

CLASSIFICAÇÃO DE CÉLULAS SANGUÍNEAS COM CNN PARA DETECÇÃO DE LEUCEMIA LINFOBLÁSTICA AGUDA UTILIZANDO O CONJUNTO DE DADOS ALL-IDB2

Rafael Augusto Hentz¹, Dr. Felipe André Zeiser², Dr. Tiago Zonta³

1. Discente do curso de graduação em Sistemas de Informação, Unoesc, Chapecó, SC
2. Software Innovation Laboratory - SOFTWARELAB, Universidade do Vale do Rio dos Sinos – UNISINOS, São Leopoldo, RS
3. Docente do curso de graduação em Sistemas de Informação, Unoesc, Chapecó, SC

Autor correspondente: Rafael Augusto Hentz, rafael.hentz@unoesc.edu.br

Área: Ciências Exatas e Tecnológicas

Introdução: A Leucemia Linfoblástica Aguda (LLA) é um dos tipos mais comuns de câncer infantil e caracteriza-se pela produção anormal de linfócitos na medula óssea, comprometendo a formação de células sanguíneas saudáveis. O diagnóstico precoce é determinante para o sucesso do tratamento, mas os métodos convencionais ainda dependem de análises laboratoriais demoradas, onerosas e sujeitas a variações entre especialistas. Nesse contexto, métodos computacionais têm se mostrado promissores no apoio ao diagnóstico, destacando-se o uso de técnicas de visão computacional e aprendizado profundo para análise de imagens médicas. **Objetivo:** Auxiliar o processo de diagnóstico da leucemia com uma aplicação de classificação de imagens de células sanguíneas baseada em deep learning. **Método:** O estudo utilizou o conjunto ALL-IDB2, composto por 260 imagens de células já recortadas e centralizadas, divididas em subconjuntos de treino (70%), validação (15%) e teste (15%). Inicialmente, avaliou-se a aplicação do modelo YOLOv11 para detecção, mas os resultados não foram satisfatórios devido às características do dataset. Em seguida, foi projetada uma CNN simples implementada em PyTorch, contendo camadas convolucionais, pooling, dropout e camadas totalmente conectadas. Para mitigar o overfitting, foram aplicadas técnicas de aumento de dados, como rotações, espelhamentos e ajustes de normalização. O treinamento foi conduzido em GPU NVIDIA T4 por 40 epochs, utilizando o otimizador Adam e a função de perda binária. O desempenho foi avaliado com métricas padrão: acurácia, precisão, recall e F1-score. **Resultados:** A CNN proposta obteve desempenho positivo no conjunto de teste, alcançando 87,5% de acurácia, 82,6% de precisão, 95% de recall e F1-score de 88,3%. Os resultados evidenciam que o modelo conseguiu generalizar de forma consistente, apresentando convergência estável no treinamento e validação. A matriz de confusão mostrou que a maioria dos erros correspondeu a falsos positivos, enquanto o alto recall reduziu a ocorrência de falsos negativos, aspecto essencial em diagnósticos clínicos. **Conclusão:** O estudo demonstrou que uma CNN leve e bem projetada pode atingir resultados competitivos na classificação de células sanguíneas com LLA mesmo em um dataset reduzido. A combinação de técnicas de pré-processamento, aumento de dados e regularização foi fundamental para a robustez do modelo. Além disso, a eficiência computacional obtida reforça o potencial de aplicação em ambientes clínicos com recursos limitados, contribuindo para o avanço de ferramentas acessíveis de apoio ao diagnóstico hematológico.

Palavras-chave: Visão computacional; Classificação de imagens; Leucemia Linfoblástica Aguda; Redes Neurais Convolucionais; Aprendizado profundo.

Agradecimentos: O autor Rafael A. Hentz agradece ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão de bolsa de iniciação científica (PIBIC).