

Refinamento e Combinação de CNNs para Melhoria na Classificação de Leucócitos Imaturos no Diagnóstico de Leucemia Mieloide Aguda

Émery F. Moriconi¹, Leonardo P. Sousa¹, Maila Claro², Ana A. F. Rocha¹,
Rodrigo M. S. Veras¹

¹Departamento de Computação,
Universidade Federal do Piauí - Teresina, Brasil

²Instituto Federal do Piauí - Teresina, Brasil

{emerymoriconi09, claromaila, allyciaroxha}@gmail.com
{leonardosousa, rveras}@ufpi.edu.br

Abstract. *This article applied techniques such as fine-tuning, data augmentation, and committees in Convolutional Neural Networks (CNNs) to classify leukemia cells in microscopic images of blood, focusing on immature leukocytes. The main objective is to classify these cells into four distinct subtypes present in the database, which contains 5,738 blood smear images. The study achieved an accuracy of 84.37% with the committee formed by the CNNs MobileNetV2, EfficientNetB3, and DenseNet201 using the majority vote committee technique in three different combinations of datasets.*

Resumo. *Este artigo aplicou técnicas, como ajuste fino, aumento de dados e comitês em Redes Neurais Convolucionais (Convolutional Neural Networks - CNNs), para a classificação de células leucêmicas em imagens microscópicas de sangue, com foco em leucócitos imaturos. O principal objetivo é classificar essas células em quatro subtipos distintos presentes na base de dados, que contém um total de 5.738 imagens de esfregaços de sangue. Utilizando a técnica de comitês com voto majoritário em três combinações diferentes de bases de dados, o estudo alcançou uma acurácia de 84,37% com o comitê formado pelas CNNs MobileNetV2, EfficientNetB3 e DenseNet201.*

1. Introdução

A leucemia é um tipo de câncer prevalente que afeta diretamente o funcionamento da medula óssea, um tecido esponjoso responsável pela produção de células sanguíneas, incluindo glóbulos brancos, essenciais para a defesa contra infecções [Travlos 2006]. Existem vários tipos de leucemia, como a Leucemia Mieloide Aguda (LMA), a Leucemia Mieloide Crônica (LMC), a Leucemia Linfoide Aguda (LLA) e a Leucemia Linfoide Crônica (LLC), cada um com características específicas. A forma aguda da doença se manifesta rapidamente com a produção de células imaturas incapazes de desempenhar suas funções normais. Em contraste, a forma crônica progride mais lentamente, com um número maior de células maduras presentes nos pacientes.

A LMA é conhecida por ser a variante mais agressiva e letal dentre os quatro tipos de leucemia. Segundo um estudo feito pelo Observatório de Oncologia, em dez anos,

entre os anos de 2008 e 2017, o Brasil teve cerca de 63 mil mortes por leucemias, sendo que 36% foram causadas pela LMA. Nessa forma da doença, há uma proliferação rápida de leucócitos imaturos, também conhecidos como blastos, que se acumulam na medula óssea e interferem na produção de células sanguíneas.

A detecção de leucócitos imaturos é crucial para o diagnóstico precoce de leucemias, principalmente de leucemias agudas, pois a presença elevada de blastos na medula óssea pode ser um indicador chave da doença em muitos casos. Os principais tipos de leucócitos imaturos são os Promielócitos, Mielócitos, Metamielócitos e os Eritroblastos, como podem ser visualizados na Figura 1. Estudos mostram que a identificação precoce de blastos pode melhorar significativamente o prognóstico dos pacientes, permitindo a implementação rápida de terapias adequadas [Bain 2010]. Assim, a distinção entre leucócitos imaturos e suas classificações, desempenha um papel fundamental na determinação precisa da natureza da doença e na escolha do tratamento mais eficaz.

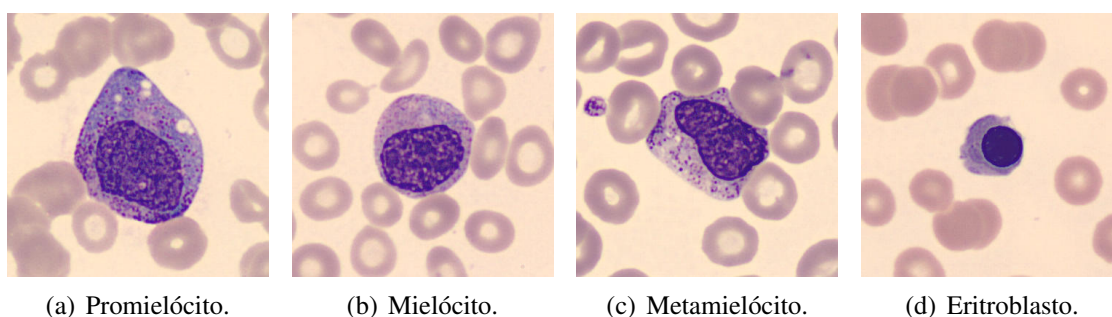


Figura 1. Exemplos de tipos de leucócitos imaturos.

O diagnóstico da leucemia, que inclui exames de sangue, biópsias de medula óssea e testes de imagem, é um processo demorado, invasivo, emocionalmente desafiador e financeiramente oneroso [Prinyakupt and Pluempitiwiriyawej 2015]. A detecção precoce é crucial para reduzir custos hospitalares e permitir intervenções terapêuticas rápidas e eficazes. Sistemas de apoio médico utilizam técnicas avançadas de processamento de imagens e inteligência artificial, como sistemas de Diagnóstico Auxiliado por Computador (CAD), para ajudar médicos a identificar padrões e anomalias em dados médicos, melhorando a precisão diagnóstica [Vogado et al. 2017].

A metodologia deste projeto propõe aprimorar técnicas de processamento de imagens e inteligência artificial para classificar células leucêmicas imaturas em imagens microscópicas de esfregaços de sangue. Serão implementados e testados modelos de aprendizagem profunda, especialmente Redes Neurais Convolucionais (CNNs), em sistemas CAD. O algoritmo avaliará diversas CNNs em diferentes conjuntos de imagens para classificar leucócitos imaturos em quatro classes: Promielócitos, Mielócitos, Metamielócitos e Eritroblastos. Isso visa melhorar a precisão e eficiência do diagnóstico, auxiliando na detecção precoce de doenças e na condução de tratamentos adequados.

O trabalho em questão está organizado da seguinte forma: Seção 2, que discorre a respeito da literatura existente a respeito do tema. A Seção 3, que descreve a metodologia aplicada para o desenvolvimento do projeto, bem como uma descrição acerca das bases de imagens médicas selecionadas e os métodos e parâmetros de avaliação propostos; e as Seções 4 e 5, que apresentam os resultados obtidos, assim como uma discussão a respeito

do avanço do projeto e de perspectivas futuras.

2. Trabalhos Relacionados

Ao longo dos anos, diversas técnicas, descritores e classificadores foram desenvolvidos e testados por pesquisadores com objetivo de apresentar soluções que envolvam o diagnóstico de leucemia em imagens de lâminas de sangue. No entanto, até o momento, a literatura sobre a divisão de classes de leucócitos imaturos é limitada, e as pesquisas realizadas não identificaram projetos que utilizem a mesma abordagem. Apesar disso, nessa Seção, serão abordadas algumas pesquisas relacionadas ao tema.

Thanh et al. [Thanh et al. 2018], propuseram um modelo de CNN para avaliar a distinção entre células normais e anormais em imagens de células sanguíneas. A CNN proposta foi treinada com a base de dados ALL-IDB1 [Labati et al. 2011], que consiste em 108 imagens (59 saudáveis e 49 portadoras de leucemia), com o auxílio de técnicas de aumento de dados, como equalização de histogramas, reflexões e rotações. Devido à aplicação dessas técnicas, o conjunto de dados resultou em 1188 imagens de células sanguíneas. O resultado para esse método foi uma acurácia de 96,6%. Entretanto, devido à proposta de classificação binária entre células sanguíneas normais e anormais, o projeto não definiu outras classes entre os leucócitos.

Rahman and Ahmad [Rahman and Ahmad 2023] desenvolveram uma metodologia para classificação de leucócitos em maduros e imaturos, empregando arquiteturas de Redes Neurais Convolucionais como a AlexNet, ResNet50, DenseNet161 e VGG-16. A base de dados utilizada foi coletada do *The Cancer Imaging Archive* [Clark et al. 2013] e reuniu 18.365 imagens de leucócitos. O modelo que apresentou o melhor desempenho foi o AlexNet modificado, que alcançou uma precisão de 96,52% e um *F1-Score* de 97,00%. O teste foi realizado com apenas um *fold* e foram aplicadas técnicas de aumento de dados. Assim como o último trabalho citado, não houve definição dos leucócitos em outros subtipos.

Dasariraju et al. [Dasariraju et al. 2020], utilizaram o algoritmo *Random Forest* para treinar a detecção e classificação de leucócitos imaturos em quatro tipos: Eritroblastos, Monoblastos, Promielócitos e Mieloblastos. O conjunto de dados pertence ao *The Cancer Imaging Archive* [Clark et al. 2013] e envolveu 1.274 imagens de leucócitos, entre eles, os quatro subtipos citados e 600 imagens de células maduras para uma classificação binária inicial. O modelo alcançou 92,99% de acurácia para a detecção de leucemia ou não, e 93,45% para a classificação de leucócitos imaturos. Contudo, essa classificação multiclasse difere em duas classes da classificação proposta pelo estudo abordado neste artigo: as classes Monoblasto e Mieloblasto. Além disso, os valores de precisão para cada uma das classes apresentaram grande variação de 65% a 100%, assim como os valores de *recall* individuais, apresentando uma variação de 75% a 96%.

No trabalho de Santos et al. [Santos et al. 2022] os autores aplicaram algoritmos de aprendizado de máquina para a classificação de células nucleadas do sangue periférico em oito classes diferentes: Neutrófilos, Eosinófilos, Basófilos, Linfócitos, Monócitos, Granulócitos Imaturos, Eritroblastos e Plaquetas. Nesse caso, subtipos de células imaturas e maduras foram agrupadas em um mesmo banco de dados. Para a classificação de saída foi escolhido o classificador *Support Vector Machine* (SVM) e foram usadas imagens disponibilizadas por um estudo de classificação de imagem do Hospital das Clínicas

de Barcelona [Acevedo et al. 2020], contendo 17.092 imagens. O modelo obteve uma acurácia média de 97,2%, e o F1-Score médio de 97%. Contudo, a definição das classes não é a mesma utilizada por profissionais da saúde, visto que, houve a união de diferentes leucócitos imaturos em uma só classe, como Metamielócitos, Mielócitos e Promielócitos, além da adição de Plaquetas ao conjunto de classes, que não são classificadas como leucócitos, mas sim, como apenas pequenos fragmentos celulares presentes no sangue.

A partir das obras analisadas, fica evidente que o desenvolvimento e o uso de Redes Neurais Convolucionais (CNNs) são essenciais para o alcance do objetivo do projeto em questão. É possível identificar modelos que obtiveram resultados benéficos e a escolha de parâmetros que contribuíram para tais resultados. Dessa forma, pode-se determinar uma direção para a atual proposta e atingir os objetivos desejados.

3. Materiais e Métodos

Nesta seção são detalhados os recursos e o método empregado no desenvolvimento deste trabalho. O objetivo em questão envolve a classificação de imagens de células leucêmicas imaturas em quatro principais subtipos presentes nas bases de imagens selecionadas: Promielócitos, Mielócitos, Metamielócitos e Eritroblastos.

Para isso, foram selecionadas, ajustadas e avaliadas as CNNs pré-treinadas MobileNet, ResNet50, DenseNet121, MobileNetV2, ResNet101, VGG19, DenseNet201 e EfficientNetB3. Para a implementação das CNNs, foi utilizada a linguagem de programação *Python*, juntamente com as bibliotecas de Aprendizado Profundo *Keras* e *TensorFlow*.

Nos tópicos seguintes abordamos detalhadamente o método proposto, a base de imagens selecionada e suas características, técnicas de aumento de dados, as CNNs selecionadas, o uso de comitê para a classificação final e as métricas de avaliação utilizadas para desempenho.

3.1. Método Proposto

O método proposto pelo projeto em questão pode ser visualizado na Figura 2. Após o treinamento e a avaliação do desempenho individual de cada CNN, os três modelos com os melhores valores de acurácia são selecionados para a formação de um comitê. Assim, foi gerado um conjunto de votação majoritária que utiliza as classificações individuais de cada modelo para a classificação final das imagens nas quatro classes trabalhadas.

Os modelos e o comitê final são treinados em três execuções utilizando três conjuntos de imagens com características e balanceamento de classes diferentes. O *Shallow Fine Tuning* foi aplicado durante 65 épocas, com um tamanho de lote (*batch size*) de 32 imagens, e o desempenho final do comitê foi avaliado por métricas comumente usadas no estado da arte, como a Acurácia, Precisão, *Recall*, F1-Score e o índice *Kappa*. Ao final, foi realizada uma comparação do desempenho desse comitê sobre as três combinações de bases de imagens utilizadas.

3.2. Base de Imagens

Para o desenvolvimento do projeto, foi utilizado um conjunto de dados contendo 5.738 imagens de lâminas de sangue. Esse conjunto é dividido em quatro principais classes de

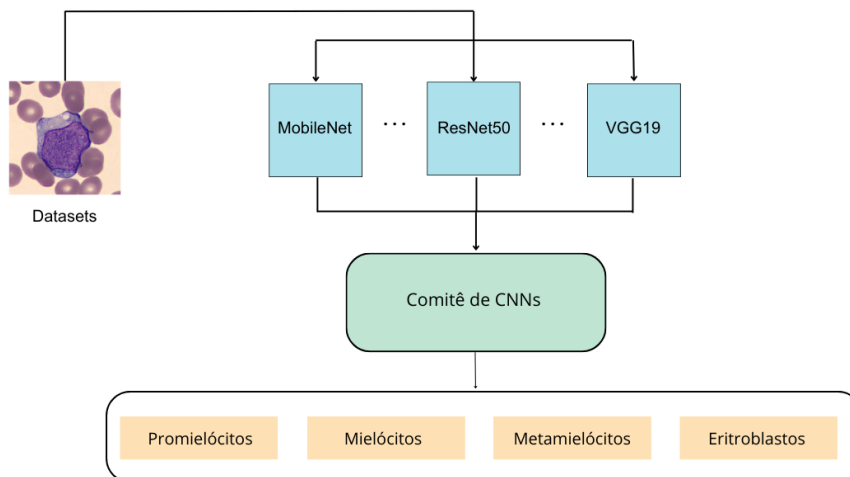


Figura 2. Fluxograma do método proposto utilizando um comitê de CNNs.

leucócitos imaturos. Os leucócitos maduros são células totalmente desenvolvidas, prontas para atuar no sistema imunológico, enquanto os leucócitos imaturos podem indicar uma resposta do organismo a infecções ou inflamações agudas. A diferenciação entre essas células e seus subtipos é crucial para o estudo.

As imagens foram organizadas em dois subconjuntos menores, o primeiro, coletado da base de imagens do Hospital Universitário de Munique [Matek et al. 2019], onde são encontradas 1.443 imagens de esfregaços de sangue com leucócitos imaturos de 100 pacientes diagnosticados com Leucemia Mieloide Aguda (LMA). Esses dados foram coletados no período entre 2014 e 2017, abrangendo 4 principais classes de leucócitos imaturos: Promielócitos, Mielócitos, Metamielócitos e Eritroblastos. O segundo conjunto foi obtido do Hospital Clínico de Barcelona [Boldú et al. 2021], na Espanha, e contém um total de 4.295 imagens de leucócitos imaturos, com mesma subdivisão do conjunto Munique. A Tabela 1 ilustra a combinação das imagens utilizadas.

Tabela 1. Resumo da combinação das bases de imagens utilizadas.

Leucócitos Imaturos	Munique	Barcelona	Total por classes
Promielócito	667	592	1259
Promielócito Biobled			
Mielócito	181	1137	1318
Metamielócito	75	1015	1090
Eritroblasto	520	1551	2071
Total	1443	4295	5738

As imagens utilizadas possuem dimensões de 360x363, e foram redimensionadas para 224x224, tamanho geralmente requerido pela maioria dos modelos pré-treinados utilizados. Com o objetivo de evitar problemáticas como o desbalanceamento de classes, e para fins de comparação de desempenho das CNNs entre as bases de imagens, as redes foram submetidas a três fases de treinamento: inicialmente utilizando a base Munique, depois a base Barcelona e por fim, combinando ambas as bases. Na Figura 3, podemos

visualizar exemplos dos tipos de leucócitos imaturos nas duas bases de imagens utilizadas.

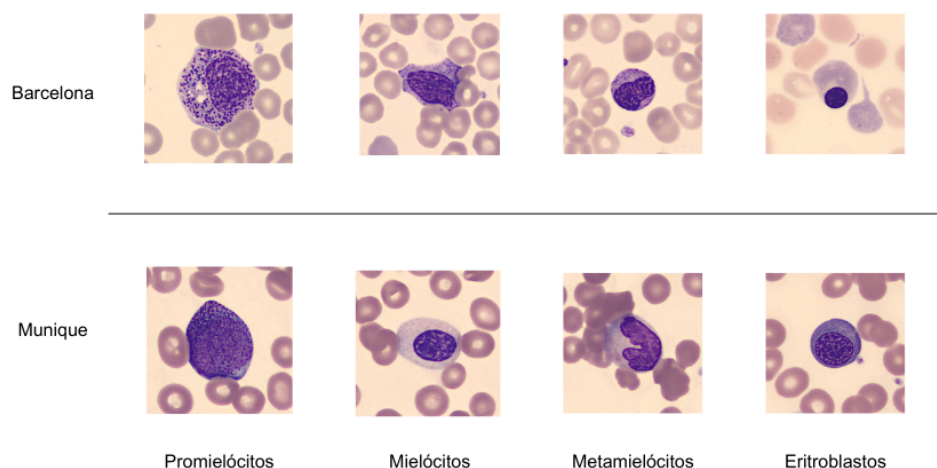


Figura 3. Tipos de leucócitos imaturos nas bases Barcelona e Munique.

3.3. Aumento de Dados

O aumento de dados é um conjunto de técnicas que visa melhorar a qualidade e diversidade dos dados para aprimorar o desempenho dos modelos de aprendizado, criando novas amostras a partir das já existentes. De acordo com Mumuni e Mumuni [Mumuni and Mumuni 2022], conjuntos de dados maiores, mais diversificados e representativos melhoram a eficácia dos modelos de aprendizado profundo em dados não observados. Para evitar problemas como *overfitting*, foram aplicadas técnicas de aumento de dados nas imagens de entrada de forma *online*, ou seja, as transformações são aplicadas dinamicamente a cada época durante o treinamento. As imagens foram rotacionadas em até 20 graus, deslocadas horizontal e verticalmente em até 20% da largura e altura, respectivamente, e sofreram cisalhamento de até 20%. Além disso, o zoom das imagens variou em até 20%, e a inversão horizontal foi utilizada para espelhar as imagens.

3.4. Redes Neurais Convolucionais

Redes Neurais Convolucionais (CNNs) são uma classe de algoritmos de aprendizado profundo projetadas para processar e analisar dados. Elas utilizam camadas convolucionais que aplicam filtros sobre a entrada, extraindo características importantes enquanto preservam relações espaciais. Em problemas de classificação, as CNNs aprendem automaticamente a identificar padrões e características distintas nas imagens, permitindo a categorização precisa dos objetos nelas contidos [Tajbakhsh et al. 2016].

Esses modelos podem ser utilizados na classificação de imagens médicas e auxiliar no diagnóstico mais preciso de doenças, como a leucemia. Nesse projeto, foram avaliadas oito CNNs pré-treinadas utilizando a base de dados ImageNet, que contém milhões de imagens rotuladas distribuídas em milhares de categorias. As CNNs selecionadas foram MobileNet, ResNet50, DenseNet121, MobileNetV2, ResNet101, VGG19, DenseNet201 e EfficientNetB3, que fazem parte das arquiteturas que obtiveram melhores performances diante do conjunto de validação da ImageNet.

3.5. Transferência de Aprendizagem

A transferência de aprendizagem é uma técnica de aprendizagem de máquina onde um modelo pré-treinado em uma grande base de dados, como o ImageNet, é adaptado para uma nova tarefa, utilizando conhecimentos adquiridos anteriormente.

O modelo proposto utiliza o *Shallow Fine-Tuning* (SFT), onde, de acordo com [Tajbakhsh et al. 2016], apenas as camadas finais são ajustadas, enquanto as primeiras camadas convolucionais, permanecem inalteradas. Esse método é menos intensivo em termos de dados e computação, e a utilização dessa técnica ajuda o novo modelo a obter melhor desempenho com menos tempo de treinamento.

3.6. Validação Cruzada

O projeto utilizou a validação cruzada estratificada *k-fold* para avaliar a performance dos modelos de aprendizado de máquina, dividindo o conjunto de dados em 5 partes ou "folds". Em cada iteração, um *fold* foi usado como conjunto de teste e os demais para treinamento com 80% dos dados, e para validação, com 20%, repetindo o processo cinco vezes para garantir que cada *fold* fosse usado uma vez como teste. Essa abordagem preserva a proporção das classes em cada *fold*, importante para conjuntos de dados desequilibrados, evitando vieses e detectando possíveis *overfitting*.

3.7. Comitês

A abordagem de comitês combina as forças de diferentes classificadores, aumentando a precisão e a robustez das previsões, melhorando a capacidade de generalização [Dietterich 2000]. Este estudo utilizou um comitê entre as três CNNs com melhores acurácias para implementar um sistema de voto majoritário, onde a decisão final é baseada na maioria dos votos dos modelos. No comitê com voto majoritário, cada CNN faz sua previsão e a classe que recebe mais votos é selecionada como a previsão final, garantindo uma decisão mais confiável e equilibrada ao considerar múltiplas perspectivas.

3.8. Métricas de Avaliação

As métricas utilizadas para avaliar o desempenho do modelo foram acurácia, precisão, *recall*, *F1-Score* e o índice *Kappa* [Powers 2020]. A acurácia mede a proporção de previsões corretas em relação ao total de previsões feitas, fornecendo uma visão geral da eficácia do modelo. A precisão indica a proporção de verdadeiros positivos entre todas as previsões positivas, refletindo a exatidão das previsões positivas do modelo. O *recall* (ou sensibilidade) é a proporção de verdadeiros positivos entre todos os casos que realmente são positivos, mostrando a capacidade do modelo de identificar corretamente os casos positivos. O *F1-Score* é a média harmônica da precisão e do *recall*, oferecendo um balanço entre ambos e sendo especialmente útil quando há um desequilíbrio entre as classes. De acordo com [Rosenfield and Fitzpatrick-Lins 1986], o índice *Kappa* é um coeficiente de concordância para escalas nominais que mede o relacionamento entre a concordância, além da casualidade e a discordância esperada.

4. Resultados e Discussão

Fazendo o uso de técnicas mencionadas como transferência de aprendizagem utilizando redes pré-treinadas e o *Shallow Fine Tuning* (SFT); técnicas de validação cruzada estratificada como o *k-fold Cross Validation* com $k = 5$ folds; foram utilizadas oito CNNs para

treinamento e seus resultados foram selecionados para a formação e avaliação de comitês. Os experimentos e resultados detalhados nessa Seção foram alcançados utilizando um computador com processador Intel(R) Core(TM) i5 de 2,90 GHz, 16 GB de RAM e placa gráfica Nvidia Zotac GeForce RTX 3060 12 GB com 3.584 núcleos.

Os modelos foram treinados individualmente e um comitê foi formado a partir dos desempenhos das melhores CNNs em três bases de imagens diferentes: a base Barcelona, a base Munique e a base Barcelona unida à base Munique. As CNNs que apresentaram os melhores resultados individuais para as métricas de avaliação utilizadas foram a MobileNetV2, com 82,05% de acurácia, a EfficientNetB3, com 82,28% e a DenseNet201, alcançando 81,63% para a acurácia. A partir da formação do comitê, a votação majoritária foi executada e os resultados obtidos podem ser visualizados na Tabela 2. Os desempenhos foram avaliados nas três combinações de bases de imagens, para fins comparativos.

Tabela 2. Melhores resultados para o comitê de CNNs avaliado.

Comitê	Base de Imagens	A (%)	P (%)	R (%)	F1-Score (%)	K (%)
MobileNetV2 EfficientNetB3 DenseNet201	Barcelona	83,03 ± 0,015	83,14 ± 0,015	83,03 ± 0,015	83,05 ± 0,015	76,62 ± 0,021
MobileNetV2 EfficientNetB3 DenseNet201	Munique	82,05 ± 0,020	83,53 ± 0,013	82,05 ± 0,020	82,59 ± 0,017	72,33 ± 0,028
MobileNetV2 EfficientNetB3 DenseNet201	Barcelona e Munique	84,37 ± 0,010	84,61 ± 0,010	84,37 ± 0,010	84,43 ± 0,010	78,66 ± 0,014

Os melhores resultados obtidos durante o treinamento estão em negrito.

Os melhores resultados foram alcançados pelo comitê proposto ao utilizar como base de imagens, a união entre a base Munique e a base Barcelona. A acurácia alcançada foi de 84,37%, a precisão foi de 84,61%, o *recall* foi de 84,37%, o F1-Score foi de 84,43% e o índice Kappa alcançado foi de 78,66%. Com base na Figura 4 podemos visualizar um gráfico comparativo entre os desempenhos do comitê quando submetido às três bases de imagens diferentes.

Embora haja proximidade nos valores das métricas, podemos observar que o melhor desempenho do comitê foi alcançado utilizando a união das bases Barcelona e Munique como conjunto de imagens de entrada. Devido à junção, a base resultante apresenta uma maior diversidade intraclasse e uma maior heterogeneidade de características em comparação às outras duas bases. Por conta de uma menor totalidade de imagens e do desbalanceamento entre as classes, o desempenho do comitê utilizando a base Munique como entrada, não foi superior aos outros resultados. Além disso, é importante ressaltar que os estudos anteriores tiveram um enfoque maior em uma classificação binária, entre leucócitos maduros e imaturos, o que difere do que foi proposto neste estudo. Os trabalhos que propuseram uma classificação multiclasse, não apresentaram a mesma abordagem quanto a divisão das classes de leucócitos imaturos nas bases de imagens. Isso conferiu um caráter inédito ao estudo, que apresentou resultados promissores conforme à abordagem proposta.

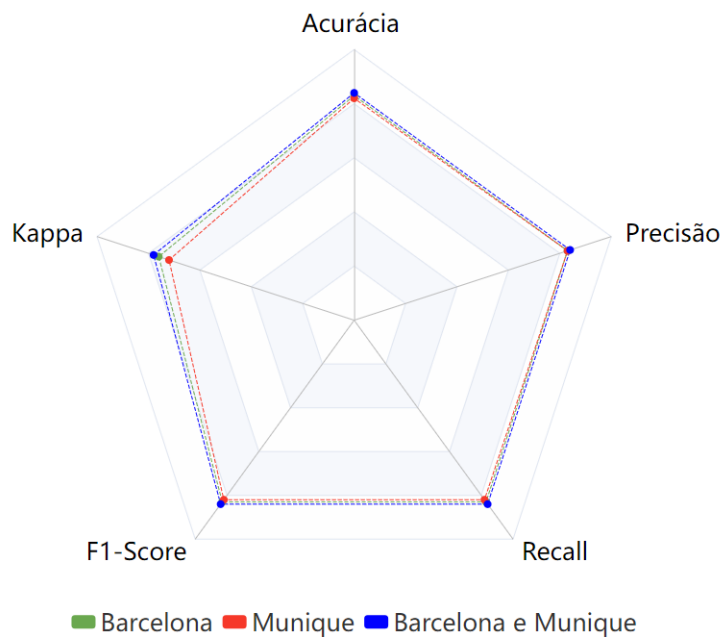


Figura 4. Comparação do desempenho do comitê com as três bases de imagens diferentes.

5. Conclusão

Este estudo apresentou uma proposta de treinamento de CNNs, e formação e avaliação de comitês com voto majoritário para a classificação de imagens de leucócitos imaturos em quatro subtipos. Três combinações de bases de imagens foram utilizadas e o comitê foi avaliado utilizando as três como entrada. Os resultados demonstraram a importância da seleção da base de imagens de entrada, tendo em vista que, conjuntos maiores e com mais diversidade tornam os modelos robustos e mais precisos durante a classificação.

É possível afirmar que resultados promissores foram alcançados e que o aperfeiçoamento futuro de algumas técnicas pode aprimorá-los. Entre eles, podemos citar o objetivo de aumentar a base de imagens utilizada aplicando técnicas de aumento de dados de forma *offline*, o que tornaria o treinamento mais vasto, além do uso de outras técnicas de refinamento, como o *Deep Fine Tuning* (DFT), visando ajustar todas as camadas das CNNs à base de imagens de entrada, e não apenas as camadas finais.

Referências

- Acevedo, A., Merino, A., Alférez, S., Ángel Molina, Boldú, L., and Rodellar, J. (2020). A dataset of microscopic peripheral blood cell images for development of automatic recognition systems. *Data in Brief*, 30:105474.
- Bain, B. J. (2010). Myeloid and lymphoid neoplasms with eosinophilia and abnormalities of *pdgfra*, *pdgfrb* or *fgfr1*. *Haematologica*.
- Boldú, L., Merino, A., Acevedo, A., Molina, A., and Rodellar, J. (2021). A deep learning model (alnet) for the diagnosis of acute leukaemia lineage using peripheral blood cell images. *Computer Methods and Programs in Biomedicine*, 202:105999.
- Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., and Prior, F. (2013). The cancer imaging archive

- (tcia): Maintaining and operating a public information repository. *Journal of Digital Imaging*.
- Dasariraju, S., Huo, M., and McCalla, S. (2020). Detection and classification of immature leukocytes for diagnosis of acute myeloid leukemia using random forest algorithm. *Bioengineering*.
- Dietterich, T. G. (2000). Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, pages 1–15. Springer.
- Labati, R. D., Piuri, V., and Scotti, F. (2011). All-idb: the acute lymphoblastic leukemia image database for image processing. *Proc. of the 2011 IEEE Int. Conf. on Image Processing (ICIP 2011)*.
- Matek, C., Schwarz, S., Spiekermann, K., and Marr, C. (2019). Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks. *Nature Machine Intelligence*, 1(11):538–544.
- Mumuni, A. and Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. *Array*, 16:100258.
- Powers, D. M. W. (2020). Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation.
- Prinyakupt, J. and Pluempitiwiriyaewej, C. (2015). Segmentation of white blood cells and comparison of cell morphology by linear and naïve bayes classifiers. *BioMed Eng OnLine*.
- Rahman, J. and Ahmad, M. (2023). *Detection of Acute Myeloid Leukemia from Peripheral Blood Smear Images Using Transfer Learning in Modified CNN Architectures*, pages 447–459.
- Rosenfield, G. H. and Fitzpatrick-Lins, K. (1986). A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering and Remote Sensing*.
- Santos, M., Bertemes, W., de Souza, I., Andrades, M., Barros, D., and Patto, V. (2022). Algoritmos de aprendizado de máquina para classificação de células nucleadas do sangue periférico - uma experiência do projeto hemovision. In *Anais da X Escola Regional de Informática de Goiás*, pages 130–140, Porto Alegre, RS, Brasil. SBC.
- Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., and Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE transactions on medical imaging*, 35(5):1299–1312.
- Thanh, T. T. P., Vununu, C., Atoev, S., Lee, S.-H., and Kwon, K.-R. (2018). Leukemia blood cell image classification using convolutional neural network. *International Journal of Computer Theory and Engineering*, 10:54–58.
- Travlos, G. S. (2006). *Normal structure, function, and histology of the bone marrow.*, volume 34.
- Vogado, L., Veras, R., Andrade, A., Santos, L., Aires, K., and Machado, V. (2017). Um sistema de diagnóstico de leucemia utilizando cnn's pré-treinadas e um comitê de classificadores. In *Anais do XVII Workshop de Informática Médica*.