UNIVERSITAS
GADJAH MADA
Penanganan Data Tidak Seimbang dengan CBAM dan NCR untuk Klasifikasi Penyakit Leukemia
Charma Novianti Hastuti, Wahyono, S.Kom., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

PENANGANAN DATA TIDAK SEIMBANG DENGAN CBAM DAN NCR UNTUK KLASIFIKASI PENYAKIT LEUKEMIA

Oleh

Charma Novianti Hastuti

24/536311/PPA/06774

Pemrosesan citra digital untuk klasifikasi penyakit medis telah berkembang pesat. Namun, ketidakseimbangan kelas dan data yang terbatas menjadi tantangan utama yang cenderung mengurangi generalisasi model dan pengenalan data minoritas. Studi ini mengusulkan kerangka kerja bebas augmentasi untuk klasifikasi leukemia yang mengatasi ketidakseimbangan kelas tanpa bergantung pada pembuatan data sintetis. Studi ini memperkenalkan kerangka kerja bebas augmentasi untuk klasifikasi leukemia yang mengurangi ketidakseimbangan tanpa menghasilkan data sintetis. Metode ini menggunakan fusi fitur *multi-stage* dengan *Concatenated Residual-Stage* ResNet-50 yang terintegrasi dengan CBAM untuk memperkuat representasi fitur diskriminatif. NCR diterapkan dalam ruang fitur untuk memperbaiki distribusi data sambil mempertahankan sampel minoritas yang informatif. Pra-pemrosesan optimal dilakukan untuk meningkatkan kualitas gambar serta klasifikasi dilakukan menggunakan SVM dalam *pipeline* pembelajaran *hybrid deep learning-machine learning.* Evaluasi dilakukan pada tiga dataset dengan ketidakseimbangan yang bervariasi.

Hasil penelitian menunjukan bahwa model mencapai kinerja maksimum ketika pada ketidakseimbangan ringan dan sedang serta akurasi 95,40% pada ketidakseimbangan ekstrem. Analisis lebih lanjut menunjukkan bahwa morfologis sel tiap kelas yang serupa menyebabkan representasi fitur yang tumpang tindih, membatasi manfaat penyempurnaan berbasis *attention*. Secara keseluruhan, hasil penelitian menunjukkan bahwa menggabungkan peningkatan fitur dengan pembersihan data selektif memungkinkan klasifikasi yang andal tanpa augmentasi sintetis, sehingga menawarkan solusi praktis dan relevan secara klinis untuk klasifikasi citra medis dalam skenario dunia nyata dengan dataset yang terbatas dan tidak seimbang.

Kata Kunci: klasifikasi leukemia, ketidakseimbangan data, *hybrid representation learning*, CBAM, *Neighborhood Cleaning Rule*

UNIVERSITAS
GADJAH MADA

Penanganan Data Tidak Seimbang dengan CBAM dan NCR untuk Klasifikasi Penyakit Leukemia
Charma Novianti Hastuti, Wahyono, S.Kom., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

HANDLING IMBALANCED DATA WITH CBAM AND NCR FOR LEUKEMIA DISEASE CLASSIFICATION

By

Charma Novianti Hastuti

24/536311/PPA/06774

Digital image processing for medical disease classification has advanced rapidly. However, class imbalance and limited data remain key challenges that reduce model generalization and recognition of minority data. This study proposes an augmentation-free framework for leukemia classification that addresses class imbalance without relying on synthetic data generation. This study introduces an augmentation-free framework for leukemia classification that mitigates imbalance without generating synthetic data. The method employs multi-stage feature fusion with a Concatenated Residual-Stage ResNet-50 integrated with CBAM to strengthen discriminative representation. NCR is applied in the deep feature space to refine data distribution while preserving informative minority samples. Optimal preprocessing improves image quality, and classification is performed using an SVM within a hybrid deep learning–machine learning pipeline. Evaluation across three datasets with varying imbalances shows that the framework achieves perfect performance under mild and moderate imbalance and 95.40% accuracy under extreme imbalance. Further analysis indicates that similar morphological classes cause overlapping feature representation, limiting the benefits of attention-based refinement. Overall, the results demonstrate that combining feature enhancement with selective data cleaning enables reliable classification without synthetic augmentation, offering a practical and clinically relevant solution for medical image classification in real-world scenarios with limited and imbalanced datasets.

Keywords: *Leukemia classification, class imbalance, hybrid representation learning, CBAM, Neighborhood Cleaning Rule*