



## REDES NEURAS PROFUNDAS APLICADAS NO DIAGNÓSTICO DE LEUCEMIA LINFOBLÁSTICA AGUDA

**Palavras-chave:** Aprendizado profundo, redes convolucionais, diagnóstico assistido por computador.

**Autores/as:**

*Marcos André Murguel de Farias*<sup>1</sup> [IMECC/Unicamp]

*Prof. Dr. Marcos Eduardo Ribeiro do Valle Mesquita (Orientador)* [IMECC/Unicamp]

---

### 1 Introdução

Nos últimos anos, o aprendizado de máquina influenciou a forma como resolvemos uma variedade de problemas do mundo real (Géron, 2019). De fato, as redes neurais artificiais superaram muitas abordagens de última geração em várias aplicações com o desenvolvimento de arquiteturas de redes neurais profundas e redes neurais convolucionais (CNN, do inglês *convolutional neural networks*). Em particular, redes neurais profundas foram aplicadas com sucesso no diagnóstico assistido por computador, incluindo diagnóstico automático da leucemia linfoblástica aguda (ALL, do inglês *acute lymphoblastic leukemia*).

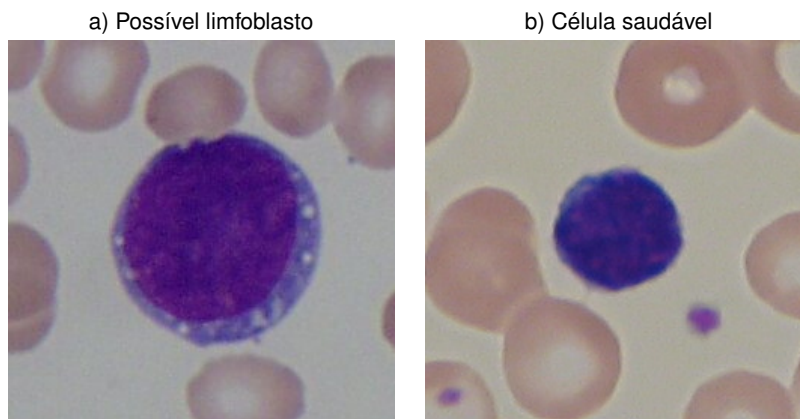
Existem vários métodos utilizados para o diagnóstico de ALL, incluindo a técnica de esfregaço de sangue periférico (Labati et al., 2011). Na técnica de esfregaço de sangue periférico, um especialista (hematologista) faz o diagnóstico contando o número de linfoblastos observados através de um microscópio numa amostra de sangue retirada do paciente. Infelizmente, a contagem manual de linfoblastos ao microscópio é uma tarefa um tanto monótona que toma muito tempo de um profissional que poderia ser mais produtivo em outros assuntos. Por esta e várias outras razões, modelos computacionais para realizar a contagem automática de linfoblastos em uma imagem de esfregaço de sangue têm sido propostos na literatura (Genovese et al., 2023).

O problema da contagem automática de linfoblastos pode ser dividido em duas etapas. A primeira etapa, geralmente chamada de fase de identificação ou segmentação, visa encontrar células na imagem de esfregaço de sangue. A rotulagem de uma célula candidata como linfoblasto ou célula saudável é realizada na segunda etapa, denominada fase de classificação. Neste projeto de iniciação científica estudamos redes neurais profundas para classificar as células brancas, ou seja, focamos na segunda etapa da análise da imagem do esfregaço de sangue. Especificamente, utilizamos redes neurais profundas para classificar uma imagem de uma célula como linfoblasto ou célula saudável. A Figura 1 ilustra exemplos das imagens utilizadas nessa tarefa de classificação binária.

---

<sup>1</sup>Bolsista de Iniciação Científica do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Figura 1: Exemplos de imagens usadas na tarefa de classificação.



Fonte: Imagens do conjunto de dados ALL-IDB (Labati et al., 2011).

## 2 Fundamentação Teórica

### 2.1 Redes Neurais Artificiais

As redes neurais artificiais são técnicas computacionais que se embasam em modelos matemáticos inspirados no cérebro humano, nas chamadas redes neurais biológicas (Géron, 2019). Do ponto de vista computacional, uma rede neural artificial é caracterizada pelo modelo dos neurônios, sua arquitetura e a regra de aprendizado. Os neurônios são as unidades básicas de processamento de uma rede neural artificial. O aprendizado de uma rede neural artificial pode ser classificado como supervisionado, não-supervisionado, entre outros. Nesse trabalho, estudamos apenas o aprendizado supervisionado no qual os parâmetros livres são ajustados minimizando uma função perda sobre um conjunto de treinamento  $\mathcal{T} = \{(\mathbf{x}_i, y_i) : i = 1, \dots, N\}$  conhecido *a priori*. Durante o treinamento num problema de classificação binária, incluindo o problema de rotulagem de uma imagem como linfoblasto ou célula saudável, podemos minimizar a entropia cruzada binária (*Binary Cross-Entropy*) dada por

$$H_p(q) = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log(p(\mathbf{x}_i)) + (1 - y_i) \cdot \log(1 - p(\mathbf{x}_i)), \quad (1)$$

em que  $N$  é o número de pares  $(\mathbf{x}_i, y_i)$  do conjunto de treinamento com  $\mathbf{x}_i$  sendo o padrão ou imagem de entrada,  $y_i \in \{0, 1\}$  seu rótulo e  $p(\mathbf{x}_i)$  a probabilidade do rótulo produzida pela rede neural artificial.

### 2.2 Camadas Convolucionais

As camadas convolucionais são primordiais numa rede neural convolucional (CNN). Essas camadas utilizam filtros, também conhecidas como *kernels*, que são usados para extrair informações dos dados de entrada. De um modo geral, os filtros percorrem toda a imagem atrás de informações em regiões específicas da imagem. De maneira informal, essas camadas captam diversas informações, primeiramente mais simples como retas, arestas e texturas brutas, informações essas que vão se tornando mais complexas até chegarmos em objetos, rostos e formas. A saída de neurônio de uma camada convolucional pode ser escrita por:

$$z_{i,j,k} = \varphi \left( b_k + \sum_{u=0}^{f_h-1} \sum_{v=0}^{f_w-1} \sum_{k'=0}^{f_n-1} x_{i',j',k'} \cdot w_{u,v,k',k} \right), \quad (2)$$

---

com

$$i' = i \times s_h + u \quad \text{e} \quad j' = j \times s_w + v,$$

em que

- $z_{i,j,k}$  é a saída do neurônio que está na localizado na linha  $i$  e na coluna  $j$  no mapa de características  $k$  de uma camada convolucional.
- $s_h$  e  $s_w$  são *strides* verticais e horizontais,  $f_h$  e  $f_w$  são, respectivamente, altura e largura do campo receptivo, já  $f_n$  é o número de mapas de características da camada anterior.
- $x_{i',j',k'}$  é a saída do neurônio que está na camada anterior, linha  $i'$ , coluna  $j'$  e no mapa de características  $k'$ .
- $b_k$  é um certo viés para o mapa de características, servindo como uma ajuste para o mapa de características  $k$ .
- $w_{u,v,k',k}$  é um certo peso da conexão entre um neurônio localizado em uma camada  $l$  com um mapa de características  $k$  e sua entrada que se localiza na coluna  $v$ , linha  $u$  e também ao mapa de características  $k'$ .
- $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  é a função de ativação, podendo ser a identidade, uma função sigmoide ou retificada linear (ReLU).

## 2.3 Camadas de Pooling

Além das camadas convolucionais, redes neurais profundas geralmente são equipadas com camadas de pooling. As camadas de pooling são usadas para redução de dimensionalidade da entrada. Seu funcionamento se dá por funções de agrupamento que simplificam as informações em áreas, normalmente  $2 \times 2$ , mantendo apenas as informações consideradas mais importantes. A camada de pooling ajuda a diminuir o número de parâmetros nas redes neurais, evitando sobre-ajuste (*overfitting*). Existem diversas funções de agrupamento, as mais comuns a serem usadas são a média, máximo e mínimo, as quais simplificam por meio das determinadas operações a região em que a área escolhida desliza pela imagem.

## 3 Implementação Computacional

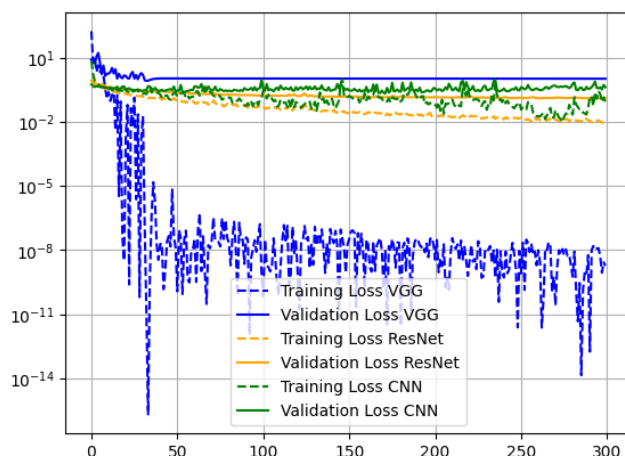
Nesse projeto de iniciação científica, consideramos três arquiteturas de redes neurais profundas: Uma variação da rede LeNet, a VGG e a ResNet. A implementação desses modelos para a classificação de linfoblastos para o diagnóstico de ALL está disponível no GitHub<sup>2</sup>.

Proposta por Yann LeCun em 1998, a LeNet foi a primeira arquitetura de rede neural convolucional para o reconhecimento de dígitos manuscritos (Lecun et al., 2015). Resumidamente, sua arquitetura conta com camadas densas, de pooling, totalmente conectadas e convolucionais. Por ter sido a primeira rede convolucional, desempenhou um papel primordial para o desenvolvimento da área. Atualmente, existem muitas variações da rede LeNet, com diversas arquiteturas que mesclam conhecimentos mais atuais com a arquitetura original. Aplicamos, nesta pesquisa, uma variação que conta com algumas diferenças da rede LeNet tradicional. A saber: A LeNet-5 é projetada para imagens em escalas de tons de cinza,

---

<sup>2</sup><https://github.com/MarcosMurguel/ICRedesNeuraisProfundasALL>

Figura 2: Função perda nos conjuntos de treino e validação pelo número de épocas no treinamento.



enquanto nossa variação trabalha com imagens RGB. Além disso, nosso modelo recebe como entrada imagens de  $126 \times 126$  pixels, enquanto a LeNet recebe imagens com tamanho de  $32 \times 32$  pixels.

Idealizada e construída pelo Grupo de Geometria Visual da Universidade de Oxford, no ano de 2014, a VGG (Simonyan and Zisserman, 2015) se destacou por sua arquitetura simples, compacta e com alto poderio de resolução para problemas de classificação de imagens. Utilizando, em sua maioria, convoluções  $3 \times 3$  e vertentes com 16 ou 19 camadas, a classe de redes VGG tornou-se uma referência para o processamento de dados não estruturados, com ênfase em imagens.

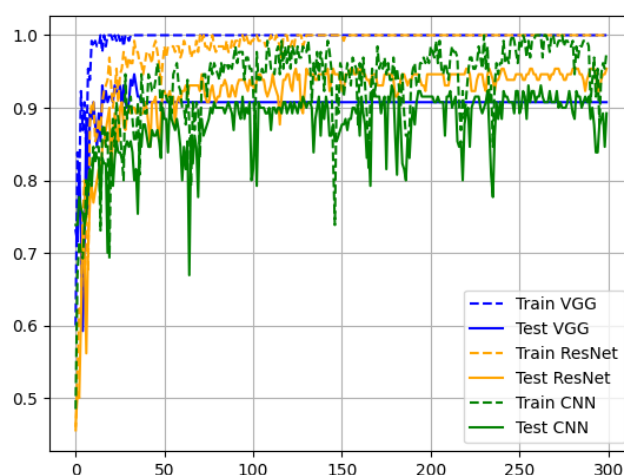
Finalmente, apresentada no ano de 2015 em um artigo chamado *"Deep Residual Learning for Image Recognition"* pelos pesquisadores da Microsoft, a ResNet surge como uma alternativa promissora para solucionar, com sua estrutura de lotes residuais, o "Problema da Dissipação do Gradiente" (He et al., 2016). Com efeito, a classe de redes profundas com blocos residuais, as chamadas ResNet, conseguiu superar um grande obstáculo do treinamento e criação de redes profundas, possibilitando a construção e redes mais profundas e complexas e, portanto, mais poderosas.

## 4 Resultados e Conclusão

Foram feitos gráficos comparativos para avaliar o desempenho dos modelos de redes neurais considerados. O gráfico na Figura 2 mostra o valor da função perda nos conjuntos de treino e validação pelo número de épocas. Note que há uma diferença significativa no valor da função perda da rede VGG nos conjuntos de treino e validação. Essa diferença sugere um sobre-ajuste da rede ao conjunto de treinamento. Ao contrário da rede VGG, não observamos uma diferença significativa na função perda produzida pelas CNN e ResNet nos conjuntos de treino e validação, sugerindo a hipótese de não existir sobre-ajuste neste caso.

A Figura 3 apresenta o gráfico da acurácia nos conjuntos de treinamento e validação em função das épocas na fase de treinamento. Em princípio, Um bom algoritmo deve "aprender" com seus dados de treino e, ao encontrar uma instância nova, conseguir classificá-la com alta acurácia. Temos um forte indicativo de sobre-ajuste quando um algoritmo que acerta muito no treino mas erra muito no conjunto de validação. Quando temos muitos erros no treino e validação, temos um forte indicativo de sub-ajuste (*underfitting*). Queremos, em um bom modelo, evitar ambos os casos tendo um bom trade-off entre o

Figura 3: Acurácia nos conjuntos de treino e validação pelo número de épocas no treinamento.



viés e a variância de nossos dados (Géron, 2019). Nesse sentido, o gráfico na Figura 3 reforça que a rede VGG apresentou um sobre-ajuste no conjunto de dados pois sua acurácia no conjunto de treino é 100% enquanto no conjunto de validação está em torno de 90%. Diferente da VGG, a rede CNN aparentemente apresentou um sob-ajuste, uma vez que o modelo não conseguiu apresentar um bom desempenho no conjunto de treinamento. Finalmente, a rede ResNet apresentou um bom desempenho em ambos conjuntos de treino e validação. Com efeito, a acurácia final nos conjuntos de treino e validação da rede ResNet foram 100% e 95,4%, respectivamente.

Concluindo, conforme nossa métrica de avaliação, a ResNet foi o modelo com melhor desempenho. Com efeito, fica nítido ao analisar os gráficos que sua arquitetura permite um ótimo valor de acurácia para nosso problema de ALL.

## Referências

- A. Genovese, V. Piuri, K. N. Plataniotis, and F. Scotti. DL4ALL: Multi-Task Cross-Dataset Transfer Learning for Acute Lymphoblastic Leukemia Detection. *IEEE Access*, pages 1–1, 2023. doi: 10.1109/ACCESS.2023.3289219. URL <https://ieeexplore.ieee.org/document/10162211/>.
- A. Géron. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O Reilly, Sebastopol, California, USA., 2nd edition, 10 2019.
- K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. doi: 10.1109/CVPR.2016.90.
- R. D. Labati, V. Piuri, and F. Scotti. All-IDB: The acute lymphoblastic leukemia image database for image processing. In *Proceedings - International Conference on Image Processing, ICIP*, pages 2045–2048, 2011. ISBN 9781457713033. doi: 10.1109/ICIP.2011.6115881.
- Y. Lecun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 5 2015. doi: 10.1038/nature14539.
- K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.