**Aprendizagem Profunda e IA Verde: Caminhos para um Futuro mais Sustentável**

**Vívian R. G. Ferraro1, Daniel da S. Costa2, Pedro Nuno de S. Moura1**

1Escola de Informática Aplicada – Universidade Federal do Estado do Rio de Janeiro (UNIRIO)  
22.290-255 – Rio de Janeiro – RJ – Brasil

2Programa de Pós-Graduação em Informática (PPGI) – Universidade Federal do Estado do Rio de Janeiro (UNIRIO)  
22.290-255 – Rio de Janeiro – RJ – Brasil

{viferraro,daniel.scosta}@edu.unirio.br, pedro.moura@uniriotec.br

**Abstract.** The objective of this article is to evaluate and compare deep learning models in terms of efficiency, energy expenditure and computational cost. To this end, a literature survey was carried out to identify and validate the main metrics used to evaluate these aspects of Deep Learning models. Additionally, experiments were conducted with prominent models in this field to evaluate their performance based on these metrics. The research aims to contribute to the understanding and optimization of the use of Deep Learning models in terms of efficiency and sustainability.

**Resumo.** O objetivo deste artigo é avaliar e comparar modelos de aprendizagem profunda em termos de eficiência, gasto energético e custo computacional. Para isso, foi realizado um levantamento da literatura para identificar e validar as principais métricas utilizadas para avaliar esses aspectos dos modelos de Aprendizagem Profunda. Além disso, foram conduzidos experimentos com modelos proeminentes nesse campo para avaliar seu desempenho com base nessas métricas. A pesquisa visa contribuir para a compreensão e otimização do uso de modelos de Aprendizagem Profunda em termos de eficiência e sustentabilidade.

# 1. Introdução

A Inteligência Artificial tem se tornado uma ferramenta cada vez mais importante em diversos setores da sociedade, incluindo saúde, educação, transporte e segurança. No entanto, o desenvolvimento e uso de modelos de Inteligência Artificial, especialmente os baseados em Aprendizagem Profunda, podem ter um impacto significativo no meio ambiente.

Modelos de Aprendizagem Profunda são capazes de realizar tarefas complexas, mas requerem um grande consumo de recursos computacionais. O treinamento desses modelos pode consumir grandes quantidades de energia. Além disso, o descarte de equipamentos que foram utilizados para treinar ou executar modelos de Aprendizagem Profunda também pode gerar um impacto ambiental.

O impacto ambiental do desenvolvimento e uso de modelos de Aprendizagem Profunda é uma questão importante que precisa ser considerada. A pesquisa nessa área está em uma fase de bastante efervescência, e alguns trabalhos recentes apontam para a necessidade de adotar medidas para minimizar esse impacto.

Para melhorar o desempenho dos modelos de aprendizagem profunda, a quantização surge como uma técnica promissora para otimizar o uso desses recursos. A quantização é uma técnica que visa reduzir o consumo de recursos computacionais e energéticos em modelos de Aprendizagem Profunda. Ao diminuir a precisão dos parâmetros do modelo, como os pesos das conexões neurais, é possível acelerar o processo de inferência e diminuir o tamanho do modelo, tornando-o mais eficiente e sustentável.

Nesse contexto, o objetivo deste artigo é avaliar e comparar modelos de aprendizagem profunda em termos de eficiência, gasto energético e custo computacional. Para isso, foi realizado um levantamento da literatura para identificar as principais métricas utilizadas para avaliar esses aspectos dos modelos de Aprendizagem Profunda. Também foram realizados experimentos com modelos relevantes nessa área para medir seu desempenho segundo essas métricas.

O artigo está organizado da seguinte forma: a seção 2 apresenta alguns trabalhos relacionados que abordam diferentes aspectos da eficiência, do consumo de energia e do custo computacional de modelos de Aprendizagem Profunda. A seção 3 descreve a metodologia proposta para avaliar e comparar modelos de Aprendizagem Profunda. A seção 4 apresenta os resultados obtidos nos experimentos realizados. A seção 5 discute as implicações e limitações dos resultados. A seção 6 conclui o artigo e sugere direções para trabalhos futuros.

# 2. Trabalhos Relacionados

Alguns dos trabalhos publicados nos últimos anos abordam diferentes aspectos da eficiência, do consumo de energia e do custo computacional de modelos de Aprendizagem Profunda. Os autores utilizaram uma variedade de métricas para avaliar a eficiência de modelos de Aprendizagem Profunda.

(i) Anthony et al. (2020) propuseram o Carbontracker, uma ferramenta que monitora e prevê o consumo de energia e as emissões de carbono do treinamento de modelos de aprendizagem profunda. A ferramenta visa conscientizar os pesquisadores sobre o impacto ambiental do seu trabalho e incentivá-los a adotar práticas mais sustentáveis. O artigo também avalia a ferramenta em diferentes arquiteturas e conjuntos de dados para segmentação de imagens médicas e fornece recomendações para diminuir as emissões de carbono.

(ii) O artigo de Douwes et al. (2021) aborda o problema do consumo de energia dos modelos geradores de áudio baseados em aprendizagem profunda. Os autores propõem uma nova medida baseada na otimização de Pareto, que considera tanto a qualidade do som sintetizado quanto o consumo de energia do modelo. Eles aplicam essa medida em vários modelos existentes, como SampleRNN, SING, WaveGAN, GANSynth e FloWaveNet, e mostram que ela pode alterar significativamente os resultados. Eles também realizam experimentos com um modelo recente chamado WaveFlow, e medem o consumo de energia real de treinamento e inferência para cinco configurações alternativas. Eles usam um espaço multiobjetivo para representar as soluções ótimas de Pareto e analisam o compromisso entre qualidade e eficiência energética.

(iii) Desislavov et al. (2023) analisaram o consumo de energia em modelos de visão computacional e processamento de linguagem natural, com um foco especial na aprendizagem profunda e em seus custos de inferência. A motivação por trás deste estudo é entender se o progresso exponencial em alguns paradigmas de IA, como o aprendizado profundo, resulta em um aumento exponencial no consumo de energia. Esta é uma questão crucial para avaliar o impacto ambiental e social da IA, bem como os limites do seu desempenho futuro. Utilizando uma metodologia que envolve a coleta de dados sobre o desempenho, os parâmetros, as operações de ponto flutuante (FLOPs) e a data de lançamento de vários modelos de visão computacional e processamento de linguagem natural, os autores estimam o consumo de energia por inferência usando dados de hardware de GPUs da Nvidia. Os resultados mostram que, embora o consumo de energia dos modelos de ponta ainda cresça exponencialmente para um aumento sustentado no desempenho, o ritmo é mais lento do que o esperado, graças às melhorias algorítmicas e à especialização do hardware. No entanto, os autores alertam para o efeito multiplicativo do aumento da penetração da IA, que pode elevar o consumo de energia global.

(iv) Lenherr et al. (2021) propuseram novas métricas para medir a eficiência e a sustentabilidade da aprendizagem profunda (DL) em plataformas de borda, que equilibram a precisão, a complexidade e o consumo de energia. As métricas são baseadas no gradiente de reconhecimento, no número de classes e na energia por inferência ou treinamento. O artigo também compara a eficiência do DL centralizado e federado, e analisa o ciclo de vida do DL, incluindo o número de vezes que os modelos são usados. Além disso, o artigo compara a eficiência da inteligência artificial e natural, e fornece insights sobre a escalabilidade do DL.

(v) Schwartz et al. (2020) introduziram o conceito de IA verde, uma abordagem de pesquisa em inteligência artificial que valoriza a eficiência computacional e promove a redução do uso de recursos. Em contraste, a IA Vermelha busca melhorar a acurácia usando grandes quantidades de poder computacional, sem levar em conta os custos ambientais, econômicos e sociais. Uma proposta dos autores é usar o número total de operações de ponto flutuante como uma medida de eficiência, incentivando os pesquisadores a reportarem essa métrica. Eles defendem que a IA Verde é mais inclusiva e sustentável do que a IA Vermelha, e apontam várias direções de pesquisa para avançar a eficiência na IA.

(vi) Strubell et al. (2019) analisaram o custo financeiro e ambiental do treinamento de redes neurais profundas para processamento de linguagem natural (PLN). Os autores estimam o consumo de energia, as emissões de carbono e o preço do treinamento de vários modelos populares de PLN, como Transformer, ELMo, BERT e GPT-2. Eles também realizam um estudo de caso do desenvolvimento de um modelo de ponta para análise sintática e semântica, mostrando que a pesquisa e o ajuste de hiperparâmetros exigem uma quantidade enorme de recursos computacionais. Com base nessas descobertas, eles propõem recomendações para reduzir os custos e melhorar a equidade na pesquisa e na prática de PLN, tais como: relatar o tempo de treinamento e a sensibilidade aos hiperparâmetros, prover acesso equitativo aos recursos computacionais e priorizar o desenvolvimento de modelos e hardware eficientes.

# 3. Metodologia

Este estudo adotou uma abordagem quantitativa e experimental para avaliar e comparar modelos de aprendizagem profunda.

Iniciamos com uma revisão sistemática da literatura, explorando os conceitos e técnicas de aprendizagem profunda, IA verde, sustentabilidade e métricas energéticas.

Identificamos e selecionamos as principais métricas usadas na literatura para avaliar modelos de aprendizagem profunda. Essas métricas incluíam medidas de desempenho, eficiência, gasto energético e custo computacional.

Na parte prática do estudo, optamos por trabalhar com dois modelos de redes neurais convolucionais amplamente referenciados na literatura: LeNet-5 e AlexNet. Esses modelos foram selecionados por apresentarem diferentes níveis de complexidade e capacidade computacional.

Utilizamos a base de dados CIFAR-10 (Krizhevsky et al., 2009), um conjunto de dados que contém 60.000 imagens coloridas em 10 classes diferentes, com 6.000 imagens por classe. As classes são avião, carro, pássaro, gato, cervo, cachorro, sapo, cavalo, barco e caminhão. O conjunto de dados tem 50.000 imagens para treinamento e 10.000 imagens para teste. Cada imagem é uma matriz de 32x32 pixels com três canais de cores (vermelho, verde e azul). O conjunto de dados CIFAR-10 é parte do conjunto de dados Tiny Images, que contém 80 milhões de imagens. Ele é frequentemente usado para avaliar a eficácia de algoritmos de aprendizado de máquina para tarefas de classificação de imagens, devido à sua complexidade e diversidade de imagens.

Os modelos foram implementados usando Python e a biblioteca PyTorch em um ambiente Linux, e treinados para classificar as imagens do conjunto CIFAR-10. Durante as fases de treinamento e inferência, utilizamos várias métricas, incluindo tempo de treinamento, tempo de inferência, número de parâmetros, número de operações de ponto flutuante (Flops), consumo energético do hardware e a emissão de carbono estimada.

Finalmente, analisamos e comparamos os resultados obtidos, utilizando técnicas estatísticas descritivas e gráficos para visualizar e comparar as métricas dos modelos. Discutimos os resultados à luz da literatura revisada e das questões de pesquisa propostas, destacando a importância da IA verde para um futuro mais sustentável.

# 4. Sections and Paragraphs

Section titles must be in boldface, 13pt, flush left. There should be an extra 12 pt of space before each title. Section numbering is optional. The first paragraph of each section should not be indented, while the first lines of subsequent paragraphs should be indented by 1.27 cm.

## 4.1. Subsections

The subsection titles must be in boldface, 12pt, flush left.

# 5. Figures and Captions

Figure and table captions should be centered if less than one line (Figure 1), otherwise justified and indented by 0.8cm on both margins, as shown in Figure 2. The caption font must be Helvetica, 10 point, boldface, with 6 points of space before and after each caption.



Figure 1. A typical figure



Figure 2. This figure is an example of a figure caption taking more than one line and justified considering margins mentioned in Section 5.

In tables, try to avoid the use of colored or shaded backgrounds, and avoid thick, doubled, or unnecessary framing lines. When reporting empirical data, do not use more decimal digits than warranted by their precision and reproducibility. Table caption must be placed before the table (see Table 1) and the font used must also be Helvetica, 10 point, boldface, with 6 points of space before and after each caption.

Table 1. Variables to be considered on the evaluation of interaction techniques



# 6. Images

All images and illustrations should be in black-and-white, or gray tones, excepting for the papers that will be electronically available (on CD-ROMs, internet, etc.). The image resolution on paper should be about 600 dpi for black-and-white images, and 150-300 dpi for grayscale images. Do not include images with excessive resolution, as they may take hours to print, without any visible difference in the result.

# 7. References

Bibliographic references must be unambiguous and uniform. We recommend giving the author names references in brackets, e.g. [Knuth 1984], [Boulic and Renault 1991]; or dates in parentheses, e.g. Knuth (1984), Smith and Jones (1999).

The references must be listed using 12 point font size, with 6 points of space before each reference. The first line of each reference should not be indented, while the subsequent should be indented by 0.5 cm.

# References

Krizhevsky, Alex and Hinton, Geoffrey and others (2009). Learning multiple layers of features from tiny images. Technical Report, University of Toronto.

Boulic, R. and Renault, O. (1991) “3D Hierarchies for Animation”, In: New Trends in Animation and Visualization, Edited by Nadia Magnenat-Thalmann and Daniel Thalmann, John Wiley & Sons ltd., England.

Dyer, S., Martin, J. and Zulauf, J. (1995) “Motion Capture White Paper”, <http://reality.sgi.com/employees/jam_sb/mocap/MoCapWP_v2.0.html>, December.

Holton, M. and Alexander, S. (1995) “Soft Cellular Modeling: A Technique for the Simulation of Non-rigid Materials”, Computer Graphics: Developments in Virtual Environments, R. A. Earnshaw and J. A. Vince, England, Academic Press Ltd., p. 449-460.

Knuth, D. E. (1984), The TeXbook, Addison Wesley, 15th edition.

Smith, A. and Jones, B. (1999). On the complexity of computing. In *Advances in Computer Science*, pages 555–566. Publishing Press.