Abordagem Híbrida CNN-Mamba para Diagnóstico Computacional de Leucemias Agudas em Imagens Hematológicas

Gilclécio S. Morais[1], Gabriel C. Sousa[1], Maíla L. Claro[1], Rodrigo M. S. Veras[2], Julian R. Valerio[1], Selles G. F. C. Araújo[1], Clésio A. Gonçalves[1], José C. J. Silva[1]

[1]Departamento de Informática, Instituto Federal do Piauí (IFPI)
Paulistana – PI – Brasil

[2]Departamento de Computação, Universidade Federal do Piauí (UFPI)
Teresina – PI – Brasil
{alguemum7668, gabrielc77877}@gmail.com, maila.claro@ifpi.edu.br, rveras@ufpi.edu.br,
{julian.valerio, selles.gustavo, clesio.araujo, justino}@ifpi.edu.br

Abstract. *Automated classification of acute leukemias from peripheral blood smear images remains challenging due to high morphological variability arising from differences in staining protocols, acquisition devices, and multicenter conditions. Although Convolutional Neural Networks (CNNs) are effective at extracting local features, they show limitations in modeling long-range spatial dependencies in highly heterogeneous scenarios. This study proposes a hybrid CNN–Mamba architecture for multiclass classification using the UNION macro-dataset, which comprises 17 independent public datasets totaling 3,157 cellular samples. The model combines local feature extraction through a pretrained MobileNetV2 backbone with global modeling based on Visual State Space Blocks (Mamba), enabling efficient capture of structural dependencies with linear complexity. The experimental protocol included stratified cross-validation, class weighting to address imbalance, and evaluation across 18 controlled scenarios. The best-performing model achieved a macro F1-score of 98.6% and an accuracy of 98.7% on an independent test set, outperforming convolutional baselines and demonstrating greater robustness to heterogeneity and reduced sample size. These findings indicate that hybrid CNN–State Space architectures represent a promising, scalable, and robust approach for AI-assisted hematological diagnostic systems.*

Resumo. *A classificação automatizada de leucemias agudas por imagens de esfregaço sanguíneo é desafiadora devido à alta variabilidade morfológica causada por diferentes protocolos de coloração, dispositivos de aquisição e contextos multicêntricos. Embora Redes Neurais Convolucionais (CNNs) sejam eficazes na extração de características locais, apresentam limitações na modelagem de dependências espaciais globais em cenários heterogêneos. Este trabalho propõe uma arquitetura híbrida CNN–Mamba para classificação multiclasse utilizando o macro-dataset UNION, composto por 17 bases públicas e 3.157 amostras celulares. O modelo combina a extração local da MobileNetV2 pré-treinada com modelagem global baseada em Visual State Space Blocks (Mamba), permitindo capturar dependências estruturais com complexidade linear. O protocolo experimental incluiu validação cruzada estratificada,*

ponderação de classes e análise de 18 cenários controlados. O melhor modelo alcançou F1-score macro de $98,6\%$ *e acurácia de* $98,7\%$ *em teste independente, superando baselines convolucionais e demonstrando maior robustez frente à heterogeneidade e à redução amostral. Os resultados indicam que arquiteturas híbridas CNN–State Space são uma alternativa promissora, escalável e robusta para sistemas de apoio ao diagnóstico hematológico assistido por inteligência artificial.*

1. Introdução

A leucemia é um grupo de neoplasias hematológicas caracterizadas pela proliferação descontrolada de células sanguíneas imaturas, afetando significativamente a mortalidade global [Siegel et al. 2023]. O diagnóstico convencional envolve análise morfológica de esfregaços sanguíneos ou aspirado de medula óssea, procedimento que depende fortemente da experiência do hematologista e pode apresentar variações inter e intraobservador. Nesse contexto, sistemas computacionais de apoio ao diagnóstico têm sido desenvolvidos para aumentar a precisão, reduzir o tempo de análise e promover maior padronização clínica.

O avanço do aprendizado profundo, especialmente após a consolidação das redes neurais convolucionais (CNNs) no reconhecimento de imagens [Krizhevsky et al. 2012], transformou o processamento de imagens médicas. Arquiteturas como ResNet [He et al. 2016] demonstraram capacidade superior de extração hierárquica de características, permitindo capturar padrões estruturais complexos presentes em tecidos biológicos. No domínio hematológico, CNNs têm sido aplicadas na classificação automática de células leucêmicas com resultados promissores, particularmente em bases públicas como a ALL-IDB [Labati et al. 2011].

Apesar do elevado desempenho alcançado por arquiteturas profundas, modelos puramente convolucionais podem apresentar limitações na modelagem de dependências globais e exigir grande volume de dados rotulados para adequada generalização. Para mitigar tais desafios, abordagens híbridas vêm sendo investigadas, combinando a extração automática de características locais por meio de CNNs com mecanismos mais eficientes de modelagem estrutural global. Recentemente, Modelos de Espaço de Estados (*State Space Models* – SSMs), como o Mamba [Ma and Wang 2024], surgiram como alternativa eficiente para captura de dependências de longo alcance com complexidade linear. Apesar desse avanço, ainda há escassez de estudos que integrem CNNs e SSMs para classificação multiclasse de leucemias sob variabilidade interlaboratorial significativa.

Diante desse cenário, este trabalho propõe uma abordagem híbrida que emprega CNNs para a extração de características morfológicas profundas de células sanguíneas e integra o Mamba como módulo de modelagem global de dependências estruturais. O objetivo central é que a combinação entre representações profundas locais e modelagem sequencial baseada em *State Space Models* pode proporcionar melhor equilíbrio entre desempenho, robustez estatística e capacidade de generalização, contribuindo para o desenvolvimento de sistemas confiáveis de apoio ao diagnóstico hematológico assistido por inteligência artificial.

Diante desse cenário, este trabalho propõe uma arquitetura híbrida CNN–Mamba para classificação multiclasse de leucemias agudas em imagens hematológicas proveni-

entes de múltiplas bases independentes. A abordagem integra extração local de características morfológicas por meio de uma CNN pré-treinada com modelagem global baseada em *Visual State Space Blocks*, permitindo aprendizado *end-to-end* eficiente e escalável.

A principal hipótese investigada é que a combinação entre representações convolucionais locais e modelagem sequencial baseada em SSMs pode aumentar a robustez estatística e a capacidade de generalização em ambientes multicêntricos heterogêneos. Para avaliar essa hipótese, conduzimos experimentos sistemáticos em um macro-*dataset* composto por 17 bases públicas independentes, analisando o impacto da resolução espacial, da disponibilidade de dados e da profundidade arquitetural no desempenho preditivo. Os resultados indicam que a integração CNN–*State Space* proporcionou leve melhora de desempenho, com indícios de maior estabilidade entre execuções, sugerindo que essa combinação merece investigação adicional em aplicações de apoio ao diagnóstico hematológico automatizado.

2. Trabalhos Relacionados

A aplicação de Inteligência Artificial ao diagnóstico hematológico evoluiu significativamente nos últimos anos, acompanhando o avanço das arquiteturas profundas para visão computacional. Inicialmente, abordagens baseadas exclusivamente em Redes Neurais Convolucionais (CNNs) dominaram o cenário, devido à sua capacidade de extrair padrões morfológicos locais relevantes em imagens microscópicas.

Entre os trabalhos pioneiros, [Vogado et al. 2020] propuseram a LeukNet, uma CNN otimizada para classificação de leucemia. Embora tenham explorado a variabilidade interlaboratorial utilizando 18 bases independentes, a abordagem puramente convolucional apresenta limitações na modelagem explícita de dependências espaciais globais, dificultando a generalização plena em cenários altamente heterogêneos.

Posteriormente, abordagens híbridas passaram a explorar a combinação entre extração profunda de características e classificadores externos. [Hidayat et al. 2023] integraram uma CNN para extração de atributos com XGBoost para classificação final. Embora essa estratégia tenha demonstrado melhorias em estabilidade preditiva em comparação a CNNs puras, o estudo permaneceu restrito a um único *dataset*, não avaliando robustez sob variabilidade interlaboratorial.

Com a consolidação dos *Transformers* Visuais [Dosovitskiy et al. 2020], novas arquiteturas passaram a modelar dependências globais de forma mais explícita. Entretanto, o alto custo computacional quadrático desses modelos limita sua escalabilidade em cenários clínicos com restrições de hardware. Mais recentemente, os Modelos de Espaço de Estados (*State Space Models* – SSMs) emergiram como alternativa eficiente para modelagem de dependências de longo alcance com complexidade linear. O modelo Mamba [Gu and Dao 2024] introduziu um mecanismo de varredura seletiva capaz de capturar relações estruturais globais mantendo eficiência computacional.

No domínio médico, Ma e Wang [Ma and Wang 2024] propuseram o Semi-Mamba-UNet para segmentação de imagens médicas, evidenciando ganhos em eficiência e desempenho. De forma complementar, [Kuang et al. 2026] integraram CNNs e Mamba para segmentação tridimensional de lesões em ressonância magnética. Ainda que tais trabalhos consolidem o uso de SSMs em imagens médicas, seu foco principal permanece

em segmentação, não explorando de forma sistemática tarefas de classificação celular multiclasse sob heterogeneidade multicêntrica.

No contexto específico da hematologia computacional, observa-se que a maioria dos estudos ainda se concentra em arquiteturas puramente convolucionais ou em combinações com classificadores tradicionais, havendo escassez de investigações que integrem modelagem sequencial baseada em SSMs para captura explícita de dependências estruturais globais em imagens celulares.

Diferentemente das abordagens que utilizam *boosting* externo como etapa final de classificação, o presente trabalho propõe uma integração arquitetural *end-to-end* entre extração convolucional profunda e modelagem sequencial baseada em *Visual State Space Blocks* (VSS), permitindo aprendizado conjunto otimizado durante todo o processo de treinamento. Além disso, a validação é conduzida em um macro-*dataset* multicêntrico composto por 17 bases independentes, abordando explicitamente o desafio de variabilidade interlaboratorial, ainda pouco explorado de maneira abrangente na literatura.

Dessa forma, este estudo posiciona-se na interseção entre eficiência computacional, modelagem global de dependências estruturais e robustez multicêntrica, contribuindo para o avanço de sistemas confiáveis de apoio ao diagnóstico hematológico automatizado. A Tabela 1 apresenta um resumo dos trabalhos encontrados na literatura.

Tabela 1. Resumo comparativo dos trabalhos.

Trabalho	Tipo de Arquitetura	Tarefa	Principal Contribuição	Limitação Central
[Vogado et al. 2020]	CNN LeukNet	Classificação	Eficiência computacional e heterogeneidade	Limitação em capturar dependências globais
[Hidayat et al. 2023]	CNN + XGBoost	Classificação	Integração deep features + boosting	Avaliação em um único *dataset*
[Ma and Wang 2024]	Mamba + U-Net	Segmentação	Introdução de SSM em imagens médicas	Foco em segmentação, não classificação
[Kuang et al. 2026]	CNN-Mamba híbrido	Segmentação 3D	Consolidação do Mamba em 3D	Pouca exploração em classificação celular

3. Materiais e Métodos

Neste capítulo, são detalhados os conceitos, as técnicas e as ferramentas empregados na elaboração da abordagem proposta. Durante a construção do modelo, utilizou-se um conjunto de operações com o objetivo de contornar os desafios impostos por imagens obtidas de múltiplas fontes, maximizando a capacidade de generalização da rede.

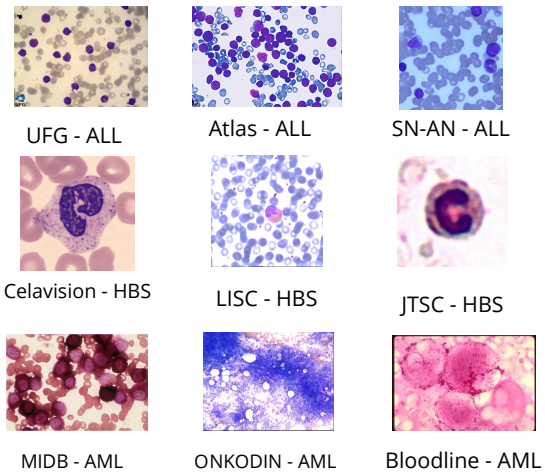
3.1. Base de Dados

Um dos maiores desafios no desenvolvimento de sistemas de auxílio ao diagnóstico de leucemia é garantir que o modelo mantenha alta precisão quando submetido a imagens com características visuais distintas daquelas vistas no treinamento. A grande maioria dos trabalhos na literatura restringe suas avaliações a conjuntos de dados únicos ou homogêneos, o que limita a comprovação da eficácia clínica dos modelos no mundo real. Para superar essa limitação e criar um ambiente de teste eficiente, este trabalho utiliza o *UNION*, um macro-dataset multicêntrico formado pelo agrupamento de 17 bases de dados públicas e independentes (cujo detalhamento e *links* oficiais estão disponíveis em `https://github.com/G1L404/UNION-Leukemia-Dataset`).

O acervo consolida um total de 3.157 amostras celulares categorizadas em três classes principais: Células Saudáveis (*Healthy Blood Slides* - HBS), Leucemia Linfoide

Aguda (ALL) e Leucemia Mieloide Aguda (AML). A fusão destas múltiplas fontes herda um desbalanceamento quantitativo natural entre as patologias e introduz uma acentuada heterogeneidade intrínseca, conforme ilustrado visualmente pela Figura 1. Nota-se que as amostras originais diferem drasticamente em níveis de iluminação, espectros de coloração química, padrões de textura, presença de artefatos de fundo e resoluções nativas.

Figura 1. Amostras extraídas do dataset UNION, ilustrando a severa variação visual inter-laboratorial.



3.2. Pré-processamento e Aumento de Dados

Para mitigar o *overfitting* e garantir a estabilidade do modelo híbrido [Chen et al. 2024], implementou-se um método sistemático de aumento de dados. Como testes empíricos revelaram que técnicas estáticas de normalização de coloração (como os métodos de Reinhard ou Macenko) introduziam artefatos de fundo severos, a invariância colorimétrica foi aprendida dinamicamente via *Color Jittering* (variações estocásticas de $\pm 20\%$ em brilho, contraste e saturação, além de alterações de matiz), forçando a rede a focar exclusivamente na morfologia estrutural do leucócito [Kuang et al. 2026].

Simultaneamente, aplicaram-se espelhamentos geométricos (*Random Horizontal/Vertical Flip*) e rotações de até $30°$. Os tensores foram padronizados via *Z-score* (pesos do *ImageNet*) e submetidos a redimensionamento em múltiplas escalas (64×64, 128×128 e 224×224 pixels). De acordo com Yu e Wang [Yu and Wang 2024], a eficácia dos Modelos de Espaço de Estados está intrinsicamente ligada ao comprimento da sequência de processamento. Portanto, a modulação da resolução afeta diretamente o campo receptivo do Mamba, fundamentando a avaliação empírica do ”Impacto da Resolução”neste estudo.

3.3. Arquitetura Híbrida CNN-*Mamba*

Para enfrentar o desafio de capturar as texturas locais do citoplasma e a morfologia global do leucócito simultaneamente, propõe-se a arquitetura híbrida CNN-*Mamba* (Figura 2). Conforme demonstrado recentemente em aplicações de imagens médicas [Yue and Li 2024, Ma and Wang 2024], a fusão da extração espacial das CNNs com a modelagem sequencial de longo alcance dos Modelos de Espaço de Estados (*State Space Models* - SSM) garante representações visuais bastante robustas. O modelo opera em

três estágios: *backbone* convolucional, adaptação sequencial e processamento estruturado (*Visual State Space Blocks* - Blocos VSS).

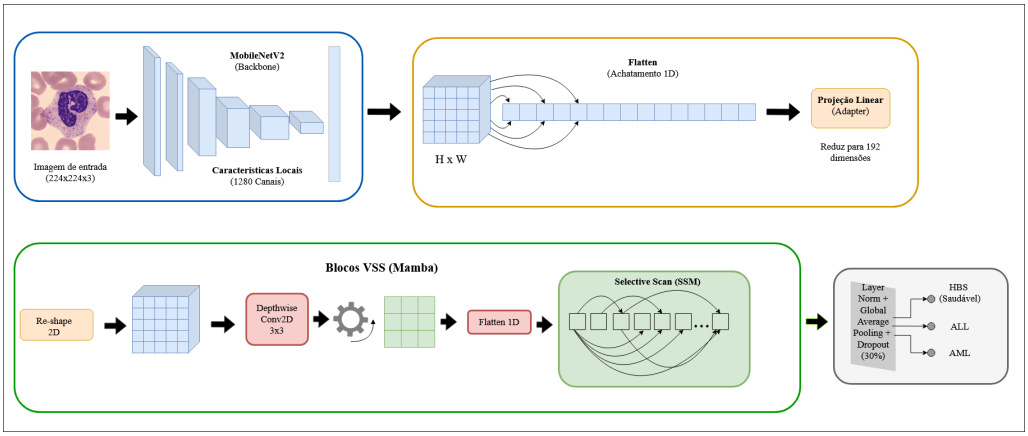


Figura 2. Visão geral da arquitetura proposta.

Como extrator local, adotou-se a arquitetura *MobileNetV2* pré-treinada (*ImageNet1K*), congelando os pesos iniciais para reaproveitar filtros genéricos de borda e liberando as camadas profundas para o *fine-tuning* nos padrões de leucemia. O tensor de saída da CNN mapeia características locais em $C = 1280$ canais nas dimensões espaciais de altura (H) e largura (W). Para adequar este mapa à natureza puramente sequencial do Mamba, o tensor $\mathbf{X} \in R^{B \times C \times H \times W}$ é achatado (*flattened*) em uma sequência 1D de comprimento $L = H \times W$. Na sequência, uma projeção linear restringe a dimensionalidade para o espaço interno do Mamba ($D_{model} = 192$).

O núcleo do modelo é fundamentado nos *Visual State Space Blocks* (VSS). Diferente de arquiteturas baseadas em *Transformers* que sofrem de gargalos computacionais, o Mamba assegura complexidade linear operando sistemas contínuos mapeados em equações diferenciais [Zhu et al. 2024]: $h'(t) = \mathbf{A}h(t) + \mathbf{B}x(t)$ e $y(t) = \mathbf{C}h(t) + \mathbf{D}x(t)$. A adaptação estrutural à CNN ocorre via varredura seletiva (*selective scan*). Sendo a dimensão do estado $N = 16$ e a interna $D = 384$, a matriz de transição $\mathbf{A} \in R^{D \times N}$ atua como um fator de esquecimento invariante no tempo, inicializado em escala logarítmica. Em contrapartida, as matrizes de projeção $\mathbf{B}, \mathbf{C} \in R^{B \times L \times N}$ e o passo $\mathbf{\Delta} \in R^{B \times L \times D}$ são gerados dinamicamente por projeções lineares sobre o *token* x_t, permitindo seleção contextual da rede. Com os parâmetros discretizados ($\bar{\mathbf{A}} = \exp(\mathbf{\Delta A})$ e $\bar{\mathbf{B}} = \mathbf{\Delta B}$), calcula-se o estado oculto e a saída iterativamente:

$$h_t = \bar{\mathbf{A}} h_{t-1} + \bar{\mathbf{B}} x_t, \quad y_t = \mathbf{C} h_t \tag{1}$$

Devido à natureza não sequencial dos dados de visão bidimensional, modelagens puras de Mamba podem perder a noção de continuidade espacial [Liu et al. 2024]. Assim, um detalhe importante da nossa implementação é a restauração momentânea da localidade 2D no Bloco VSS. Antes de injetada nas equações do Mamba, a sequência é reformatada para a grade $H \times W$ original e processada por uma *Depthwise Conv2D* (3×3), garantindo que o SSM não perca o viés indutivo de proximidade dos pixels da célula. Por fim, a sequência enriquecida passa por *LayerNorm*, *Global Average Pooling* e *Dropout* (30%), alimentando o classificador linear final para as classes HBS, ALL e AML.

3.4. Desenho Experimental e Métricas de Avaliação

Para assegurar a confiabilidade estatística e mitigar vieses da heterogeneidade do dataset *UNION*, o desenho experimental adotou a Validação Cruzada Estratificada em K-grupos [Li et al. 2023]. O treinamento investigou 18 cenários controlados, combinando: resoluções espaciais (64×64, 128×128 e 224×224 pixels), disponibilidade percentual de dados (25%, 50% e 100%) e profundidade do módulo Mamba (2 e 4 blocos).

A otimização utilizou o algoritmo *AdamW* aliado ao agendador dinâmico *Cosine Annealing*. Para contornar o severo desbalanceamento intrínseco gerado pela fusão das 17 bases, implementou-se a Entropia Cruzada Ponderada (*Weighted Cross-Entropy Loss*), aplicando pesos inversamente proporcionais à frequência de cada classe para penalizar majoritariamente erros em patologias minoritárias [Vogado et al. 2020].

O desempenho preditivo dos modelos foi quantificado em um conjunto de teste independente através das métricas: **Acurácia Global**, **Precisão**, **Sensibilidade (*Recall*)** e **F1-*Score* (Macro)** [Powers 2020, Steyerberg et al. 2010, Metz 1978]. Devido ao alto desbalanceamento, o F1-*Score* foi definido como métrica primária para avaliar a eficácia da arquitetura [Saito and Rehmsmeier 2015]. Adicionalmente, a **Matriz de Confusão** foi empregada para a análise detalhada dos padrões de erro anatômicos entre as classes de maior similaridade [Hastie et al. 2009].

4. Resultados e Discussões

Nesta seção são apresentados e discutidos os resultados experimentais obtidos a partir do treinamento e avaliação de dezoito modelos distintos baseados na arquitetura híbrida CNN–Mamba. Os experimentos foram estruturados de modo a investigar sistematicamente o impacto de três fatores principais no desempenho de classificação: (i) a resolução de entrada, (ii) a proporção de dados de treinamento e (iii) a profundidade do módulo Mamba (2 ou 4 blocos). Para cada combinação desses parâmetros, foram analisadas métricas quantitativas de desempenho, permitindo avaliar não apenas a eficácia individual de cada configuração, mas também as interações entre as variáveis experimentais.

4.1. Visão Geral dos Resultados

O melhor desempenho foi obtido pelo modelo com 2 camadas mamba e resolução 224×224, utilizando 100% do *dataset*, alcançando $98,6\%$ de F1-*score* e $98,7\%$ de Acurácia, superando todas as demais variações testadas.

Os resultados foram obtidos exclusivamente no conjunto de teste independente, garantindo uma estimativa imparcial do desempenho. As métricas selecionadas, Acurácia, Precisão, *Recall* e F1-*score*, foram escolhidas por fornecerem uma avaliação robusta em cenários de classificação multiclasse e desbalanceamento entre categorias celulares [Hastie et al. 2009].

De modo geral, a análise dos modelos com melhor desempenho evidencia que a combinação entre alta resolução de entrada e maior disponibilidade de dados exerce papel determinante na performance final. O modelo com resolução de 224 pixels treinado com 100% dos dados e 2 blocos Mamba apresentou o melhor resultado global (F1 = 0,986), indicando que representações espaciais mais detalhadas associadas a um volume completo de treinamento favorecem a capacidade discriminativa. Em seguida, o modelo com

Tabela 2. Desempenho dos três melhores modelos CNN-*Mamba* no conjunto de teste

Resolução	*Dataset* (%)	Blocos	Acurácia	Precisão	*Recall*	F1-*score*
224	100	2	**0,987**	**0,987**	**0,985**	**0,986**
128	100	4	0,977	0,975	0,974	0,974
224	50	4	0,968	0,966	0,966	0,966

resolução intermediária (128 pixels), 100% dos dados e 4 blocos Mamba alcançou F1 = 0,974, e o modelo com resolução de 224 pixels e 50% do dataset obteve F1 = 0,966, sugerindo que o aumento da profundidade do módulo sequencial pode compensar parcialmente reduções na resolução de entrada e na quantia de dados disponíveis.

4.2. Análise Detalhada do Melhor Modelo

Com o objetivo de avaliar de forma aprofundada o comportamento do modelo com melhor desempenho global, foi realizada uma análise abrangente considerando comparação com métodos de referência e matriz de confusão. Essa análise permite examinar não apenas o desempenho agregado, mas também a capacidade discriminativa individual para cada categoria celular.

A Tabela 3 apresenta a comparação entre o modelo proposto e métodos baseline selecionados da literatura. Os modelos com (*) são apenas para contextualização, pois foram treinados e testados em um dataset diferente do utilizado neste trabalho.

Tabela 3. Comparação com modelos baseline

Modelo	Acurácia	Precisão	*Recall*	F1-*score*
LeukNet [Vogado et al. 2020]*	0,982	0,982	**0,987**	-
CNN-XGBoost [Hidayat et al. 2023]*	0,854	0,883	0,904	-
ResNet50 com Fine-Tuning	0,973	0,968	0,968	0,967
Método Proposto	**0,987**	**0,987**	0,985	**0,986**

A matriz de confusão apresentada na Figura 3 permite analisar os padrões de erro do modelo. Observa-se elevada taxa de acerto para todas as classes, com destaque para HBS, classificada corretamente em todos os casos. Os erros concentram-se principalmente na distinção entre ALL e AML, indicando maior similaridade morfológica entre essas categorias.

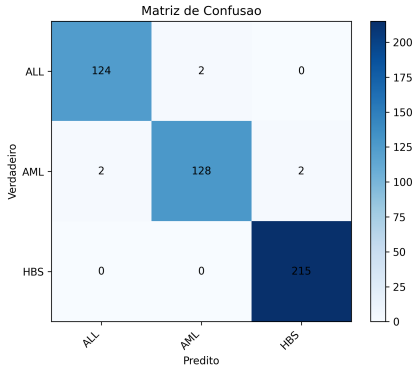


Figura 3. Matriz de confusão do melhor modelo

Para avaliar a coerência morfológica das decisões do modelo, aplicou-se a técnica Grad-CAM++ sobre o backbone convolucional. Na Figura 4 observa-se que as ativações concentram-se predominantemente na região nuclear e citoplasmática das células, indicando que o modelo baseia suas decisões em características biologicamente plausíveis, e não em artefatos de fundo ou padrões de coloração.

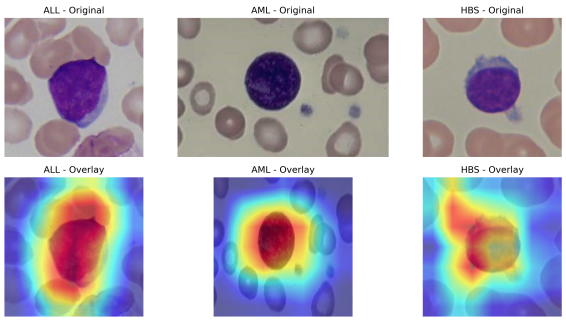


Figura 4. Mapas de ativação para amostras representativas das classes ALL, AML e HBS.

4.3. Análise Estatística Comparativa

Com o objetivo de verificar se as melhorias de desempenho do modelo proposto são estatisticamente significativas e representam uma mudança real na natureza das classificações, foi conduzida análise estatística formal utilizando as predições geradas no conjunto de teste independente.

Diferentemente da avaliação de normalidade de métricas globais, a comparação direta das predições de classificação exige a avaliação pareada das matrizes de confusão. Por isso, optou-se pela aplicação do teste multiclasse de McNemar-Bowker para a comparação entre as predições do modelo proposto e dos principais baselines implementados sob o mesmo protocolo experimental.

A hipótese nula (H_0) postula a simetria marginal das classificações, assumindo que as discordâncias entre os modelos ocorrem nas mesmas proporções, enquanto a hipótese alternativa (H_1) indica uma assimetria estatisticamente significativa nesses erros.

Os resultados indicaram que o modelo CNN-Mamba (F1-*score* = 0,986) apresentou classificação estatisticamente superior à ResNet50 congelada (p = 0,017 $<$ 0,05), evidenciando a capacidade dos módulos híbridos em capturar dependências globais. Contudo, quando comparado à ResNet50 com *fine-tuning* parcial, o teste indicou simetria nas discordâncias (p = 0,39 $>$ 0,05). Embora o modelo proposto apresente maior precisão absoluta, seus erros espelham os do *baseline* treinado. Teoricamente, isso ocorre porque a configuração rasa do módulo sequencial (2 blocos Mamba) atua como refinadora das características convolucionais, otimizando o desempenho sem forçar representações divergentes que causariam *overfitting*. Assim, a simetria residual reflete a forte ambiguidade biológica inerente entre as classes ALL e AML.

Além disso, foram calculados intervalos de confiança de 95% para o F1-*score* médio(0,9855; IC95%:0,9731–0,9954), reforçando a consistência estatística dos resulta-

dos observados. A diferença média absoluta de desempenho foi de ΔF1 = 1,83%, indicando ganho prático relevante sob condições multicêntricas heterogêneas.

Esses resultados confirmam que o desempenho superior do modelo proposto não decorre de flutuações aleatórias do processo de treinamento, mas sim de melhorias estruturais introduzidas pela arquitetura híbrida baseada em *State Space Models*.

5. Contribuições do Trabalho

Este trabalho contribui para o avanço da hematologia computacional ao propor uma arquitetura híbrida CNN–Mamba especificamente projetada para classificação multiclasse de leucemias agudas em cenários multicêntricos heterogêneos. A principal inovação reside na integração arquitetural *end-to-end* entre extração convolucional profunda de características morfológicas locais e modelagem sequencial baseada em Visual *State Space Blocks*, permitindo captura eficiente de dependências estruturais globais com complexidade linear. Diferentemente de abordagens que utilizam classificadores externos ou módulos desacoplados, a estratégia proposta promove aprendizado conjunto otimizado, favorecendo maior coesão representacional e estabilidade preditiva.

Além da proposição metodológica, o estudo realiza validação sistemática em um macro-*dataset* composto por 17 bases públicas independentes, incorporando variações substanciais de coloração, iluminação, resolução e protocolos laboratoriais. Essa configuração experimental aproxima o modelo de condições clínicas reais, superando a limitação recorrente de avaliações restritas a bases homogêneas e ampliando evidências sobre capacidade de generalização sob deslocamento de domínio interlaboratorial.

O trabalho também apresenta análise quantitativa estruturada do impacto de fatores arquiteturais e amostrais críticos, investigando de forma controlada os efeitos da resolução espacial, do volume de dados de treinamento e da profundidade do módulo sequencial no desempenho preditivo. Tal investigação fornece evidências empíricas sobre estabilidade, saturação informacional em altas resoluções e sensibilidade da profundidade arquitetural, aspectos ainda pouco explorados de maneira sistemática na literatura de classificação hematológica.

Adicionalmente, o modelo proposto obteve melhor F1-*score* macro em comparação aos baselines avaliados, com suporte de testes estatísticos, indicando potencial para cenários com desbalanceamento e múltiplas fontes de dados. Por fim, a análise de interpretabilidade baseada em Grad-CAM++ indica que as decisões do modelo concentram-se predominantemente em regiões nuclear e citoplasmática das células, sugerindo alinhamento entre os padrões aprendidos e critérios morfológicos clinicamente plausíveis, fortalecendo a confiabilidade da abordagem como ferramenta de apoio ao diagnóstico.

6. Conclusão

Este trabalho apresentou uma arquitetura híbrida CNN–Mamba para classificação multiclasse de leucemias agudas em imagens hematológicas multicêntricas, abordando explicitamente o desafio da variabilidade interlaboratorial. Diferentemente de abordagens puramente convolucionais, o modelo proposto integra extração local de características morfológicas com modelagem sequencial baseada em *State Space Models*, permitindo captura eficiente de dependências estruturais globais com complexidade linear.

A validação experimental foi conduzida sob protocolo sistemático, incluindo validação cruzada estratificada, conjunto de teste independente, ponderação de classes para mitigação de desbalanceamento e avaliação sistemática em dezoito cenários controlados. Os resultados demonstraram que o modelo alcança desempenho superior (F1-*score* macro de $98,6\%$) mantendo estabilidade mesmo sob redução de dados e variações de resolução, evidenciando forte capacidade de generalização.

A análise comparativa com modelos baseline e a avaliação estatística formal confirmaram que os ganhos observados não são decorrentes de variações aleatórias do treinamento, mas sim da arquitetura proposta. Além disso, a robustez frente à heterogeneidade multicêntrica reforça o potencial de aplicabilidade prática em ambientes clínicos reais, onde protocolos de aquisição e coloração variam significativamente.

Como perspectivas futuras, pretende-se explorar estratégias de adaptação de domínio interlaboratorial, validação externa em bases adicionais independentes e investigação de técnicas de interpretabilidade para suporte à tomada de decisão médica. Os resultados obtidos indicam que arquiteturas híbridas baseadas em CNN e *State Space Models* representam uma alternativa promissora para o desenvolvimento de sistemas confiáveis e escaláveis de apoio ao diagnóstico hematológico assistido por inteligência artificial.

Agradecimentos

Os autores agradecem ao Instituto Federal do Piauí (IFPI) pelo apoio ao desenvolvimento deste trabalho.

Referências

Chen, C.-S., Zhou, D., Chen, G.-Y., Jiang, D., and Chen, D.-S. (2024). Res-vmamba: Fine-grained food category visual classification using selective state space models with deep residual learning. *arXiv preprint arXiv:2402.15761*.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Gu, A. and Dao, T. (2024). Mamba: Linear-time sequence modeling with selective state spaces. In *First conference on language modeling*.

Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Hidayat, T., Hadinata, E., Damanik, I. S., Vikki, Z., and Irvanizam, I. (2023). An implementation of hybrid cnn-xgboost method for leukemia detection problem. *Infolitika Journal of Data Science*, 1(1):15–21.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.

Kuang, Z., Yan, X., Yu, J., Sun, D., Zhao, J., and Sun, L. (2026). Lbmnet: a hybrid multi-scale cnn–mamba framework for enhanced 3d stroke lesion segmentation in mri. *Frontiers in Medicine*, 13:1759114.

Labati, R. D., Piuri, V., and Scotti, F. (2011). All-idb: The acute lymphoblastic leukemia image database for image processing. In *2011 18th IEEE international conference on image processing*, pages 2045–2048. IEEE.

Li, M., Jiang, Y., Zhang, Y., and Zhu, H. (2023). Medical image analysis using deep learning algorithms. *Frontiers in public health*, 11:1273253.

Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., Jiao, J., and Liu, Y. (2024). Vmamba: Visual state space model. *Advances in neural information processing systems*, 37:103031–103063.

Ma, C. and Wang, Z. (2024). Semi-mamba-unet: Pixel-level contrastive and cross-supervised visual mamba-based unet for semi-supervised medical image segmentation. *Knowledge-Based Systems*, 300:112203.

Metz, C. E. (1978). Basic principles of roc analysis. In *Seminars in nuclear medicine*, volume 8, pages 283–298. Elsevier.

Powers, D. M. (2020). Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*.

Saito, T. and Rehmsmeier, M. (2015). The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, 10(3):e0118432.

Siegel, R. L., Miller, K. D., Wagle, N. S., and Jemal, A. (2023). Cancer statistics, 2023. *CA: a cancer journal for clinicians*, 73(1):17–48.

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J., and Kattan, M. W. (2010). Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*, 21(1):128–138.

Vogado, L. H., Veras, R. M., and Aires, K. R. (2020). "leuknet-um modelo de rede neural convolucional para o diagnóstico de leucemia. *Anais do ... Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*.

Yu, W. and Wang, X. (2024). Mambaout: Do we really need mamba for vision? *arXiv preprint arXiv:2405.07992*.

Yue, Y. and Li, Z. (2024). Medmamba: Vision mamba for medical image classification. *arXiv preprint arXiv:2403.03849*.

Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., and Wang, X. (2024). Vision mamba: Efficient visual representation learning with bidirectional state space model. *arXiv preprint arXiv:2401.09417*.