

# “LeukNet” - Um Modelo de Rede Neural Convolutacional para o Diagnóstico de Leucemia

Luis H. S. Vogado, Rodrigo M. S. Veras, Kelson R. T. Aires

<sup>1</sup>Departamento de Computação,  
Universidade Federal do Piauí - Teresina, Brasil

lhvogado@gmail.com, {rveras, kelson}@ufpi.edu.br

**Abstract.** *A leucemia é um tipo de câncer que afeta a produção de células sanguíneas na medula óssea o que dificulta a coagulação do sangue e o combate a infecções. Nesse trabalho propomos um método para o diagnóstico automático de leucemia utilizando Redes Neurais Convolutacionais (CNNs). Nós utilizamos CNNs pré-treinadas e técnicas de transferência de aprendizagem na construção do método proposto. Empregamos a técnica modified Deeply Fine Tuning (mDFT) combinada com operações de aumento de dados para refinar um modelo pré-treinado. Para treinar e testar o método proposto, utilizamos um conjunto de 3.536 imagens de 18 bases diferentes. A técnica de validação Leave-One-Dataset-Out Cross-Validation (LODOCV) foi proposta para avaliar a capacidade de generalização do modelo. Os principais resultados obtidos utilizando o LODOCV em três bases de dados foram 97,04, 82,46 e 70,24% de acurácia.*

## 1. Introdução

Leucemia é uma das doenças mais perigosas de acordo com o *American Cancer Society*<sup>1</sup>, tendo sido estimados cerca de 61.780 novos casos e 22.840 mortes em 2019. A leucemia é uma doença sem origem definida que afeta a produção de glóbulos brancos na medula óssea. Com a presença da enfermidade, células jovens ou blastos são produzidos de forma anormal, substituindo células sanguíneas normais (glóbulos brancos, vermelhos e plaquetas). Consequentemente o indivíduo sofre com problemas no transporte de oxigênio e no combate a infecções. Dentre as formas de diagnóstico da leucemia, temos a punção lombar, mielograma, hemograma e citometria de fluxo [Khosravan et al. 2019].

No intuito de auxiliar médicos no diagnóstico de doenças como a leucemia, são desenvolvidos sistemas conhecidos como *Computer-aided diagnosis* (CAD) [Shanthi and Sabeenian 2019] os quais utilizam imagens resultantes de exames de sangue, resultados de biópsia e ou outros que geralmente estão disponíveis em uma base de dados com exemplos. Os sistemas CAD empregam técnicas de processamento de imagens e aprendizado de máquina para produzir um diagnóstico suplementar. A classificação provida geralmente é em “saúdável” ou “doente”, “benigno” ou “maligno” e são frequentemente empregados para a triagem de doenças. Dessa forma, fornecendo um primeiro diagnóstico, ou oferecendo uma segunda opinião baseada em exemplos previamente rotulados.

Neste contexto, propomos o LeukNet, que se baseia em uma rede neural convolutacional (CNN) que utiliza conceitos de transferência de aprendizagem selecionados de

---

<sup>1</sup><https://cancerstatisticscenter.cancer.org/#!/cancer-site/Leukemia>

acordo com um extenso estudo de arquiteturas, estratégias de treinamento e análise do espaço de características. Portanto, o método *modified Deeply Fine-Tuning* (mDFT) foi empregado no treinamento do modelo proposto. A LeukNet foi avaliada em 3.536 imagens de lâminas de sangue pertencentes a diferentes bases de dados, incluindo hospitais e outras instituições. Cada conjunto de dados inclui imagens adquiridas sob diferentes condições, dimensões e características de cor, contraste e textura. Os resultados experimentais indicaram a necessidade de um protocolo de avaliação utilizando o *leave-one-dataset-out cross validation* (LODOCV), onde o teste é realizado em um conjunto de dados. Além disso, os conjuntos de dados restantes são usados no processo de treinamento. Este procedimento é executado até que todos os conjuntos de dados sejam testados individualmente. Isso garante que o CNN não seja treinado com nenhuma imagem dos conjuntos de dados a serem testados.

## 2. Materiais e Métodos

Durante a construção do modelo proposto, utilizamos conjuntos de técnicas e operações que buscam maximizar a sua acurácia. As bases de dados utilizadas nos experimentos propostos foram obtidas de diferentes fontes e contam com uma alta taxa de heterogeneidade, tornando a tarefa de classificação mais desafiadora.

### 2.1. Bases de dados

Um dos desafios na criação de um sistema de assistência médica é a capacidade de diagnosticar com precisão a partir de bases de imagens com características distintas. Para avaliar a capacidade de generalização dos sistemas, propõe-se o uso de 18 conjuntos de dados públicos, divididos em conjuntos de desenvolvimento e desempenho. Por meio da experimentação, utiliza-se o conjunto de desenvolvimento para definir a configuração ideal do modelo proposto. Esse conjunto foi formado por 17 bases de dados com 3.417 imagens. Essas imagens possuem heterogeneidade em termos de cor, contraste, resolução e textura, em que cada base possui um balanceamento diferente entre as classes, o que põe em teste a robustez do classificador.

O conjunto de desempenho é formado por uma base de dados nova, adquirida na Universidade Federal de Goiás (UFG), denominado conjunto de dados UFG<sup>2</sup>. Além disso, todas as bases avaliadas possuem imagens com diferentes resoluções. Sendo assim, realizamos o redimensionamento de todas as imagens para o tamanho 224x224, sendo esse o padrão da CNN utilizada.

### 2.2. Ajuste-fino

Técnicas de transferência de aprendizado frequentemente empregadas em aprendizado profundo utilizam pesos que são pré-treinados em grandes conjuntos de dados, como o conjunto de dados do *ImageNet challenge*, apresentado em [Russakovsky et al. 2015]. Esse procedimento diminui a necessidade de retrainar todos os parâmetros da CNN do zero [Yosinski et al. 2014]. A Figura 1 descreve esse procedimento. Pode-se observar que algumas camadas são copiadas da rede pré-treinada, formando uma arquitetura base, enquanto outras camadas são inicializadas aleatoriamente e personalizadas para a tarefa em questão.

---

<sup>2</sup>Link para o conjunto de dados da UFG: <https://hematologia.farmacia.ufg.br>

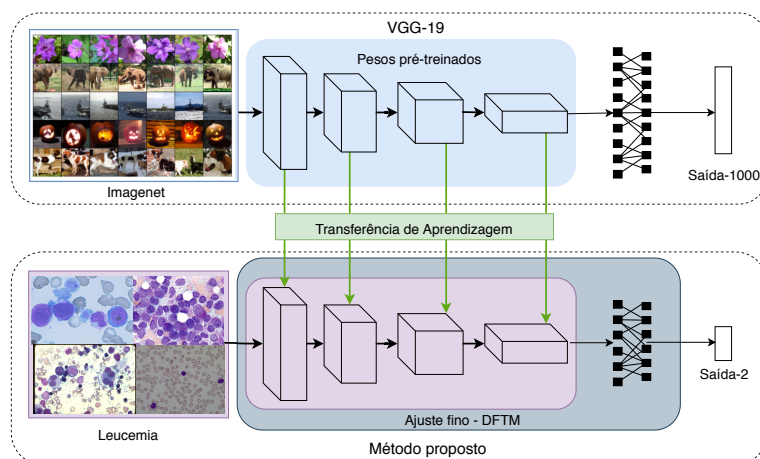


Figura 1. Fluxograma exemplificando a transferência de aprendizagem e ajuste fino utilizados no desenvolvimento do método proposto.

A transferência de aprendizagem pode se dividir em duas técnicas, a extração de características com camadas totalmente conectadas e o ajuste-fino. De acordo com a literatura existem dois tipos de ajuste fino: *Shallow Fine Tuning* (SFT) e *Deeply Fine-Tuning* (DFT). Sendo o primeiro com o congelamento das camadas convolucionais e o segundo consistindo no retreinamento de toda a arquitetura. Com base nisso, apresentamos versões modificadas das duas técnicas, sendo demoninadas da seguinte forma: *modified Shallow Fine-Tuning* (mSFT) e *modified Deeply Fine-Tuning* (mDFT). O mSFT e mDFT apresentam como principal característica a alteração nas camadas totalmente conectadas das arquiteturas originais, diminuindo a quantidade de parâmetros da rede, fornecendo assim um treinamento mais rápido e menos propenso ao *overfitting*.

### 2.3. Arquiteturas Avaliadas

Foram exploradas e avaliadas as arquiteturas de CNN's projetadas para o *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) [Russakovsky et al. 2015]. Redes sequenciais como VGG-16 e VGG-19 [Simonyan and Zisserman 2014] facilitam mudanças na estrutura da arquitetura, enquanto as redes Residuais e Inception apresentaram melhores resultados no ILSVRC. A ResNet50 [He et al. 2016], InceptionV3 [Szegedy et al. 2016] e Xception [Chollet 2017] possuem menos parâmetros que os VGGNets, mas têm mais profundidade.

### 3. Resultados e Discussão

Para definir o modelo proposto, foram efetuados experimentos LODOCV utilizando as bases de dados ALL-IDB 1 e ALL-IDB 2. O experimento consiste em separar uma única base de dados para o conjunto de teste enquanto as demais são utilizadas no treinamento da arquitetura. Os resultados são apresentados por meio das seguintes métricas: acurácia (A), precisão (P), *recall* (R) e especificidade (E). Como as novas camadas são treinadas com pesos inicializados aleatoriamente, um mesmo modelo foi executado cinco vezes para calcular a média e o desvio padrão das métricas.

Na Tabela 1 são apresentados os melhores resultados para as combinações entre as arquiteturas avaliadas e as técnicas de ajuste-fino. Dentre as abordagens, observa-se

que as técnicas de ajuste em que todos os pesos foram reajustados, obtiveram melhores resultados. Dentre essas, a abordagem mDFT alcançou os melhores resultados em todas as métricas calculadas para a base ALL-IDB 2 e em quatro das obtidas na base ALL-IDB 1. Ainda, destaca-se a arquitetura VGG-16 com os melhores resultados em ambas as bases de dados.

Tabela 1. Melhores resultados obtidos por arquiteturas e técnicas de ajuste-fino nas bases ALL-IDB 1 e ALL-IDB 2.

Arq.	Ajuste-fino	A(%)	P(%)	R(%)	E(%)
<b>ALL-IDB 1</b>					
VGG-16	mDFT	97,04±1,21	96,42±2,45	97,14±1,83	96,95±2,21
VGG-19	DFT	97,04±0,41	95,98±0,04	97,55±0,91	96,61±0,00
Inception V3	mDFT	65,56±9,79	58,47±17,18	73,92±16,99	60,91±19,90
Xception	mDFT	77,41±8,65	69,40±9,76	94,69±3,10	63,05±18,11
ResNet50	mDFT	87,96±2,70	91,59±8,56	82,04±4,42	92,88±8,08
<b>ALL-IDB 2</b>					
VGG-16	mDFT	82,46±0,02	77,59±0,04	92,30±0,08	72,61±0,09
VGG-19	mDFT	79,62±6,31	77,54±8,53	85,38±9,53	73,85±12,93
Inception V3	mDFT	58,38±3,09	56,95±2,42	70,00±12,38	46,77±11,66
Xception	mDFT	64,92±2,60	63,58±2,54	70,31±7,06	59,54±6,45
ResNet50	mDFT	69,46±6,26	66,96±7,05	80,31±10,17	58,62±16,71

O modelo final do LeukNet usa o backbone da VGG-16, com novas camadas densas com menor dimensionalidade e uma estratégia de treinamento baseada na transferência de aprendizagem com mDFT. Considerando os resultados obtidos por meio da definição do modelo proposto e da escolha da VGG-16 como arquitetura base. Foi realizada uma análise do espaço de características para justificar o porquê do resultado obtido com essa arquitetura e ajuste-fino foram superiores as outras avaliadas. Durante a análise foi verificado que o mDFT forneceu uma separabilidade linear superior dos dados para os experimentos cross-dataset.

Os resultados apresentados na seção anterior foram obtidos utilizando o LO-DOCV. Entretanto, os trabalhos na literatura não utilizaram esse tipo de validação. Isso decorre provavelmente da falta de múltiplas bases de dados para construir um conjunto de treinamento representativo. Portanto, foi aplicado o *k-fold cross validation*, com  $k = 5$ , para comparar o modelo proposto com métodos do estado da arte.

Na Tabela 2, comparamos o modelo proposto com o trabalho apresentado por Vogado et al. [Vogado et al. 2018] em dois experimentos distintos. Primeiramente, comparamos os resultados obtidos por cada metodologia utilizando o conjunto de dados apresentado neste trabalho e a validação por meio do *k-fold*. Observamos ainda que o melhor resultado foi obtido pelo modelo proposto com 98,24% de acurácia contra 92,79%.

Tabela 2. Comparação entre a LeukNet e o método proposto por Vogado et al. [Vogado et al. 2018] com k-fold e LODOCV.

Approach	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)
k-fold				
Vogado et al. [Vogado et al. 2018]	92,79	92,90	92,80	92,22
<b>LeukNet</b>	<b>98,24</b>	<b>98,20</b>	<b>98,76</b>	<b>97,34</b>
LODOCV - UFG				
Vogado et al. [Vogado et al. 2018]	52,06	49,90	52,10	47,70
<b>LeukNet</b>	<b>70,24±5,51</b>	<b>70,54±8,62</b>	<b>80,31±15,17</b>	<b>58,94±22,24</b>

O segundo experimento apresentado na Tabela 2, consiste na utilização da base

de dados UFG como conjunto de teste por meio do LODOCV. O conjunto UFG é um conjunto de dados novo, que nunca foi usado em estudos anteriores, e é particularmente desafiador por três razões: (1) o conjunto de dados é formado por imagens adquiridas por diferentes microscópios e de acordo com diferentes resoluções e condições de iluminação; (2) tem imagens com lâminas de sangue com diversos leucócitos e imagens com apenas um leucócito; e (3) entre todos os conjuntos de dados usados neste estudo, o conjunto de dados UFG é aquele com a maior diversidade dentro da classe de leucemia, uma vez que possui exemplos de imagens com os quatro principais subtipos de leucemia.

Da Tabela 2, pode-se concluir que as duas abordagens obtiveram resultados inferiores quando comparadas àquelas obtidas pelo k-fold. Porém, a decadência do modelo proposto foi mais moderada (de 98,24 a 70,24%) do que a de Vogado et al. [Vogado et al. 2018] (de 92,79 a 52,06 %). Esse resultado sugere que o LeukNet pode generalizar melhor do que os métodos da literatura. Assim, pode-se concluir que essa generalização superior se deve ao uso de maior quantidade de dados e à definição precisa dos parâmetros da rede neural convolucional como foi realizado neste estudo.

#### **4. Conclusão**

Neste trabalho, é proposta uma nova arquitetura de CNN e uma estratégia de treinamento para o diagnóstico de leucemia em imagens de lâminas de sangue. Diferentes arquiteturas, parâmetros e técnicas de ajuste fino foram estudados para definir nosso modelo que resulta em diagnósticos mais precisos e robustos do que as reportadas no estado da arte.

A partir da comparação com estudos anteriores, algumas conclusões podem ser tiradas para o diagnóstico de leucemia usando imagens: primeiro, o ajuste fino é uma técnica de transferência de aprendizado mais eficiente que a extração de características por permitir melhor generalização entre conjuntos de dados diferentes. Também foi constatado que os modelos CNN sequenciais e de menor profundidade e menor complexidade na quantidade de parâmetros são mais adequados para cenários em que se precisa classificar imagens de diferentes fontes. Além disso, a escolha da técnica de ajuste fino é essencial para obter melhores resultados. Como as imagens de amostras de sangue pertencem a um domínio diferente daquele usado para pré-treinar as camadas, é preferível ajustar todas as camadas.

Por meio da análise do espaço de características, foi constatado que o modelo proposto consegue generalizar melhor as características quando comparado com outros modelos pré-treinados, inclusive na ImageNet. Outra contribuição proposta foi a união de 18 bases de dados públicas, totalizando uma base com 3536 imagens que até então não foi observada em trabalhos da literatura. A validação de dados por meio do LODOCV demonstrou que métodos da literatura que utilizaram outros meios de validação, talvez não desempenhem bem em conjuntos de dados nunca vistos. Uma forma de verificar esse problema é com a utilização do LODOCV. Realizar a validação com essa técnica permite a criação de abordagens mais eficientes no reconhecimento de dados nunca vistos ou de bases distintas.

#### **Agradecimentos**

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Também foi

recebido uma GPU Nvidia GeForce Titan XP a partir do *NVIDIA GPU Grant Program* por meio do projeto ao qual este trabalho pertence.

## Artigos Publicados

Artigos resultantes do trabalho apresentado foram publicados em evento e revista.<sup>3</sup>

## Referências

- [Chollet 2017] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 1800–1807.
- [He et al. 2016] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778. IEEE Computer Society.
- [Khosravan et al. 2019] Khosravan, N., Celik, H., Turkbey, B., Jones, E. C., Wood, B., and Bagci, U. (2019). A collaborative computer aided diagnosis (c-cad) system with eye-tracking, sparse attentional model, and deep learning. *Medical Image Analysis*, 51:101–115.
- [Russakovsky et al. 2015] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252.
- [Shanthi and Sabeenian 2019] Shanthi, T. and Sabeenian, R. (2019). Modified alexnet architecture for classification of diabetic retinopathy images. *Computers and Electrical Engineering*, 76:56–64.
- [Simonyan and Zisserman 2014] Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556.
- [Szegedy et al. 2016] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 2818–2826.
- [Vogado et al. 2018] Vogado, L. H. S., Veras, R. M. S., Araújo, F. H. D., e Silva, R. R. V., and Aires, K. R. T. (2018). Leukemia diagnosis in blood slides using transfer learning in cnns and SVM for classification. *Engineering Applications of Artificial Intelligence*, 72:415–422.
- [Yosinski et al. 2014] Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). How transferable are features in deep neural networks? In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2, NIPS'14*, pages 3320–3328, Cambridge, MA, USA.

---

<sup>3</sup>SBCAS 2019: <https://doi.org/10.5753/sbcas.2019.6241>  
Sensors: <https://doi.org/10.3390/s21092989>