Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 560−569
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9831

Fusi Cross-Attention CNN–Transformer untuk Klasifikasi Multi-Kelas Acute Lymphoblastic Leukemia

Wahyuni Fithratul Zalmi[1,*], Rahmi Putri Kurnia[2], Yulia Jihan Sy[2]

[1] Fakultas Teknik, Program Studi Teknik Informatika, Universitas Sam Ratulangi, Manado, Indonesia
[2] Teknologi Informasi, Politeknik Negeri Padang, Padang, Indonesia
Email: [1,*]wahyuni.fithratul.zalmi@unsrat.ac.id [2]rahmiputri@pnp.ac.id [2]yulia@pnp.ac.id
Email Penulis Korespondensi: wahyuni.fithratul.zalmi@unsrat.ac.id
Submitted: **03/05/2026**; Accepted: **02/06/2026**; Published**: 30/06/2026**

Abstrak−Acute Lymphoblastic Leukemia (ALL) merupakan keganasan hematologi yang memerlukan pemeriksaan awal secara cepat dan akurat. Citra Peripheral Blood Smear (PBS) dapat digunakan sebagai sumber informasi morfologi sel, namun klasifikasi berbasis citra masih menghadapi tantangan akibat variasi bentuk, warna, dan struktur sel darah. Penelitian ini mengusulkan model klasifikasi multi-kelas ALL berbasis fusi cross-attention CNN–Transformer dengan dua input citra, yaitu citra PBS original dan citra hasil segmentasi yang telah tersedia pada dataset. Kontribusi utama penelitian ini terletak pada integrasi fitur lokal dari CNN dan fitur global dari Transformer melalui mekanisme cross-attention, serta evaluasi komponen model melalui baseline comparison dan ablation study. Dataset yang digunakan terdiri atas 3.256 pasangan citra PBS pada empat kelas, yaitu Benign, Early, Pre, dan Pro. Hasil pengujian menunjukkan bahwa model memperoleh accuracy sebesar 0.9980 dan macro F1-score sebesar 0.9975 pada data uji. Meskipun demikian, performa yang sangat tinggi ini perlu diinterpretasikan secara hati-hati karena penelitian belum melibatkan validasi eksternal berbasis institusi berbeda maupun penilaian langsung oleh ahli patologi. Oleh karena itu, model yang diusulkan lebih tepat diposisikan sebagai pendekatan komputasional potensial untuk klasifikasi citra PBS, bukan sebagai sistem diagnosis klinis final. Evaluasi lanjutan pada dataset eksternal, skema pembagian berbasis pasien, serta validasi pakar diperlukan untuk menilai generalisasi dan relevansi klinis model.

Kata Kunci: Acute Lymphoblastic Leukemia; Peripheral Blood Smear; Deep Learning; Vision Transformer; Cross-Attention

Abstract−Acute Lymphoblastic Leukemia (ALL) is a hematological malignancy that requires a quick and accurate initial examination. Peripheral Blood Smear (PBS) imaging can be used as a source of cell morphology information, but image-based classification still faces challenges due to variations in the shape, color, and structure of blood cells. This study proposes an ALL multi-class classification model based on CNN–Transformer cross-attention fusion with two image inputs, namely the original PBS image and the segmented image that is already available in the dataset. The main contribution of this study lies in the integration of the local features of the CNN and the global features of the Transformer through the cross-attention mechanism, as well as the evaluation of the model components through baseline comparison and ablation studies. The dataset used consisted of 3,256 pairs of PBS images in four classes, namely Benign, Early, Pre, and Pro. The test results showed that the model obtained an accuracy of 0.9980 and a macro F1-score of 0.9975 on the test data. Nonetheless, this very high performance needs to be interpreted with caution as the research has not involved external validation based on different institutions or direct assessments by pathologists. Therefore, the proposed model is more appropriately positioned as a potential computational approach to the classification of PBS images, rather than as a final clinical diagnostic system. Advanced evaluation of external datasets, patient-based allocation schemes, and expert validation are required to assess the generalization and clinical relevance of the model.

Keywords: Acute Lymphoblastic Leukemia; Peripheral Blood Smear; Deep Learning; Vision Transformer; Cross-Attention

1. PENDAHULUAN

Acute Lymphoblastic Leukemia (ALL) merupakan keganasan hematologi agresif yang ditandai oleh proliferasi abnormal sel limfoblas pada sumsum tulang dan darah perifer [1], [2]. Deteksi dini ALL sangat penting karena keterlambatan diagnosis dapat memperburuk prognosis pasien dan meningkatkan kompleksitas terapi [3], [4]. Secara klinis, diagnosis definitif ALL umumnya memerlukan pemeriksaan seperti flow cytometry, aspirasi sumsum tulang, dan analisis molekuler yang bersifat invasif, mahal, membutuhkan waktu, serta bergantung pada ketersediaan tenaga ahli [5], [6]. Oleh karena itu, citra Peripheral Blood Smear (PBS) menjadi alternatif penting untuk skrining awal karena lebih mudah diperoleh, lebih cepat, dan dapat memberikan informasi morfologi sel darah yang relevan.

Namun, interpretasi citra PBS masih menghadapi tantangan besar. Kemiripan morfologi antara sel benign dan limfoblas malignan, variasi pewarnaan, perbedaan kualitas citra mikroskopis, serta subjektivitas pemeriksa dapat menyebabkan kesalahan klasifikasi. Tantangan ini semakin kompleks pada klasifikasi multi-kelas karena model tidak hanya harus membedakan kelas benign dan malignant, tetapi juga membedakan subtipe ALL seperti Early Pre-B, Pre-B, dan Pro-B yang memiliki perbedaan visual halus. Dengan demikian, diperlukan pendekatan komputasional yang mampu mengekstraksi fitur lokal dan global secara simultan serta memberikan interpretasi visual terhadap keputusan model.

Meskipun pendekatan CNN telah menunjukkan hasil yang menjanjikan dalam deteksi dan klasifikasi leukemia berbasis citra, masih terdapat keterbatasan dalam menangkap hubungan global antarbagian sel, terutama pada kasus klasifikasi subtipe ALL yang memiliki perbedaan morfologi halus [7], [8], [9]. Di sisi lain, Transformer memiliki kemampuan untuk memodelkan dependensi spasial dan hubungan global secara lebih luas, tetapi penggunaannya secara tunggal pada dataset medis berukuran terbatas masih menghadapi tantangan generalisasi, kebutuhan data besar, dan kompleksitas komputasi [10], [11]. Oleh karena itu, diperlukan pendekatan hybrid yang mampu menggabungkan keunggulan CNN dalam mengekstraksi fitur lokal dan Transformer dalam menangkap konteks global secara adaptif.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 560−569
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9831

Perkembangan terbaru menunjukkan mulai digunakannya Vision Transformer (ViT) dan mekanisme attention dalam diagnosis leukemia [12]. Model seperti VisTA menggabungkan Vision Transformer dan CNN berbasis attention untuk meningkatkan deteksi leukemia pada citra darah beresolusi tinggi [13]. Seiring berkembangnya pendekatan berbasis attention, beberapa penelitian mulai mengombinasikan CNN dan Transformer untuk meningkatkan kemampuan representasi fitur pada citra medis [14]. CNN berperan dalam menangkap fitur lokal, sedangkan Transformer mampu memodelkan hubungan global antarbagian citra [15]. Selain itu, aspek interpretabilitas juga mulai menjadi perhatian penting dalam pengembangan sistem diagnosis berbasis deep learning, karena model medis tidak hanya dituntut memiliki akurasi tinggi, tetapi juga mampu memberikan penjelasan terhadap area citra yang berkontribusi terhadap keputusan prediksi [16].

Meskipun berbagai pendekatan tersebut menunjukkan kemajuan signifikan, masih terdapat beberapa gap penelitian. Pertama, sebagian besar studi masih menggunakan arsitektur single-stream sehingga belum memanfaatkan hubungan komplementer antara citra asli dan citra hasil segmentasi. Kedua, pendekatan CNN kuat dalam mengekstraksi fitur lokal seperti tekstur, warna, dan batas sel, tetapi kurang optimal dalam menangkap hubungan global. Ketiga, Transformer mampu menangkap dependensi global, tetapi sering kali membutuhkan integrasi yang lebih kuat dengan fitur lokal agar efektif pada dataset medis berukuran terbatas. Keempat, banyak penelitian hanya menggunakan feature concatenation sederhana tanpa mekanisme interaksi adaptif antarfitur. Kelima, masih terbatas penelitian yang menggabungkan dual-stream input, DenseNet121, Transformer, cross-attention, ablation study, validasi statistik, dan Explainable AI dalam satu kerangka evaluasi yang komprehensif.

Meskipun berbagai arsitektur CNN modern seperti ConvNeXt dan EfficientNetV2 menunjukkan performa yang sangat baik pada berbagai tugas klasifikasi citra, penelitian ini memilih DenseNet121 sebagai backbone utama karena karakteristiknya yang lebih sesuai untuk dataset medis berukuran relatif terbatas. DenseNet121 memiliki mekanisme *dense connectivity* yang memungkinkan setiap layer menerima informasi dari layer sebelumnya secara langsung, sehingga mendukung *feature reuse*, memperkuat propagasi gradien, dan mengurangi jumlah parameter dibandingkan arsitektur yang lebih kompleks. Karakteristik ini penting pada klasifikasi citra Peripheral Blood Smear (PBS), karena fitur morfologi seperti tekstur nukleus, batas sel, dan distribusi sitoplasma sering kali bersifat halus dan memerlukan representasi lokal yang stabil. Selain itu, penggunaan DenseNet121 berbasis *transfer learning* juga telah banyak dilaporkan efektif pada berbagai tugas klasifikasi citra medis dengan ukuran dataset terbatas. Di sisi lain, penelitian ini tidak hanya berfokus pada pemilihan backbone, tetapi juga pada mekanisme integrasi fitur antara CNN dan Transformer.

Pendekatan *feature fusion* sederhana seperti concatenation memang mampu menggabungkan fitur dari dua representasi berbeda, namun metode tersebut memperlakukan seluruh fitur secara statis tanpa mempertimbangkan hubungan kontekstual antarrepresentasi. Dalam konteks morfologi sel darah, hubungan antara fitur lokal dan global bersifat kompleks, karena perubahan kecil pada struktur nukleus atau distribusi kromatin dapat memiliki keterkaitan dengan pola morfologi sel secara keseluruhan. Oleh karena itu, digunakan mekanisme *cross-attention* untuk memungkinkan fitur CNN dan Transformer saling berinteraksi secara adaptif. Melalui mekanisme ini, model dapat mempelajari bagian fitur global yang paling relevan terhadap fitur lokal tertentu, sehingga integrasi representasi menjadi lebih selektif dibandingkan penggabungan fitur konvensional. Pendekatan ini diharapkan mampu meningkatkan kemampuan model dalam membedakan subtipe ALL yang memiliki perbedaan visual halus pada citra PBS.

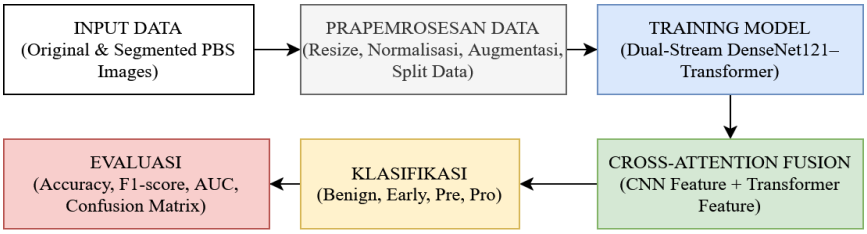
Berdasarkan gap tersebut, penelitian ini mengusulkan arsitektur Dual-Stream DenseNet121–Vision Transformer berbasis Cross-Attention untuk klasifikasi multi-kelas ALL pada citra PBS. DenseNet121 digunakan untuk mengekstraksi fitur lokal dari citra asli, sedangkan Vision Transformer digunakan untuk mengekstraksi fitur global dari citra hasil segmentasi. Mekanisme cross-attention diterapkan untuk memungkinkan interaksi adaptif antara representasi CNN dan Transformer, sehingga model dapat memadukan informasi morfologi lokal dan konteks global secara lebih efektif. Kontribusi utama penelitian ini adalah: pertama, mengembangkan arsitektur dual-stream yang menggabungkan citra original dan segmented untuk klasifikasi multi-kelas ALL. Kedua, mengintegrasikan DenseNet121 dan Vision Transformer melalui mekanisme cross-attention, bukan hanya concatenation sederhana. Ketiga, melakukan baseline comparison dan ablation study untuk mengukur kontribusi setiap komponen model. Keempat, melakukan validasi statistik menggunakan repeated experiment, 5-fold cross-validation, paired t-test, Wilcoxon test, dan McNemar test untuk mengevaluasi konsistensi serta signifikansi performa model. Secara keseluruhan, penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan sistem diagnosis berbantuan komputer yang lebih akurat, interpretatif, dan relevan secara klinis untuk skrining awal Acute Lymphoblastic Leukemia berbasis citra Peripheral Blood Smear.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian ini disusun untuk menggambarkan alur kerja secara menyeluruh, mulai dari input data, prapemrosesan, pelatihan model, integrasi fitur, klasifikasi, hingga evaluasi performa. Alur tahapan penelitian secara ringkas ditunjukkan pada Gambar 1.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 560−569
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9831

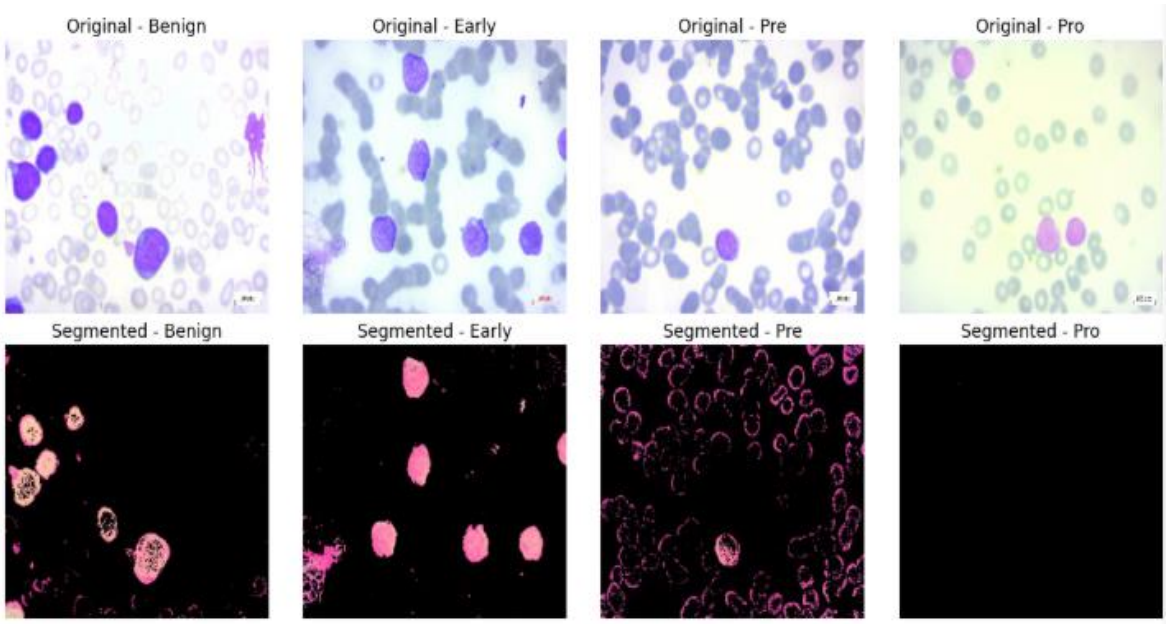


Gambar 1. Alur tahapan penelitian

Gambar 1 menunjukkan alur penelitian yang dimulai dari input data berupa citra PBS original dan segmented. Data kemudian melalui tahap prapemrosesan, yaitu resize, normalisasi, augmentasi, dan pembagian dataset. Selanjutnya, data dilatih menggunakan model Dual-Stream DenseNet121–Transformer, di mana fitur CNN dan Transformer digabungkan melalui mekanisme cross-attention fusion. Hasil fusi fitur digunakan untuk klasifikasi empat kelas, yaitu Benign, Early, Pre, dan Pro. Pada tahap akhir, performa model dievaluasi menggunakan accuracy, F1-score, AUC, dan confusion matrix.

2.2 Input Data

Input data yang digunakan dalam penelitian ini berupa citra Peripheral Blood Smear (PBS) Acute Lymphoblastic Leukemia (ALL) yang diperoleh dari bone marrow laboratory Taleqani Hospital, Tehran, Iran. Dataset terdiri dari dua jenis citra, yaitu Original dan Segmented, dengan total 3.256 pasangan citra valid yang terbagi ke dalam empat kelas: Benign sebanyak 504 citra, Early sebanyak 985 citra, Pre sebanyak 963 citra, dan Pro sebanyak 804 citra. Setiap citra original memiliki pasangan segmented yang sesuai, sehingga dataset dapat digunakan sebagai input pada arsitektur dual-stream DenseNet121–Transformer. Setelah proses pengecekan dataset dilakukan, citra dari setiap kelas ditampilkan untuk memberikan gambaran visual terhadap karakteristik data yang digunakan. Visualisasi ini mencakup citra Original dan Segmented dari masing-masing kelas, sehingga perbedaan morfologi sel pada kelas Benign, Early, Pre, dan Pro dapat diamati secara lebih jelas. Contoh tampilan citra dataset ditunjukkan pada Gambar 2.



Gambar 2. Visualisasi Citra PBS Original dan Segmentasi

Gambar 2 menampilkan contoh citra Peripheral Blood Smear (PBS) pada setiap kelas, yang terdiri dari citra original (baris atas) dan citra hasil segmentasi (baris bawah). Terlihat bahwa citra original menunjukkan variasi morfologi sel darah secara keseluruhan, sedangkan citra segmented menyoroti bagian penting seperti area sel dan nukleus. Perbedaan karakteristik visual antar kelas, khususnya pada ukuran, bentuk, dan distribusi sel, menjadi dasar bagi model dalam melakukan proses klasifikasi.

2.3 Prapemrosesan Data

Prapemrosesan data dilakukan untuk memastikan seluruh citra siap digunakan sebagai input model dual-stream. Pada tahap awal, seluruh file pada folder Original dan Segmented diperiksa berdasarkan validitas file, format, mode warna, dan ukuran citra. Hasil pemeriksaan menunjukkan bahwa terdapat 6.512 file citra valid yang terdiri dari 3.256 citra original dan 3.256 citra segmented, tanpa ditemukan file rusak. Seluruh citra memiliki format JPEG, mode warna RGB, dan ukuran paling sering 224 × 224 piksel. Selain itu, proses sinkronisasi dilakukan untuk memastikan setiap

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 560−569
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9831

citra original memiliki pasangan segmented yang sesuai. Hasilnya, diperoleh 3.256 pasangan data valid dengan distribusi kelas yang seimbang antara original dan segmented, yaitu Benign 504, Early 985, Pre 963, dan Pro 804.

Setelah data dinyatakan valid, citra diproses menggunakan transformasi standar sebelum masuk ke model. Seluruh citra disiapkan pada ukuran input 224 × 224 piksel dan dinormalisasi menggunakan standar ImageNet dengan nilai mean [0.485, 0.456, 0.406] dan standard deviation [0.229, 0.224, 0.225]. Dataset kemudian dibagi menggunakan metode stratified split menjadi training set sebanyak 2.279 data, validation set sebanyak 488 data, dan test set sebanyak 489 data. Pada data training diterapkan augmentasi berupa horizontal flip, vertical flip, rotasi, penyesuaian brightness, contrast, saturation, dan normalisasi ImageNet, sedangkan data validation dan test hanya melalui resize dan normalisasi tanpa augmentasi acak. Data kemudian dimuat menggunakan data loader dengan ukuran batch 16, menghasilkan 143 batch training, 31 batch validation, dan 31 batch test, dengan bentuk input torch.Size([16, 3, 224, 224]) untuk masing-masing citra original dan segmented.

2.4 Training Model

2.4.1 Arsitektur Dual-Stream CNN–Transformer

Model yang diusulkan menggunakan pendekatan *dual-stream architecture* yang memproses dua jenis citra secara paralel, yaitu citra original dan citra segmented [17]. Pendekatan ini bertujuan untuk memanfaatkan informasi komplementer, di mana citra original mengandung detail visual lengkap, sedangkan citra segmented menyoroti struktur penting sel. Arsitektur ini terdiri dari dua jalur utama, yaitu CNN berbasis DenseNet121 dan Transformer berbasis Vision Transformer, yang kemudian diintegrasikan melalui mekanisme cross-attention.

2.4.2 DenseNet121 untuk Ekstraksi Fitur Lokal

DenseNet121 digunakan sebagai *backbone* CNN untuk mengekstraksi fitur lokal dari citra original. Arsitektur DenseNet memiliki keunggulan dalam *feature reuse* melalui koneksi padat (*dense connectivity*), sehingga mampu menangkap informasi tekstur, warna sitoplasma, dan bentuk nukleus secara efektif. Model ini diinisialisasi menggunakan bobot pretrained ImageNet, dan layer klasifikasi akhir dihapus untuk menghasilkan representasi fitur berdimensi tinggi sebesar $F_{cnn} \in \mathbb{R}^{1024}$ [18].

2.4.3 Vision Transformer untuk Ekstraksi Fitur Global

Vision Transformer (ViT) digunakan untuk mengekstraksi fitur global dari citra segmented. Citra dibagi menjadi beberapa patch berukuran tetap, kemudian diproyeksikan menjadi embedding vektor yang diproses menggunakan mekanisme self-attention. Transformer memungkinkan model untuk menangkap hubungan spasial antar bagian sel dan pola morfologi global yang tidak dapat ditangkap secara optimal oleh CNN. Output dari Transformer direpresentasikan sebagai $F_{trans} \in \mathbb{R}^{D}$, di mana Dmerupakan dimensi embedding [19].

2.4.4 Penyamaan Dimensi Fitur (Feature Projection)

Karena output DenseNet121 dan Transformer memiliki dimensi yang berbeda, dilakukan proses penyamaan dimensi menggunakan *fully connected projection layer*. Kedua fitur diproyeksikan ke dimensi yang sama [20], yaitu:

$$F'_{cnn} = W_c F_{cnn}, F'_{trans} = W_t F_{trans} \tag{1}$$

dengan $F'_{cnn}, F'_{trans} \in \mathbb{R}^{512}$. Proses ini bertujuan untuk mempermudah integrasi fitur pada tahap selanjutnya.

2.4.5 Mekanisme Cross-Attention

Integrasi fitur dilakukan menggunakan mekanisme *cross-attention*, di mana fitur dari CNN berperan sebagai *query* (Q), sedangkan fitur dari Transformer berperan sebagai *key* (K) dan *value* (V) [21]:

$$Q = F'_{cnn}, K = F'_{trans}, V = F'_{trans} \tag{2}$$

Output cross-attention dihitung sebagai:

$$F_{cross} = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{3}$$

Mekanisme ini memungkinkan model untuk mempelajari hubungan antara fitur lokal dan global secara adaptif, sehingga meningkatkan kualitas representasi fitur.

2.4.6 Feature Fusion dan Klasifikasi

Fitur dari CNN, Transformer, dan cross-attention kemudian digabungkan menggunakan teknik *concatenation* [20]:

$$F_{final} = [F'_{cnn}, F'_{trans}, F_{cross}] \tag{4}$$

Fitur gabungan ini diproses melalui beberapa layer fully connected dengan tambahan *batch normalization* dan *dropout* untuk meningkatkan generalisasi model. Output akhir dihasilkan melalui layer klasifikasi dengan fungsi aktivasi softmax untuk menentukan kelas input ke dalam empat kategori, yaitu Benign, Early, Pre, dan Pro.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 560−569
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9831

2.4.7 Strategi Training dan Fine-Tuning

Model dilatih menggunakan optimizer AdamW dengan learning rate awal sebesar 1×10^{-4}dan scheduler *Cosine Annealing*. Untuk mencegah overfitting, digunakan teknik *early stopping* berdasarkan performa validation. Selain itu, dilakukan *fine-tuning* pada DenseNet121 dengan membekukan sebagian layer pada tahap awal dan membuka layer tertentu pada tahap lanjutan menggunakan learning rate yang lebih kecil. Fungsi loss yang digunakan adalah CrossEntropyLoss, serta dibandingkan dengan Focal Loss untuk mengatasi potensi ketidakseimbangan data.

2.5 Evaluasi Model

Evaluasi model dilakukan untuk mengukur performa klasifikasi multi-kelas Acute Lymphoblastic Leukemia (ALL) secara komprehensif. Model yang telah dilatih diuji menggunakan test set yang tidak terlibat dalam proses training maupun validation, sehingga hasil evaluasi mencerminkan kemampuan generalisasi model terhadap data baru. Beberapa metrik evaluasi yang digunakan dalam penelitian ini meliputi accuracy, precision, recall, dan F1-score, baik dalam bentuk *macro* maupun *weighted*, untuk mengakomodasi distribusi kelas yang tidak sepenuhnya seimbang. Accuracy digunakan untuk mengukur proporsi prediksi yang benar, sedangkan precision dan recall digunakan untuk mengevaluasi kemampuan model dalam mengidentifikasi kelas secara spesifik. F1-score digunakan sebagai metrik utama karena merupakan keseimbangan antara precision dan recall, terutama pada kasus multi-kelas.
Secara matematis, metrik evaluasi dapat dirumuskan sebagai berikut:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

$$\text{Precision} = \frac{TP}{TP+FP}, \text{Recall} = \frac{TP}{TP+FN} \tag{6}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision}+\text{Recall}} \tag{7}$$

Selain itu, digunakan Area Under Curve (AUC) berbasis *Receiver Operating Characteristic (ROC)* untuk mengukur kemampuan model dalam membedakan antar kelas. Evaluasi juga dilengkapi dengan confusion matrix untuk menganalisis distribusi kesalahan prediksi antar kelas, sehingga dapat diketahui kelas mana yang paling sering mengalami mis-klasifikasi. Untuk memperkuat validitas hasil, dilakukan baseline comparison dengan beberapa model pembanding, seperti CNN-only, Transformer-only, serta dual-stream tanpa cross-attention. Selain itu, dilakukan ablation study dengan menghilangkan komponen tertentu dalam model untuk mengukur kontribusi masing-masing bagian. Evaluasi juga dilengkapi dengan pendekatan Explainable AI (XAI) menggunakan Grad-CAM dan attention map untuk memvisualisasikan area fokus model. Sebagai tahap akhir, dilakukan validasi statistik menggunakan repeated experiment, 5-fold cross-validation, serta uji statistik seperti *paired t-test*, *Wilcoxon test*, dan *McNemar test* untuk memastikan bahwa peningkatan performa model signifikan secara statistik.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil eksperimen dari model yang diusulkan serta analisis terhadap performa yang diperoleh. Evaluasi dilakukan untuk mengukur kemampuan model dalam melakukan klasifikasi multi-kelas Acute Lymphoblastic Leukemia (ALL) berdasarkan citra Peripheral Blood Smear (PBS). Selain itu, dilakukan pembahasan terhadap hasil yang diperoleh melalui perbandingan dengan model baseline, analisis kontribusi komponen model, serta interpretasi menggunakan pendekatan Explainable AI untuk memahami pola keputusan model secara lebih mendalam.

3.1 Hasil

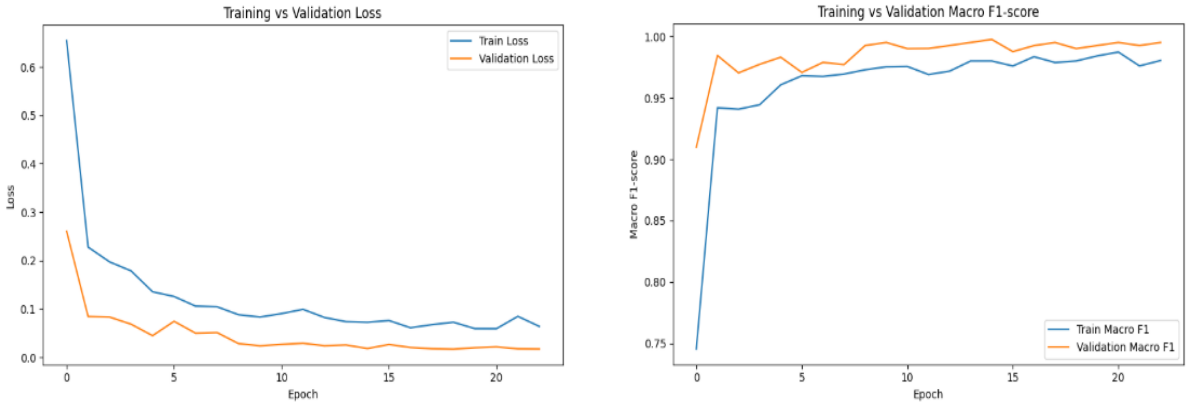
3.1.1 Hasil Pelatihan Model

Hasil pelatihan model disajikan untuk menunjukkan performa selama proses training dan fine-tuning. Pelatihan dilakukan dalam dua tahap, yaitu tahap pertama dengan membekukan (freeze) parameter DenseNet121 dan tahap kedua dengan melakukan fine-tuning pada layer tertentu. Berdasarkan hasil pelatihan, model menunjukkan peningkatan performa yang signifikan pada setiap epoch, yang ditandai dengan penurunan nilai loss serta peningkatan nilai F1-score. Pada tahap pertama, model mencapai performa terbaik pada epoch ke-10 dengan nilai training loss sebesar 0.0833, training F1-score sebesar 0.9752, validation loss sebesar 0.0238, dan validation F1-score sebesar 0.9950. Selanjutnya, pada tahap fine-tuning, performa model meningkat lebih lanjut dan mencapai hasil terbaik pada epoch ke-5 dengan nilai training loss sebesar 0.0724, training F1-score sebesar 0.9800, validation loss sebesar 0.0183, dan validation F1-score sebesar 0.9975. Ringkasan hasil pelatihan model ditunjukkan pada Tabel 2.

Tabel 2. Hasil Pelatihan Model

Stage	Epoch Terbaik	Train Loss	Train F1	Val Loss	Val F1
Frozen DenseNet121	10	0.0833	0.9752	0.0238	0.9950
Fine-tuning DenseNet121	5	0.0724	0.9800	0.0183	0.9975

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 560−569
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9831



Gambar 3. Kurva training dan validation loss serta macro F1-score selama proses pelatihan model.

Gambar 3 menunjukkan kurva perkembangan *training loss*, *validation loss*, serta *macro F1-score* selama proses pelatihan model. Terlihat bahwa nilai *training loss* dan *validation loss* mengalami penurunan secara konsisten seiring bertambahnya epoch, yang mengindikasikan proses pembelajaran berjalan dengan baik. Selain itu, nilai *macro F1-score* pada data training dan validation menunjukkan tren peningkatan dan stabil pada nilai tinggi, yang menunjukkan bahwa model mampu mencapai performa optimal tanpa mengalami overfitting yang signifikan

3.1.2 Hasil Evaluasi Model Final

Hasil evaluasi model dilakukan menggunakan *test set* untuk mengukur performa akhir model dalam melakukan klasifikasi multi-kelas Acute Lymphoblastic Leukemia (ALL). Evaluasi dilakukan menggunakan beberapa metrik, yaitu *accuracy*, *precision*, *recall*, *F1-score*, *specificity*, dan *Area Under Curve (AUC)*. Berdasarkan hasil pengujian, model yang diusulkan memperoleh nilai *accuracy* sebesar 0.9980, *precision macro* sebesar 0.9983, *recall macro* sebesar 0.9967, dan *F1-score macro* sebesar 0.9975. Selain itu, nilai *specificity macro* mencapai 0.9993, dan *AUC macro* sebesar 1.0000, yang menunjukkan kemampuan model dalam membedakan antar kelas secara sangat baik. Ringkasan hasil evaluasi model ditunjukkan pada Tabel 3.

Tabel 3. Hasil Evaluasi Model Final

Metrik	Nilai
Accuracy	0.9980
Precision Macro	0.9983
Recall Macro	0.9967
F1-score Macro	0.9975
F1-score Weighted	0.9980
Specificity Macro	0.9993
AUC Macro	1.0000

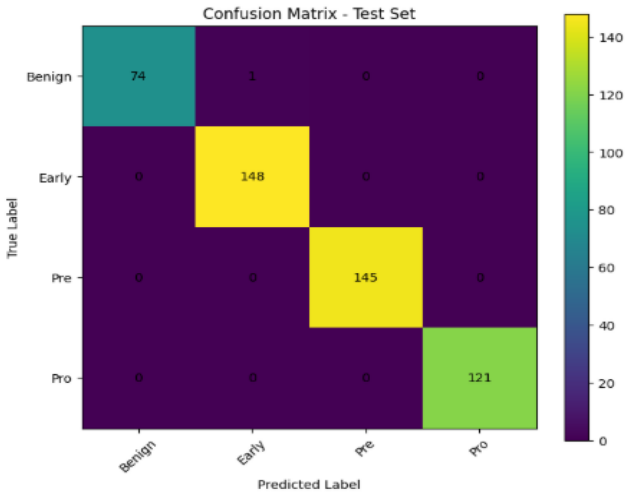
3.1.3 Classification Report

Classification report digunakan untuk mengevaluasi performa model pada masing-masing kelas, yaitu Benign, Early, Pre, dan Pro. Metrik yang digunakan meliputi *precision*, *recall*, *F1-score*, dan jumlah data (*support*) pada setiap kelas. Berdasarkan hasil pengujian, model menunjukkan performa yang sangat tinggi pada seluruh kelas. Kelas Pre dan Pro mencapai nilai *precision*, *recall*, dan *F1-score* sebesar 1.0000, yang menunjukkan bahwa seluruh data pada kelas tersebut berhasil diklasifikasikan dengan benar. Sementara itu, kelas Early memiliki nilai *precision* sebesar 0.9933, *recall* sebesar 1.0000, dan *F1-score* sebesar 0.9966. Pada kelas Benign, model memperoleh nilai *precision* sebesar 1.0000, *recall* sebesar 0.9867, dan *F1-score* sebesar 0.9933. Secara keseluruhan, model mencapai nilai *macro average F1-score* sebesar 0.9975 dan *weighted F1-score* sebesar 0.9980, yang menunjukkan bahwa model mampu melakukan klasifikasi secara konsisten pada seluruh kelas. Ringkasan classification report ditunjukkan pada Tabel 4.

Tabel 4. Classification Report

Kelas	Precision	Recall	F1-score	Support
Benign	1.0000	0.9867	0.9933	75
Early	0.9933	1.0000	0.9966	148
Pre	1.0000	1.0000	1.0000	145
Pro	1.0000	1.0000	1.0000	121
Macro Avg	0.9983	0.9967	0.9975	489
Weighted Avg	0.9980	0.9980	0.9980	489

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 560−569
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9831



Gambar 4. Confusion matrix hasil klasifikasi pada test set.

Gambar 4 menunjukkan confusion matrix hasil klasifikasi pada data uji. Terlihat bahwa sebagian besar prediksi berada pada diagonal utama, yang menunjukkan bahwa model mampu mengklasifikasikan data dengan sangat baik pada seluruh kelas. Hanya terdapat satu kesalahan klasifikasi pada kelas Benign yang diprediksi sebagai kelas Early, sementara kelas lainnya berhasil diklasifikasikan secara sempurna. Hal ini menunjukkan bahwa model memiliki tingkat akurasi yang sangat tinggi dan kesalahan klasifikasi yang minimal.

3.1.4 Perbandingan dengan Model Baseline

Perbandingan dengan model baseline dilakukan untuk mengevaluasi keunggulan model yang diusulkan dibandingkan dengan pendekatan lain. Beberapa model baseline yang digunakan dalam penelitian ini meliputi model berbasis citra tunggal (original image only dan segmented image only), model berbasis arsitektur tunggal (DenseNet121 only dan Transformer only), serta model dual-stream tanpa mekanisme cross-attention. Berdasarkan hasil pengujian, model yang diusulkan menunjukkan performa yang lebih unggul dibandingkan dengan sebagian besar model baseline. Model dual-stream dengan mekanisme cross-attention memperoleh nilai accuracy sebesar 0.9979, recall macro sebesar 0.9967, dan F1-score macro sebesar 0.9975, yang lebih tinggi dibandingkan dengan model tanpa cross-attention yang memiliki F1-score sebesar 0.9925. Selain itu, model berbasis citra tunggal seperti DenseNet121 only dan Transformer only menunjukkan performa yang lebih rendah, masing-masing dengan F1-score sebesar 0.9900 dan 0.9786. Hasil ini menunjukkan bahwa penggunaan pendekatan dual-stream yang menggabungkan citra original dan segmented mampu meningkatkan performa model, dan integrasi menggunakan mekanisme cross-attention memberikan kontribusi tambahan dalam meningkatkan kualitas klasifikasi. Ringkasan perbandingan model ditunjukkan pada Tabel 5.

Tabel 5. Baseline Comparison

Model	Accuracy	Recall Macro	F1 Macro	AUC Macro
Original Image Only	0.997955	0.996667	0.997480	1.000000
Dual-stream + Cross-Attention	0.997955	0.996667	0.997480	1.000000
Dual-stream tanpa Cross-Attention	0.993865	0.994897	0.992532	0.999987
DenseNet121 Only	0.991820	0.991529	0.989996	0.999971
Segmented Image Only	0.989775	0.987889	0.988827	0.999932
Transformer Only	0.981595	0.980755	0.978601	0.999681

3.1.5 Ablation Study

Ablation study dilakukan untuk mengevaluasi kontribusi masing-masing komponen dalam model yang diusulkan, yaitu DenseNet121, Transformer, citra segmented, citra original, mekanisme cross-attention, serta penggunaan fungsi loss. Pengujian dilakukan dengan menghilangkan satu komponen pada setiap eksperimen dan membandingkan performanya dengan model penuh. Berdasarkan hasil pengujian, model dengan konfigurasi lengkap (full model) menunjukkan performa terbaik dibandingkan dengan sebagian besar variasi lainnya. Model dengan penggunaan focal loss bahkan mencapai performa maksimum dengan nilai accuracy, recall macro, dan F1-score macro sebesar 1.0000. Sementara itu, penghapusan komponen tertentu menyebabkan penurunan performa yang bervariasi. Penghilangan DenseNet121 menghasilkan penurunan paling signifikan dengan F1-score sebesar 0.9896, diikuti oleh penghilangan citra original dengan F1-score sebesar 0.9930. Selain itu, model tanpa cross-attention mengalami penurunan performa dengan F1-score sebesar 0.9975, yang menunjukkan bahwa mekanisme tersebut memberikan kontribusi terhadap peningkatan performa model. Penggunaan citra segmented dan Transformer juga terbukti memberikan kontribusi positif, meskipun penurunannya relatif lebih kecil dibandingkan komponen utama lainnya. Ringkasan hasil ablation study ditunjukkan pada Tabel 6.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 560−569
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9831

Tabel 6. Ablation Study

Model	Accuracy	Recall Macro	F1 Macro	AUC Macro
Full Model + Focal Loss	1.000000	1.000000	1.000000	1.000000
Tanpa Cross-Attention	0.997955	0.998311	0.997497	1.000000
Tanpa Transformer	0.997955	0.996667	0.997480	1.000000
Tanpa Segmented Image	0.997955	0.996667	0.997480	0.999987
Full Model tanpa Focal Loss	0.995910	0.994943	0.994960	0.999979
Tanpa Original Image	0.993865	0.991644	0.993045	0.999829
Full Model + Cross-Attention	0.993865	0.993288	0.992491	0.999974
Tanpa DenseNet121	0.991820	0.990845	0.989649	0.999712

Hasil ablation study menunjukkan bahwa penggunaan focal loss menghasilkan performa sangat tinggi hingga mencapai nilai 1.0000 pada seluruh metrik evaluasi. Namun, capaian ini perlu diinterpretasikan secara hati-hati. Skor sempurna pada dataset medis tidak selalu menunjukkan bahwa model telah memiliki kemampuan generalisasi yang kuat, melainkan dapat dipengaruhi oleh karakteristik dataset yang relatif bersih, variasi citra yang terbatas, atau kemungkinan kemiripan distribusi antara data latih dan data uji. Oleh karena itu, hasil tersebut tidak digunakan sebagai satu-satunya dasar klaim keunggulan model, tetapi diposisikan sebagai indikasi bahwa focal loss mampu meningkatkan separabilitas kelas pada dataset yang digunakan. Evaluasi tambahan menggunakan validasi silang, repeated experiment, dan pengujian pada dataset eksternal tetap diperlukan untuk memastikan bahwa performa model tidak disebabkan oleh overfitting.

3.1.6 Validasi Statistik

Validasi statistik dilakukan untuk mengevaluasi konsistensi dan signifikansi performa model yang diusulkan dibandingkan dengan model baseline tanpa mekanisme cross-attention. Pengujian dilakukan menggunakan beberapa pendekatan, yaitu repeated experiment dengan tiga random seed (42, 123, dan 2024), 5-fold cross-validation, serta uji statistik berupa paired t-test, Wilcoxon signed-rank test, dan McNemar test. Berdasarkan hasil repeated experiment, model yang diusulkan menunjukkan nilai F1-score yang konsisten tinggi, yaitu sebesar 0.9956, 0.9975, dan 0.9983, yang secara umum lebih tinggi dibandingkan model tanpa cross-attention. Pada pengujian 5-fold cross-validation, model yang diusulkan juga menunjukkan performa yang stabil dengan nilai F1-score berkisar antara 0.9929 hingga 1.0000, serta secara umum lebih unggul dibandingkan model baseline. Hasil uji statistik menunjukkan bahwa pada skenario 5-fold cross-validation, peningkatan performa model signifikan pada metrik accuracy dan F1-score dengan nilai p-value masing-masing sebesar 0.0341 dan 0.0352 ($p < 0.05$). Sementara itu, pada pengujian repeated experiment, perbedaan performa tidak menunjukkan signifikansi statistik. Selain itu, hasil McNemar test menunjukkan tidak terdapat perbedaan signifikan dalam distribusi kesalahan prediksi antara kedua model. Ringkasan hasil validasi statistik ditunjukkan pada Tabel 7.

Tabel 7. Validasi Statistik

Sumber	Metrik	Full Mean	Baseline Mean	Δ Mean	p-value	Signifikan
Repeated Seed	F1 Macro	0.997125	0.996068	0.001057	0.295475	Tidak
5-Fold CV	Accuracy	0.998157	0.993550	0.004607	0.034148	Ya
5-Fold CV	F1 Macro	0.997830	0.992732	0.005098	0.035164	Ya
McNemar	Prediction Error	-	-	-	1.000000	Tidak

3.2 Pembahasan

Hasil penelitian menunjukkan bahwa model yang diusulkan mampu mencapai performa yang sangat tinggi dalam klasifikasi multi-kelas Acute Lymphoblastic Leukemia (ALL), dengan nilai accuracy sebesar 0.9980 dan macro F1-score sebesar 0.9975. Performa tinggi ini sejalan dengan penelitian terbaru yang menunjukkan bahwa model berbasis deep learning, khususnya CNN, transfer learning, Vision Transformer, ensemble, dan model hybrid, memiliki kemampuan yang menjanjikan dalam klasifikasi leukemia berbasis citra darah maupun citra mikroskopis [7], [11], [22]. Stabilitas selama proses pelatihan, yang ditunjukkan oleh penurunan loss dan peningkatan F1-score, mengindikasikan bahwa model mampu mempelajari representasi data secara efektif tanpa mengalami overfitting yang signifikan. Hasil confusion matrix menunjukkan bahwa hampir seluruh prediksi berada pada diagonal utama, dengan hanya satu kesalahan klasifikasi. Hal ini menunjukkan kemampuan diskriminatif model yang sangat baik dalam membedakan karakteristik morfologi antar kelas. Temuan ini konsisten dengan penelitian sebelumnya yang menunjukkan bahwa pendekatan deep learning pada citra sel darah mampu menangkap karakteristik morfologi leukosit, termasuk tekstur, bentuk, dan struktur sel, yang relevan dalam deteksi serta klasifikasi leukemia [23].

Dari sisi arsitektur, penggunaan pendekatan dual-stream yang menggabungkan citra original dan segmented terbukti meningkatkan performa model. Kombinasi dua jenis input memungkinkan model memanfaatkan informasi komplementer, yaitu detail tekstur dari citra original dan struktur morfologi dari citra segmented. Pendekatan ini didukung oleh penelitian sebelumnya yang menunjukkan bahwa integrasi multi-input, multimodal, atau feature fusion dapat meningkatkan performa klasifikasi citra medis dibandingkan penggunaan satu jenis data saja [24]. Selain itu,

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 560−569
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9831

mekanisme cross-attention berperan dalam mengintegrasikan fitur lokal dan global secara adaptif. Hasil validasi statistik menunjukkan bahwa peningkatan performa model signifikan pada skenario 5-fold cross-validation ($p<0.05$), yang mengindikasikan bahwa cross-attention meningkatkan kemampuan generalisasi model. Temuan ini sejalan dengan literatur terbaru yang menunjukkan bahwa mekanisme attention dan model hybrid CNN–Transformer mampu menggabungkan keunggulan CNN dalam menangkap fitur lokal dengan kemampuan Transformer dalam memodelkan hubungan global pada citra medis [9], [11].

Hasil ablation study menunjukkan bahwa DenseNet121 dan citra original merupakan komponen yang memberikan kontribusi paling signifikan terhadap performa model. Hal ini menunjukkan pentingnya fitur lokal dalam klasifikasi leukemia, karena tekstur dan bentuk sel merupakan faktor utama dalam diagnosis berbasis citra mikroskopis. Sementara itu, penggunaan Transformer dan citra segmented memberikan kontribusi tambahan dalam menangkap hubungan spasial dan struktur global, yang juga telah didukung oleh penelitian mengenai Vision Transformer pada analisis citra medis [10], [25]. Meskipun model menunjukkan performa yang sangat tinggi, terdapat beberapa keterbatasan yang perlu diperhatikan. Dataset yang digunakan memiliki karakteristik yang relatif terstruktur dan bersih, sehingga memungkinkan model mencapai performa mendekati sempurna. Oleh karena itu, diperlukan pengujian lebih lanjut pada dataset eksternal untuk memastikan kemampuan generalisasi model dalam skenario klinis yang lebih kompleks. Hal ini sejalan dengan rekomendasi literatur terbaru yang menekankan pentingnya dataset yang lebih besar, beragam, validasi eksternal, standardisasi evaluasi, serta peningkatan interpretabilitas sebelum model deep learning dapat diterapkan secara klinis [11], [26].

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa arsitektur dual-stream DenseNet121–Transformer berbasis cross-attention merupakan pendekatan yang efektif dalam klasifikasi multi-kelas Acute Lymphoblastic Leukemia. Integrasi fitur lokal dan global secara adaptif terbukti meningkatkan performa dan stabilitas model, serta memberikan kontribusi yang signifikan dibandingkan pendekatan konvensional.

4. KESIMPULAN

Penelitian ini mengusulkan model klasifikasi multi-kelas Acute Lymphoblastic Leukemia (ALL) berbasis arsitektur dual-stream DenseNet121–Transformer dengan mekanisme cross-attention. Model dirancang untuk memanfaatkan dua jenis input, yaitu citra Peripheral Blood Smear (PBS) original dan segmented, sehingga mampu menggabungkan informasi fitur lokal dan global secara komplementer. Berdasarkan hasil eksperimen, model yang diusulkan mampu mencapai performa yang sangat tinggi dengan nilai accuracy sebesar 0.9980 dan macro F1-score sebesar 0.9975 pada data uji. Hasil ini didukung oleh stabilitas selama proses pelatihan, performa klasifikasi yang konsisten pada seluruh kelas, serta kesalahan prediksi yang sangat minimal berdasarkan confusion matrix. Selain itu, hasil perbandingan dengan model baseline menunjukkan bahwa pendekatan dual-stream dengan cross-attention memberikan performa yang lebih baik dibandingkan model berbasis single-stream maupun tanpa mekanisme attention. Hasil ablation study menunjukkan bahwa komponen DenseNet121, citra original, dan mekanisme cross-attention memberikan kontribusi signifikan terhadap peningkatan performa model. Validasi statistik melalui repeated experiment, 5-fold cross-validation, serta uji statistik menunjukkan bahwa peningkatan performa model bersifat konsisten dan signifikan pada beberapa metrik evaluasi. Meskipun demikian, penelitian ini masih memiliki keterbatasan, terutama pada penggunaan dataset yang relatif terstruktur dan terbatas. Oleh karena itu, penelitian selanjutnya disarankan untuk melakukan evaluasi pada dataset eksternal yang lebih beragam serta menguji model dalam skenario klinis nyata untuk memastikan kemampuan generalisasi. Selain itu, pengembangan model dengan integrasi metode interpretabilitas yang lebih mendalam juga dapat menjadi arah penelitian selanjutnya.

REFERENCES

[1] A. Krętowska-Grunwald *et al.*, “Significance of Th17 and Treg in Treatment Efficacy and Outcome in Pediatric Acute Lymphoblastic Leukemia,” *Int. J. Mol. Sci.*, vol. 24, no. 15, p. 12323, Aug. 2023, doi: 10.3390/ijms241512323.

[2] E. Kassa *et al.*, “Clinical profile and treatment outcomes of acute lymphoblastic leukemia among children attending at the University of Gondar comprehensive specialized hospital and Tikur Anbessa Specialized Hospital in Ethiopia,” *PLOS One*, vol. 20, no. 6, p. e0322747, Jun. 2025, doi: 10.1371/journal.pone.0322747.

[3] P. K. Das, D. V. A, S. Meher, R. Panda, and A. Abraham, “A Systematic Review on Recent Advancements in Deep and Machine Learning Based Detection and Classification of Acute Lymphoblastic Leukemia,” *IEEE Access*, vol. 10, pp. 81741–81763, 2022, doi: 10.1109/ACCESS.2022.3196037.

[4] I. Ahmad *et al.*, “Pediatric Acute Lymphoblastic Leukemia: Clinical Characteristics, Treatment Outcomes, and Prognostic Factors: 10 Years’ Experience From a Low- and Middle-Income Country,” *JCO Glob. Oncol.*, no. 9, p. e2200288, Jun. 2023, doi: 10.1200/GO.22.00288.

[5] A. D. Saucedo-Campos *et al.*, “Peripheral Blood as a Diagnostic Alternative to Bone Marrow in Immunophenotyping Pediatric B-Cell Acute Lymphoblastic Leukemia,” *Int. J. Mol. Sci.*, vol. 27, no. 1, p. 193, Dec. 2025, doi: 10.3390/ijms27010193.

[6] K. S. Al-Badrani *et al.*, “Immunophenotyping in diagnosing pediatric acute leukemia after setting up the first flow cytometry unit in Mosul City in Iraq: an observational study of the project performed through a contribution from Japan,” *Transl. Pediatr.*, vol. 14, no. 5, pp. 900–914, May 2025, doi: 10.21037/tp-2025-24.

[7] R. F. Oybek Kizi, T. P. Theodore Armand, and H.-C. Kim, “A Review of Deep Learning Techniques for Leukemia Cancer Classification Based on Blood Smear Images,” *Appl. Biosci.*, vol. 4, no. 1, p. 9, Feb. 2025, doi: 10.3390/applbiosci4010009.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 560−569
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9831

[8] T. Mustaqim, C. Fatichah, and N. Suciati, "Deep Learning for the Detection of Acute Lymphoblastic Leukemia Subtypes on Microscopic Images: A Systematic Literature Review," *IEEE Access*, vol. 11, pp. 16108–16127, 2023, doi: 10.1109/ACCESS.2023.3245128.
[9] G. Papanastasiou, N. Dikaios, J. Huang, C. Wang, and G. Yang, "Is Attention all You Need in Medical Image Analysis? A Review," *IEEE J. Biomed. Health Inform.*, vol. 28, no. 3, pp. 1398–1411, Mar. 2024, doi: 10.1109/JBHI.2023.3348436.
[10] S. Takahashi *et al.*, "Comparison of Vision Transformers and Convolutional Neural Networks in Medical Image Analysis: A Systematic Review," *J. Med. Syst.*, vol. 48, no. 1, p. 84, Sep. 2024, doi: 10.1007/s10916-024-02105-8.
[11] F. A. Mohammed, K. K. Tune, B. G. Assefa, M. Jett, and S. Muhie, "Medical Image Classifications Using Convolutional Neural Networks: A Survey of Current Methods and Statistical Modeling of the Literature," *Mach. Learn. Knowl. Extr.*, vol. 6, no. 1, pp. 699–736, Mar. 2024, doi: 10.3390/make6010033.
[12] R. F. Oybek Kizi, T. P. Theodore Armand, and H.-C. Kim, "A Review of Deep Learning Techniques for Leukemia Cancer Classification Based on Blood Smear Images," *Appl. Biosci.*, vol. 4, no. 1, p. 9, Feb. 2025, doi: 10.3390/applbiosci4010009.
[13] R. Azad *et al.*, "Advances in Medical Image Analysis with Vision Transformers: A Comprehensive Review," 2023, *arXiv*. doi: 10.48550/ARXIV.2301.03505.
[14] G. Papanastasiou, N. Dikaios, J. Huang, C. Wang, and G. Yang, "Is Attention all You Need in Medical Image Analysis? A Review," *IEEE J. Biomed. Health Inform.*, vol. 28, no. 3, pp. 1398–1411, Mar. 2024, doi: 10.1109/JBHI.2023.3348436.
[15] M. Aamir, Z. Rahman, N. Choudhry, J. Ahmed Bhutto, W. Ahmed Abro, and Z. Zhu, "From CNNs to Transformers: A Review of Evolving Deep Learning Architectures for Brain Tumor Classification," *IEEE Access*, vol. 13, pp. 184918–184936, 2025, doi: 10.1109/ACCESS.2025.3625607.
[16] X. Jiang, Z. Hu, S. Wang, and Y. Zhang, "Deep Learning for Medical Image-Based Cancer Diagnosis," *Cancers*, vol. 15, no. 14, p. 3608, Jul. 2023, doi: 10.3390/cancers15143608.
[17] Y. Li *et al.*, "A review of deep learning-based information fusion techniques for multimodal medical image classification," *Comput. Biol. Med.*, vol. 177, p. 108635, Jul. 2024, doi: 10.1016/j.compbiomed.2024.108635.
[18] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, "Transfer learning for medical image classification: a literature review," *BMC Med. Imaging*, vol. 22, no. 1, p. 69, Dec. 2022, doi: 10.1186/s12880-022-00793-7.
[19] K. He *et al.*, "Transformers in medical image analysis," *Intell. Med.*, vol. 3, no. 1, pp. 59–78, Feb. 2023, doi: 10.1016/j.imed.2022.07.002.
[20] Y. Li *et al.*, "A review of deep learning-based information fusion techniques for multimodal medical image classification," *Comput. Biol. Med.*, vol. 177, p. 108635, Jul. 2024, doi: 10.1016/j.compbiomed.2024.108635.
[21] G. Papanastasiou, N. Dikaios, J. Huang, C. Wang, and G. Yang, "Is Attention all You Need in Medical Image Analysis? A Review," *IEEE J. Biomed. Health Inform.*, vol. 28, no. 3, pp. 1398–1411, Mar. 2024, doi: 10.1109/JBHI.2023.3348436.
[22] T. Mustaqim, C. Fatichah, and N. Suciati, "Deep Learning for the Detection of Acute Lymphoblastic Leukemia Subtypes on Microscopic Images: A Systematic Literature Review," *IEEE Access*, vol. 11, pp. 16108–16127, 2023, doi: 10.1109/ACCESS.2023.3245128.
[23] M. Shahzad *et al.*, "Blood cell image segmentation and classification: a systematic review," *PeerJ Comput. Sci.*, vol. 10, p. e1813, Feb. 2024, doi: 10.7717/peerj-cs.1813.
[24] Y. Li *et al.*, "A review of deep learning-based information fusion techniques for multimodal medical image classification," *Comput. Biol. Med.*, vol. 177, p. 108635, Jul. 2024, doi: 10.1016/j.compbiomed.2024.108635.
[25] K. He *et al.*, "Transformers in medical image analysis," *Intell. Med.*, vol. 3, no. 1, pp. 59–78, Feb. 2023, doi: 10.1016/j.imed.2022.07.002.
[26] S. Q. Alhamrani *et al.*, "Machine Learning for Multi-Omics Characterization of Blood Cancers: A Systematic Review," *Cells*, vol. 14, no. 17, p. 1385, Sep. 2025, doi: 10.3390/cells14171385.