] Kusumoto, D. e Yuasa, S., "The application of convolutional neural networks to stem cell biology," *Inflammation and Regeneration*, vol. 39, n.14. <https://doi.org/10.1186/s41232-019-0103-3>
- [4] A G Santiago, C C Santos, M M G Macedo, J K M B Daguano e A C D Rodas (2021) "Non-destructive fibroblast NIH-3T3 spheroid classification using convolutional neural network", *Proc. SPIE 11600, Medical Imaging 2021: Biomedical Applications in Molecular, Structural, and Functional Imaging*, 11600N. <https://doi.org/10.1117/12.2580916>
- [5] Kingma D, P.; Ba, J. (2014) Adam: A method for stochastic optimization 3 International Conference for Learning Representations, San Diego, USA, 2015, pp 1–15