

Klasifikasi *Acute Lymphoblastic Leukemia* Berbasis Citra Mikroskopis Sel Darah Putih Menggunakan ResNet-50V2 dengan *Squeeze-and-Excitation* dan Grad-CAM

Harnelia^{*1}, Rizal Adi Saputra², Asa Hari Wibowo³

^{1,2,3}Informatika, Fakultas Teknik, Universitas Haluoleo, Indonesia

Email: ¹harnelia_e1e122056@student.uho.ac.id, ²rizaladisaputra@uho.ac.id, ³asa.hari@uho.ac.id

Abstrak

Acute Lymphoblastic Leukemia (ALL) merupakan salah satu penyakit keganasan hematologi yang memerlukan identifikasi secara dini dan akurat. Penelitian ini mengembangkan model klasifikasi berbasis citra menggunakan ResNet-50V2 dengan *Squeeze-and-Excitation* (SE) Block serta Grad-CAM untuk mengklasifikasikan citra sel darah putih ke dalam kelas ALL dan Non-ALL sekaligus memberikan visualisasi area yang berkontribusi terhadap prediksi model. Dataset yang digunakan terdiri dari 11.210 citra yang dikumpulkan dari beberapa sumber, termasuk dataset publik dan data klinis dari RS Kanker Dharmais. Model dilatih menggunakan strategi transfer learning dua fase dan dievaluasi menggunakan metrik akurasi, AUC-ROC, *precision*, *recall*, dan *F1-score*. Hasil pengujian menunjukkan bahwa model memperoleh akurasi sebesar 0,89 dan AUC-ROC sebesar 0,95 pada data uji. Evaluasi interpretabilitas menggunakan Grad-CAM melalui *Pointing Game Accuracy* (PGA) memperoleh nilai 100% pada sampel purposif dan 65% pada sampel acak. Hasil tersebut menunjukkan bahwa model dapat memberikan informasi visual mengenai area yang berkontribusi terhadap prediksi, meskipun konsistensi area yang ditunjukkan masih bervariasi pada beberapa citra. Model juga diimplementasikan ke dalam aplikasi berbasis web menggunakan *Flask* untuk menampilkan hasil klasifikasi dan visualisasi Grad-CAM. Sistem yang dikembangkan ditujukan sebagai alat bantu skrining awal ALL dan bukan sebagai alat diagnosis definitif.

Kata kunci: Citra Sel Darah Putih, Grad-CAM, Klasifikasi Leukemia, ResNet-50V2, *Squeeze-and-Excitation*.

Abstract

Acute Lymphoblastic Leukemia (ALL) is a hematological malignancy that requires early and accurate identification. This study develops an image-based classification model using ResNet-50V2 with a *Squeeze-and-Excitation* (SE) Block and Grad-CAM to classify white blood cell images into ALL and Non-ALL classes while providing visualizations of regions contributing to model predictions. The dataset consisted of 11,210 images collected from multiple sources, including public datasets and clinical data from RS Kanker Dharmais. The model was trained using a two-phase transfer learning strategy and evaluated using accuracy, AUC-ROC, *precision*, *recall*, and *F1-score*. The test results showed that the model achieved an accuracy of 0.89 and an AUC-ROC of 0.95 on the test set. Interpretability evaluation using Grad-CAM through *Pointing Game Accuracy* (PGA) achieved 100% on purposively selected samples and 65% on randomly selected samples. These results indicate that the model can provide visual information about regions contributing to its predictions, although the consistency of the highlighted regions varied across images. The model was also implemented in a web-based application using *Flask* to display classification results and Grad-CAM visualizations. The developed system is intended as an early screening aid for ALL and not as a definitive diagnostic tool.

Keywords: *Acute Lymphoblastic Leukemia*, Grad-CAM, ResNet-50V2, *Squeeze-and-Excitation*, White Blood Cell Images.

This work is an open access article and licensed under a Creative Commons Attribution-NonCommercial ShareAlike 4.0 International (CC BY-NC-SA 4.0)



1. PENDAHULUAN

Perkembangan teknologi informasi di bidang kesehatan, khususnya kedokteran komputasi, telah memberikan kontribusi besar dalam upaya penyelamatan jiwa manusia. Melalui teknologi ini, proses diagnosis penyakit dapat dilakukan secara lebih cepat dan akurat. Salah satu penyakit yang masih menjadi perhatian dunia medis adalah leukemia atau kanker darah, yaitu kelainan pada sistem hematologi yang ditandai oleh produksi sel darah putih abnormal dalam jumlah berlebihan [1]. Leukemia tidak hanya menyerang orang dewasa, tetapi juga merupakan salah satu kanker paling umum pada anak-anak secara global [2].

Berdasarkan data *World Health Organization* (WHO) melalui *Global Cancer Observatory* tahun 2020, tercatat sekitar 474.519 kasus baru leukemia di dunia dengan 311.594 kematian [3]. Sementara itu, data *International Agency for Research on Cancer* (IARC) mencatat 13.959 kasus baru dan 10.370 kematian akibat leukemia di Indonesia pada tahun 2022, menempatkan leukemia pada peringkat ke-9 berdasarkan jumlah kasus [4].

Di antara berbagai sub tipe leukemia, *Acute Lymphoblastic Leukemia* (ALL) merupakan jenis yang paling sering ditemukan pada anak-anak dan remaja. ALL memiliki karakteristik morfologi sel darah putih yang khas, terutama pada perubahan ukuran inti sel, rasio inti sitoplasma, dan tekstur kromatin, yang dapat diamati melalui citra apusan darah tepi [5]. Proses identifikasi leukemia secara konvensional masih bergantung pada pemeriksaan mikroskopis manual oleh analis laboratorium yang membutuhkan ketelitian tinggi, waktu yang relatif lama, dan berpotensi menimbulkan subjektivitas hasil [6].

Berbagai metode telah diterapkan dalam analisis citra medis untuk klasifikasi sel darah putih, seperti *K-Nearest Neighbor* (KNN) [7] dan *Multilayer Perceptron* (MLP) [8]. Namun, pendekatan *Convolutional Neural Network* (CNN) banyak digunakan dalam klasifikasi citra karena mampu mempelajari dan mengekstraksi fitur visual secara otomatis dari citra. Studi literatur menunjukkan bahwa kombinasi CNN dengan *attention mechanism* mampu meningkatkan kualitas ekstraksi fitur penting pada citra sel darah putih [5]. Selain itu, penelitian menggunakan arsitektur ResNet-50 dengan teknik *Explainable AI* seperti Grad-CAM memungkinkan visualisasi area citra yang menjadi dasar keputusan model [9].

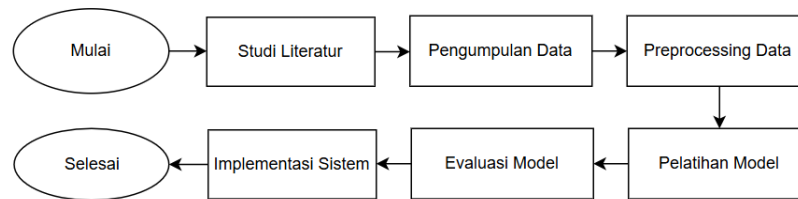
Beberapa penelitian terkini telah mengeksplorasi pendekatan serupa namun dengan cakupan yang masih terbatas. Jawahar dkk. [10] mengembangkan model CNN berbasis atensi untuk klasifikasi ALL dan berhasil mencapai akurasi tinggi, namun pengujian dilakukan secara eksklusif pada dataset publik ALL-IDB yang bersifat *homogen* sehingga kemampuan generalisasi model terhadap data klinis nyata belum tervalidasi. Diaz dkk. [9] menggabungkan ResNet-50 dengan Grad-CAM dan segmentasi U-Net, memperoleh performa yang sangat baik, tetapi arsitektur segmentasi berbasis *deep learning* yang digunakan memerlukan anotasi *ground truth* yang ekstensif dan beban komputasi tinggi sehingga kurang praktis untuk implementasi di fasilitas kesehatan dengan keterbatasan sumber daya. Bukhari dkk. [11] mengintegrasikan SE Block pada CNN dan mencapai akurasi 0,99 pada dataset ALL-IDB, namun tidak menyertakan mekanisme *explainability* sehingga keputusan model tidak dapat diverifikasi secara klinis oleh tenaga medis. Islam dkk. [12] menerapkan *Explainable AI* pada klasifikasi sel darah putih dengan pendekatan CNN yang dioptimasi, tetapi validasi kuantitatif terhadap ketepatan fokus atensi model belum dilakukan secara formal. Dari penelitian-penelitian tersebut, masih terdapat keterbatasan dalam mengintegrasikan mekanisme *attention*, *explainability*, dan validasi kuantitatif fokus atensi dalam satu pendekatan, terutama pada penggunaan dataset dengan sumber yang beragam. Berdasarkan identifikasi kesenjangan tersebut, penelitian ini mengusulkan pendekatan yang mengintegrasikan ResNet-50V2 dengan SE Block yang ringan secara komputasi, Grad-CAM sebagai mekanisme *explainability*, serta validasi kuantitatif fokus atensi model melalui *Pointing Game Accuracy* (PGA). Penelitian ini menggunakan dataset multi-sumber yang menggabungkan dataset publik dengan data klinis RS Kanker Dharmais, dengan sebagian data Dharmais digunakan sebagai bagian dari dataset utama untuk meningkatkan variasi data dan sebagian lainnya dipisahkan untuk pengujian sistem setelah model selesai dilatih dan tidak digunakan dalam pelatihan, validasi, maupun pengujian model pada dataset utama.

Berdasarkan hal tersebut, penelitian ini bertujuan: (1) merancang dan mengimplementasikan sistem klasifikasi *Acute Lymphoblastic Leukemia* (ALL) berbasis citra mikroskopis sel darah putih dengan arsitektur ResNet-50V2 pada dataset multi-sumber yang mencakup data klinis nyata; (2) menganalisis pengaruh integrasi *Squeeze-and-Excitation* (SE) Block terhadap peningkatan performa klasifikasi dibandingkan baseline tanpa mekanisme atensi; dan (3) mengevaluasi interpretabilitas keputusan model secara kuantitatif menggunakan metode *Pointing Game Accuracy* (PGA) untuk menilai kesesuaian fokus atensi Grad-CAM dengan area morfologi sel yang relevan.

2. METODE PENELITIAN

Metode penelitian dapat diartikan sebagai teknik ataupun cara-cara melaksanakan penelitian sampai menyusun laporannya berdasarkan fakta-fakta atau gejala-gejala secara ilmiah. Tahapan penelitian meliputi studi literatur, pengumpulan data, *preprocessing* data, pelatihan model menggunakan

arsitektur ResNet-50V2 dengan penambahan SE Block, evaluasi model menggunakan metrik kinerja serta visualisasi Grad-CAM, hingga tahap akhir berupa implementasi sistem dalam bentuk aplikasi berbasis *web*. Alur lengkap tahapan penelitian ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1. Studi Literatur

Studi Literatur yaitu suatu teknik yang digunakan untuk mencari ide atau sumber referensi dalam penelitian [13]. Pada penelitian ini dilakukan untuk mengkaji penelitian terdahulu yang relevan dengan topik klasifikasi leukemia berbasis citra digital, mencakup metode segmentasi sel darah putih, arsitektur *Convolutional Neural Network* (CNN), penerapan *attention mechanism* dalam klasifikasi citra medis, serta teknik *Explainable Artificial Intelligence* (XAI) seperti Grad-CAM. Selain itu, penelitian ini juga melibatkan wawancara dengan dokter spesialis patologi sebagai sumber informasi primer untuk memperoleh pemahaman klinis terkait karakteristik sel darah putih dan proses diagnosis leukemia.

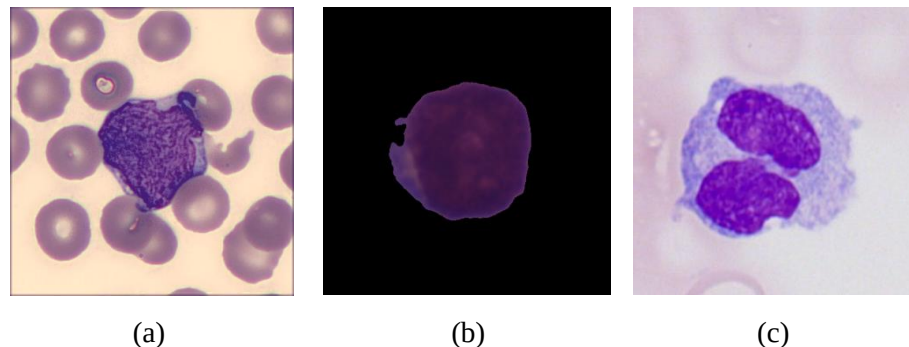
2.2. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini berupa citra mikroskopis sel darah putih yang diperoleh dari dua dataset dari sumber terbuka, yaitu (1) dataset *Acute Lymphoblastic Leukemia* (ALL) yang mencakup 5.555 citra ALL dan 3.000 citra Non-ALL [14] serta (2) dataset *Acute Myeloid Leukemia* (AML) yang terdiri atas 2.555 citra AML [15]. Selain itu, penelitian ini juga memanfaatkan data sekunder (3) diperoleh dari RS Kanker Dharmais, Jakarta, yang terdiri atas 100 citra ALL dan 100 citra Non-ALL. Dari 200 citra tersebut, 100 citra yang terdiri atas 50 citra ALL dan 50 citra Non-ALL dimasukkan ke dalam dataset utama, sedangkan 100 citra lainnya yang terdiri atas 50 citra ALL dan 50 citra Non-ALL dipisahkan sebagai data uji eksternal. Dengan demikian, dataset utama terdiri atas 11.210 citra, sedangkan 100 citra uji eksternal merupakan data terpisah yang tidak termasuk dalam jumlah tersebut. Dalam proses pengolahan data, kelas ALL dibentuk melalui penggabungan citra berlabel ALL dari dataset (1) dan 50 citra ALL dari dataset (3), sehingga diperoleh 5.605 citra ALL. Sementara itu, kelas Non-ALL dibentuk dari penggabungan 3.000 citra Non-ALL dari dataset (1), 2.555 citra AML dari dataset (2), dan 50 citra Non-ALL dari dataset (3), sehingga diperoleh 5.605 citra Non-ALL. Berikut rincian jumlah citra berdasarkan sumber data, kelas, dan penggunaannya dalam penelitian disajikan pada Tabel 1.

Tabel 1. Distribusi Citra Berdasarkan Sumber, Kelas, dan Penggunaan

Sumber data	ALL	Non-ALL	Total	Penggunaan
Dataset ALL publik	5.555	3.000	8.555	Dataset utama
Dataset AML publik	0	2.555	2.555	Dataset utama
RS Kanker Dharmais – dataset utama	50	50	100	Dataset utama
RS Kanker Dharmais – data uji eksternal	50	50	100	Data uji eksternal
Total dataset utama	5.605	5.605	11.210	-
Total seluruh citra	5.655	5.655	11.310	-

Contoh citra dari masing-masing sumber dataset yang digunakan dalam penelitian ditunjukkan pada Gambar 2.

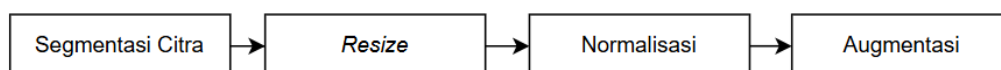


Gambar 2. (a) Data klinis RS Kanker Dharmais (kelas ALL); (b) Dataset *Acute Lymphoblastic Leukemia* (ALL); (c) Dataset *Acute Myeloid Leukemia* (AML).

Kelas Non-ALL didefinisikan sebagai citra yang bukan ALL, sehingga citra AML tetap termasuk dalam kelas Non-ALL meskipun AML merupakan jenis leukemia. Dataset utama kemudian dibagi secara *stratified* dengan perbandingan 70:15:15 untuk data latih, validasi, dan uji menggunakan *random seed* 42. Pembagian dilakukan pada tingkat citra, sehingga setiap citra hanya ditempatkan pada salah satu subset data. Namun, penelitian ini belum menerapkan pemisahan berbasis pasien atau preparat karena informasi identitas pasien atau preparat tidak tersedia pada dataset yang digunakan. Hasil pembagian menghasilkan 7.847 citra data latih yang terdiri atas 3.923 citra ALL dan 3.924 citra Non-ALL, 1.681 citra validasi yang terdiri atas 841 citra ALL dan 840 citra Non-ALL, serta 1.682 citra uji yang terdiri atas 841 citra ALL dan 841 citra Non-ALL. Selain dataset utama, Sebanyak 100 citra RS Kanker Dharmais lainnya, yang terdiri atas 50 citra ALL dan 50 citra Non-ALL, merupakan data uji eksternal yang tidak termasuk dalam dataset utama dan tidak digunakan dalam proses pelatihan, validasi, maupun pengujian model pada dataset utama. Data tersebut hanya digunakan setelah proses pelatihan model selesai untuk mengevaluasi kinerja sistem pada citra baru yang tidak pernah digunakan dalam proses pembelajaran model. Dengan demikian, 100 citra Dharmais yang digunakan dalam dataset utama dan 100 citra yang digunakan sebagai data uji eksternal merupakan dua kelompok citra yang berbeda dan tidak digunakan secara bersamaan. Proses pelatihan dan pengujian model dilakukan menggunakan *platform Kaggle* dengan akselerator GPU T4 $\times 2$, sedangkan pengembangan sistem dilakukan pada laptop HP dengan prosesor *Intel Core i5-1115G4*, RAM 8 GB, dan penyimpanan 512 GB. Implementasi menggunakan *Python 3.12*, *TensorFlow 2.14*, *OpenCV*, dan *scikit-learn* sebagai perangkat lunak dan pustaka utama.

2.3. Preprocessing Data

Pengolahan citra (image processing) merupakan salah satu proses awal dalam Perbaikan kualitas citra [16]. Seluruh citra melalui tahap preprocessing yang meliputi segmentasi, *resize*, normalisasi, dan augmentasi sebelum digunakan dalam pelatihan model.



Gambar 3. Tahapan *Preprocessing* Data

2.3.1. Segmentasi

Segmentasi citra dilakukan untuk memfokuskan area pengamatan pada bagian sel darah putih dengan mengurangi informasi latar belakang yang tidak relevan. Proses ini diterapkan pada data yang berasal dari RS Kanker Dharmais serta sebagian data dari *Kaggle* yang belum dalam kondisi tersegmentasi. Sementara itu, data *Kaggle* yang telah tersedia dalam bentuk tersegmentasi tidak diproses kembali. Langkah ini bertujuan untuk menyeragamkan area objek yang dianalisis pada seluruh dataset.

Proses segmentasi dilakukan dengan menerapkan *thresholding* pada ruang warna HSV untuk memisahkan objek sel dari latar belakang. Hasil *thresholding* direpresentasikan dalam bentuk mask biner yang didefinisikan sebagai berikut:

$$\text{Mask}(x, y) = \begin{cases} 1, & \text{jika } H, S, V \in \text{rentang } threshold \\ 0, & \text{lainnya} \end{cases} \quad (1)$$

Keterangan :

Mask(x,y) : citra biner hasil segmentasi pada koordinat piksel (x,y)

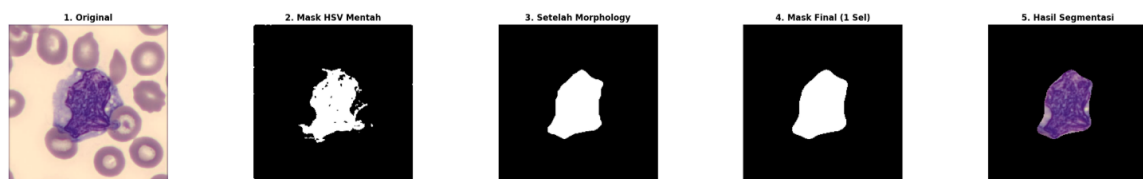
1 : piksel termasuk objek (sel leukosit)

0 : piksel termasuk latar belakang

H, S, V : komponen warna pada ruang HSV (Hue, Saturation, Value)

Rentang threshold : batas nilai HSV yang digunakan untuk mendeteksi objek

Proses segmentasi dilakukan dengan menerapkan *thresholding* pada ruang warna HSV untuk memisahkan objek sel dari latar belakang. Berdasarkan karakteristik pewarnaan Giemsa pada citra sel darah putih yang cenderung menghasilkan warna ungu-kebiruan pada area inti leukosit, rentang nilai *threshold* HSV yang digunakan ditetapkan sebagai berikut: H: 125–155, S: 80–255, V: 50–255 (skala OpenCV, $H \div 2$ dari rentang 0–360°). Rentang ini dipilih melalui observasi empiris terhadap sampel citra dari kedua sumber dataset untuk memastikan konsistensi segmentasi lintas sumber data. Hasil *thresholding* direpresentasikan dalam mask biner sebagaimana Persamaan (1). Selanjutnya diterapkan operasi morfologi *opening* (kernel elips 5×5, 2 iterasi) untuk menghilangkan noise kecil, diikuti operasi *closing* (kernel elips 25×25, 3 iterasi) untuk menutup lubang pada objek sel. Pengisian kontur internal dan seleksi kontur terbesar dilakukan untuk mengekstraksi satu objek sel darah putih dominan per citra. Contoh hasil segmentasi ditunjukkan pada Gambar 4.



Gambar 4. Proses Segmentasi HSV

Evaluasi segmentasi secara kuantitatif menggunakan IoU tidak dilakukan karena tidak tersedia anotasi *ground truth* tingkat piksel dari ahli patologi. Sebagai alternatif, validasi dialihkan menggunakan *Explainable AI* melalui metrik *Pointing Game Accuracy* (PGA) pada Grad-CAM. Pendekatan ini digunakan untuk mengevaluasi kesesuaian lokasi puncak aktivasi Grad-CAM terhadap area sel yang teridentifikasi melalui mask segmentasi HSV. Karena *mask* tersebut bukan anotasi *ground truth* dari ahli patologi, hasil PGA diinterpretasikan sebagai evaluasi otomatis berbasis mask segmentasi dan bukan sebagai validasi klinis terhadap ketepatan fokus atensi model. Penggunaan PGA dinilai lebih tepat karena karakteristik *heatmap* Grad-CAM yang bersifat menyebar dapat menurunkan skor IoU meskipun fokus model sudah akurat. Dengan demikian, PGA digunakan sebagai metrik untuk mengevaluasi kesesuaian fokus atensi Grad-CAM terhadap area sel yang menjadi acuan segmentasi.

2.3.2. Resize

Citra hasil segmentasi diubah ukurannya dari resolusi asli menjadi 224×224 piksel guna menyesuaikan dengan spesifikasi dimensi matriks masukan standar pada arsitektur model ResNet-50V2. Proses pengubahan ukuran dimensi ini diimplementasikan dengan menerapkan metode interpolasi bilinear. Pemilihan interpolasi bilinear dilakukan untuk mengubah dimensi citra agar sesuai dengan masukan ResNet-50V2, dengan tetap mempertahankan representasi visual citra sejauh dimungkinkan oleh proses pengubahan ukuran [17].

2.3.3. Normalisasi

Normalisasi data merupakan suatu tahapan untuk menyelaraskan rentang nilai dari beberapa variabel agar berada pada skala yang seragam. Proses ini berfungsi untuk mengeliminasi kesenjangan nilai yang terlalu ekstrem, baik terlalu besar maupun terlalu kecil, sehingga mempermudah jalannya analisis statistik [18]. Normalisasi dilakukan dalam dua tahap: (1) konversi nilai piksel ke rentang [0,1], dan (2) normalisasi menggunakan nilai mean dan standar deviasi ImageNet menggunakan persamaan Z-score sebagai berikut:

$$X' = \frac{\left(\frac{x}{255}\right) - \text{mean}(x)}{\text{std}(x)} \quad (2)$$

Keterangan:

X' : nilai piksel yang telah dinormalisasi

mean(x) : nilai mean channel

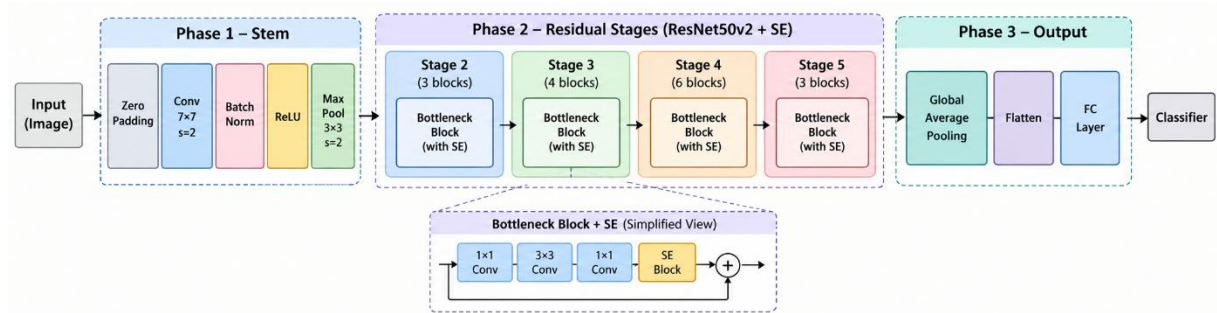
Std(x) : standar deviasi channel

2.3.4. Augmentasi

Augmentasi data merupakan teknik manipulasi citra yang diterapkan untuk memperbanyak jumlah sampel data latih secara efisien. Metode ini bertujuan untuk mengoptimalkan proses *training* model tanpa memerlukan tahapan pengumpulan atau pengambilan sampel citra baru di lapangan [19]. Proses augmentasi hanya diterapkan pada data latih dengan melakukan rotasi acak $\pm 15^\circ$, pergeseran horizontal dan vertikal sebesar 8% dari dimensi citra, *zoom* acak hingga 8%, serta *horizontal flip* dengan probabilitas 0,5. Parameter augmentasi ini ditetapkan secara konservatif untuk mempertahankan karakteristik morfologi sel darah putih, mengingat perubahan geometris yang terlalu ekstrem berpotensi mengubah ciri diagnostik yang khas seperti bentuk inti sel dan rasio nukleus-sitoplasma.

2.4. Perancangan Model

Model CNN yang digunakan adalah ResNet-50V2 dengan integrasi mekanisme atensi *Squeeze-and-Excitation* (SE) Block. ResNet-50V2 dipilih karena desain *pre-activation residual block* yang menempatkan *Batch Normalization* dan ReLU sebelum konvolusi, sehingga menghasilkan aliran gradien yang lebih stabil dan kemampuan ekstraksi fitur yang lebih baik dibandingkan ResNet-50 standar [20]. *Base model* diinisialisasi menggunakan bobot pralatih ImageNet (*include_top = False*) untuk mempertahankan representasi fitur hierarkis. SE Block dipilih sebagai mekanisme atensi dalam penelitian ini berdasarkan tiga pertimbangan utama. Pertama, SE Block bersifat ringan secara komputasi (*plug-and-play*) karena hanya menambahkan dua lapisan *fully connected* dengan rasio reduksi $r = 16$, sehingga peningkatan jumlah parameter model bersifat minimal dibandingkan arsitektur atensi berbasis spasial seperti CBAM atau *Non-Local Block*. Kedua, SE Block tidak mengubah struktur residual ResNet-50V2 secara drastis karena disisipkan setelah keluaran *base model* tanpa memodifikasi aliran gradien pada blok residual yang ada, sehingga kestabilan pelatihan tetap terjaga. Ketiga, SE Block terbukti efektif pada domain citra hematologi sebagaimana ditunjukkan oleh Bukhari dkk. [11]. Mekanisme ini bekerja melalui dua tahap: (1) *Squeeze* GlobalAveragePooling2D menghasilkan vektor deskriptor $z \in \mathbb{R}^C$; dan (2) *Excitation* dua lapisan Dense (2048 $\rightarrow$ 128 $\rightarrow$ 2048) dengan aktivasi ReLU dan sigmoid menghasilkan bobot kanal yang dikalikan secara elemen-per-elemen dengan *feature map* asli, sehingga fitur yang informatif diperkuat dan yang kurang relevan ditekan. *Classification head* terdiri dari GlobalAveragePooling2D, Dropout(0,5), Dense(128, ReLU), Dropout(0,3), dan Dense(2, softmax). Integrasi SE Block bertujuan untuk meningkatkan fokus model terhadap fitur penting pada citra. Struktur lengkap arsitektur ResNet-50V2 yang digunakan dalam penelitian ini ditunjukkan pada gambar 5.



Gambar 5. Arsitektur ResNet-50V2 + SE Block

2.5. Optimasi dan Pelatihan Model

Pada tahap pelatihan model, digunakan *optimizer* Adam yang mampu memperbarui bobot secara adaptif berdasarkan estimasi momen pertama dan kedua dari gradien, sehingga proses pembelajaran menjadi lebih stabil dan efisien [21], dengan fungsi *loss categorical cross-entropy* dan *metrik accuracy*. Proses pelatihan menggunakan strategi *transfer learning* dua fase. Pada fase pertama (*warm-up*), seluruh lapisan base model dibekukan dan hanya SE Block serta *classification head* yang dilatih selama maksimal 15 epoch dengan *learning rate* 1×10^{-3} . Pada fase kedua (*fine-tuning*), 50 lapisan terakhir dari base model diaktifkan untuk dilatih dengan *learning rate* 5×10^{-6} selama maksimal 100 epoch. Kedua fase menggunakan *batch size* 32 dan data validasi terpisah yang terdiri atas 15% dari dataset utama.

Untuk mengatur pembobotan kelas, digunakan *class weight* yang dihitung berdasarkan distribusi kelas pada data latih menggunakan metode *compute_class_weight* dengan *parameter balanced*. Selama pelatihan, *accuracy* digunakan sebagai metrik kinerja, sedangkan *val_accuracy* digunakan untuk memilih bobot model terbaik. Pada fase *fine-tuning*, diterapkan *EarlyStopping* berdasarkan *val_loss* dengan *patience* 8 dan *restore_best_weights*, serta *ReduceLROnPlateau* berdasarkan *val_loss* dengan faktor 0,3, *patience* 3, dan minimum *learning rate* 1×10^{-7} . *ModelCheckpoint* digunakan untuk menyimpan bobot dengan *val_accuracy* terbaik. Pada fase *warm-up*, hanya *ReduceLROnPlateau* yang diterapkan. Setelah proses pelatihan selesai, model diharapkan dapat mempelajari pola-pola representatif dari data latih dan siap digunakan untuk melakukan klasifikasi pada data baru [22]. Penelitian ini menggunakan satu skema pembagian data stratified dengan perbandingan 70:15:15 tanpa *k-fold cross-validation*. Oleh karena itu, variasi performa model pada pembagian data yang berbeda belum dievaluasi dan menjadi salah satu keterbatasan penelitian.

2.6. Evaluasi Model

Evaluasi model dilakukan untuk mengukur kinerja sistem dalam mengklasifikasikan citra sel darah putih ke dalam dua kelas, yaitu *Acute Lymphoblastic Leukemia* (ALL) dan Non-ALL. Evaluasi akhir model dilakukan menggunakan data uji yang tidak digunakan dalam proses pelatihan maupun pemilihan bobot model. Data validasi digunakan selama proses pelatihan untuk memantau performa model dan menentukan bobot terbaik. *Confusion matrix* yang diterapkan merupakan matriks berordo 2×2 yang menyajikan jumlah prediksi benar dan keliru untuk masing-masing kelas, sebagaimana ditunjukkan pada Tabel 2. Matriks ini digunakan untuk mengevaluasi kinerja model dalam mengklasifikasikan citra berdasarkan kesesuaian dengan label ground truth [23].

Tabel 2. *Confusion Matrix*

		Actual Class	
		ALL	Non_ALL
Predict Class	ALL	True Positive (TP) Sel ALL diprediksi benar sebagai ALL	False Positive (FP) Sel Non-ALL salah diprediksi sebagai ALL

	Non_ALL	<i>False Negative (FN)</i> Sel ALL salah diprediksi sebagai Non-ALL	<i>True Negative (TN)</i> Sel Non-ALL diprediksi benar sebagai Non-ALL
--	----------------	--	---

Selain menggunakan *confusion matrix*, penelitian ini juga memanfaatkan *classification report* untuk memberikan analisis yang lebih mendalam terhadap kinerja model pada data pengujian. Laporan ini mencakup metrik evaluasi seperti *precision*, *recall*, dan *f1-score*. *Precision* menunjukkan proporsi prediksi positif yang benar, sedangkan *recall* mengukur kemampuan model dalam mengidentifikasi seluruh data positif secara tepat. Adapun *f1-score* merupakan rata-rata harmonik dari *precision* dan *recall* yang digunakan untuk menyeimbangkan kedua metrik tersebut. Evaluasi juga dilakukan terhadap hasil visualisasi Grad-CAM menggunakan metode *pointing game*. Metode ini digunakan untuk menilai sejauh mana model mampu memfokuskan perhatian pada area yang relevan dalam citra. Dengan demikian, evaluasi yang dilakukan tidak hanya mengukur performa klasifikasi, tetapi juga interpretabilitas model dalam membedakan citra ALL dan Non-ALL secara lebih komprehensif.

2.7. Implementasi Sistem

Implementasi sistem pada penelitian ini berupa pengembangan aplikasi berbasis web yang digunakan untuk mengklasifikasikan citra menjadi kategori ALL dan Non-ALL. Aplikasi ini dibangun dengan mengintegrasikan model *deep learning* ResNet-50V2 yang telah dilatih sebelumnya ke dalam sistem *backend*. Pengguna dapat mengunggah citra melalui antarmuka web, kemudian sistem akan memproses citra tersebut dan menghasilkan prediksi kelas beserta tingkat kepercayaannya. Selain itu, sistem juga menampilkan visualisasi Grad-CAM untuk menunjukkan area citra yang menjadi fokus model dalam proses klasifikasi. Aplikasi ini dirancang untuk memudahkan pengguna dalam melakukan pengujian model secara langsung serta memberikan interpretasi visual terhadap hasil prediksi, sehingga meningkatkan transparansi dan pemahaman terhadap keputusan model. Untuk meningkatkan kemudahan penggunaan, aplikasi ini juga dilengkapi dengan fitur unduh (download) sehingga pengguna dapat menyimpan hasil prediksi dan visualisasi yang dihasilkan.

3. HASIL DAN PEMBAHASAN

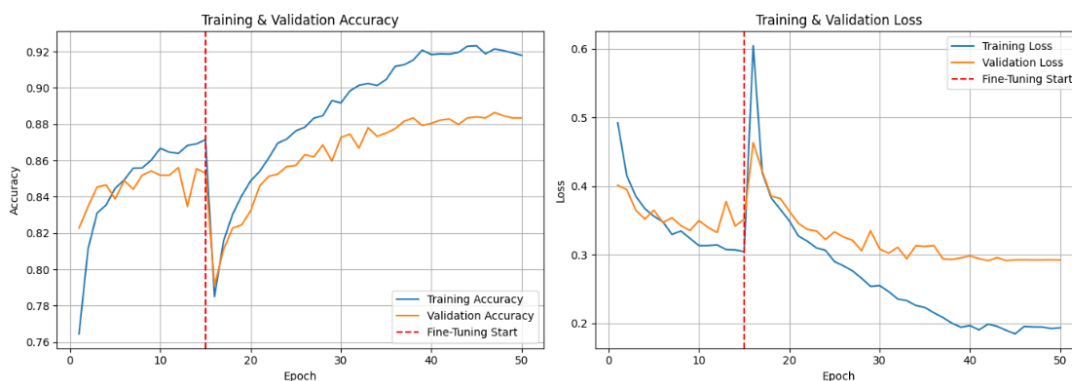
3.1. Hasil Pelatihan Model

Proses pelatihan model dilakukan dalam dua fase menggunakan strategi *transfer learning* berbasis bobot pra-latih ImageNet. Fase pertama (pemanasan) membekukan seluruh lapisan *base model* ResNet-50V2 dan hanya melatih SE Block serta *classification head* selama 15 epoch dengan *learning rate* 1×10^{-3} . Fase kedua (*fine-tuning*) membuka 50 lapisan terakhir *base model* dengan *learning rate* yang diturunkan menjadi 5×10^{-6} dan dibatasi hingga maksimal 100 epoch, disertai mekanisme *EarlyStopping* dengan *patience* 8 apabila *validation loss* tidak menunjukkan penurunan.

Berdasarkan grafik akurasi pelatihan dan validasi yang ditunjukkan pada Gambar 6, kurva *training accuracy* dan *validation accuracy* secara umum menunjukkan tren peningkatan yang konsisten. Memasuki epoch 15, tepat saat fase *fine-tuning* dimulai, sempat terjadi penurunan akurasi sesaat akibat guncangan adaptasi bobot jaringan baru, sebelum akhirnya model berhasil melakukan pemulihan (*recovery*) secara progresif. Model terbaik (ResNet-50V2 dengan *fine-tuning*) mencapai akurasi pelatihan sebesar 0,91 dan akurasi validasi sebesar 0,88. Selisih antara kedua kurva sebesar 3,15 poin persentase menunjukkan kedekatan performa antara data latih dan validasi. Nilai akurasi validasi sebesar 0,88 yang mendekati akurasi pengujian sebesar 0,89 menunjukkan bahwa performa model pada kedua subset tersebut relatif berdekatan. Namun, hasil ini belum cukup untuk menyimpulkan kemampuan generalisasi model pada data dari sumber atau fasilitas kesehatan lain.

Sementara itu, kurva *loss* menunjukkan penurunan yang stabil pada kedua set data. Sama halnya dengan kurva akurasi, pada awal fase *fine-tuning* (epoch 15) kurva *loss* sempat mengalami lonjakan tajam (*spike*) hingga menyentuh nilai ~ 0.60 , yang kemudian melandai secara gradual seiring berjalannya epoch. Tidak terdapat perbedaan yang mencolok antara *training loss* dan *validation loss* di akhir proses

pelatihan, yang menunjukkan bahwa strategi pelatihan dua fase yang diterapkan mampu mengoptimalkan representasi fitur secara bertahap tanpa menyebabkan *overfitting*. Mekanisme *ReduceLROnPlateau* dan pembobotan kelas (*class weight*) yang diterapkan selama pelatihan turut berkontribusi dalam menstabilkan proses pembelajaran, terutama mengingat adanya variasi distribusi kelas pada dataset yang digunakan



Gambar 6. Grafik *Training* dan *Validation Accuracy*

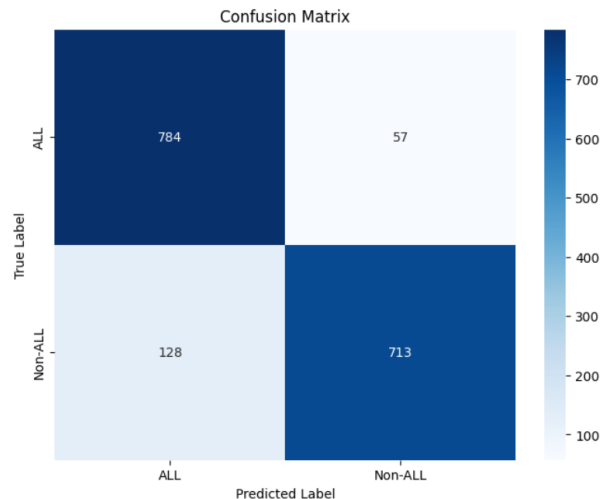
Secara keseluruhan, peningkatan performa model dari tahap awal hingga proses *fine-tuning*, serta kestabilan pergerakan kurva yang dihasilkan pada epoch akhir, menunjukkan bahwa pendekatan pelatihan yang diterapkan berhasil meningkatkan kinerja model secara konsisten. Dengan demikian, implementasi metode *transfer learning* yang dikombinasikan dengan teknik *fine-tuning* menunjukkan peningkatan kinerja pada eksperimen yang dilakukan.

3.2. Evaluasi Model

3.2.1. Confusion Matrix

Evaluasi performa model dilakukan menggunakan model terbaik ResNet-50V2 yang telah melalui tahap *fine-tuning*, dengan pengujian pada 1.682 citra dari subset data uji yang terbagi secara seimbang antara kelas ALL dan Non-ALL. Subset ini tidak digunakan dalam proses pelatihan maupun pemilihan bobot model. Pengujian ini bertujuan untuk melihat kemampuan model dalam mempertahankan akurasi serta melakukan generalisasi terhadap data baru yang tidak terlibat selama proses pelatihan. Penilaian performa dilakukan menggunakan dua pendekatan utama, yaitu *confusion matrix* untuk melihat distribusi prediksi secara rinci, serta *classification report* untuk mengukur nilai *precision*, *recall*, dan *F1-score*.

Persebaran hasil prediksi model terhadap data uji tersebut secara mendetail dapat dilihat pada Gambar 7. Berdasarkan visualisasi grafik *confusion matrix* pada gambar tersebut, model ResNet-50V2 menunjukkan kemampuan mengidentifikasi sebagian besar sampel pada kedua kelas dalam mengidentifikasi objek target dengan mencatatkan nilai *True Positive* (TP) sebesar 784 sampel pada kelas ALL dan *True Negative* (TN) sebesar 713 sampel pada kelas Non-ALL. Meskipun demikian, masih ditemukan adanya margin kesalahan berupa 57 sampel *False Negative* (FN) dan 128 sampel *False Positive* (FP). *False Positive*, yaitu citra Non-ALL yang diprediksi sebagai ALL, dapat terjadi karena adanya kemiripan karakteristik visual antara sel pada citra Non-ALL dan karakteristik morfologi limfoblas yang dipelajari model. Sementara itu, *False Negative*, yaitu citra ALL yang diprediksi sebagai Non-ALL, dapat terjadi akibat variasi karakteristik morfologi sel limfoblas pada citra yang menyulitkan model dalam mengenali pola yang relevan. Penyebab tersebut merupakan interpretasi terhadap pola kesalahan klasifikasi dan tidak didasarkan pada pemeriksaan manual terhadap setiap citra yang mengalami kesalahan prediksi.



Gambar 7. Hasil *Confusion Matrix*

Hasil tersebut kemudian direpresentasikan dalam bentuk persentase pada Tabel 3 melalui *classification report*. Kelas ALL memperoleh nilai *recall* sebesar 93% dan *precision* sebesar 86%, yang menunjukkan kemampuan model yang baik dalam mendeteksi kasus positif. Nilai *recall* yang tinggi pada kelas ini penting karena dapat meminimalkan kemungkinan terlewatnya kasus leukemia. Sementara itu, kelas Non-ALL memiliki nilai *precision* sebesar 93% dan *recall* sebesar 85%, yang menunjukkan bahwa prediksi pada kelas ini cenderung akurat meskipun masih terdapat beberapa kesalahan klasifikasi. Secara keseluruhan, model mencapai akurasi sebesar 0,89 dengan nilai *macro average F1-score* sebesar 0,89. Hasil ini menunjukkan bahwa model memiliki performa yang relatif seimbang pada kedua kelas dalam subset data uji yang digunakan. Namun, hasil tersebut belum cukup untuk menyimpulkan kemampuan generalisasi model pada pasien, preparat, atau fasilitas kesehatan lain. Hal ini mengindikasikan bahwa pendekatan *transfer learning* yang dikombinasikan dengan *finetuning* memberikan hasil yang efektif dalam tugas klasifikasi citra yang dilakukan.

Tabel 3. *Classification Report*

Kelas	Precision	Recall	F1-Score	Support
ALL	0,86	0,93	0,89	841
Non-ALL	0,93	0,85	0,89	841
Akurasi			0,89	1.682
Macro Average	0,89	0,89	0,89	1.682
Weighted Average	0,89	0,89	0,89	1.682

Berdasarkan Tabel 4, model ResNet-50V2 dengan SE Block memperoleh peningkatan akurasi dari 0,86 menjadi 0,89, AUC-ROC dari 0,91 menjadi 0,95, dan *macro F1-score* dari 0,85 menjadi 0,89 dibandingkan model *baseline* tanpa SE Block. Penggunaan SE Block juga menurunkan jumlah *False Positive* dari 198 menjadi 128, meskipun jumlah *False Negative* meningkat dari 45 menjadi 57. Hasil ini menunjukkan peningkatan performa keseluruhan setelah integrasi SE Block, terutama dalam mengidentifikasi kelas Non-ALL, dengan catatan bahwa jumlah kesalahan *False Negative* pada kelas ALL juga meningkat..

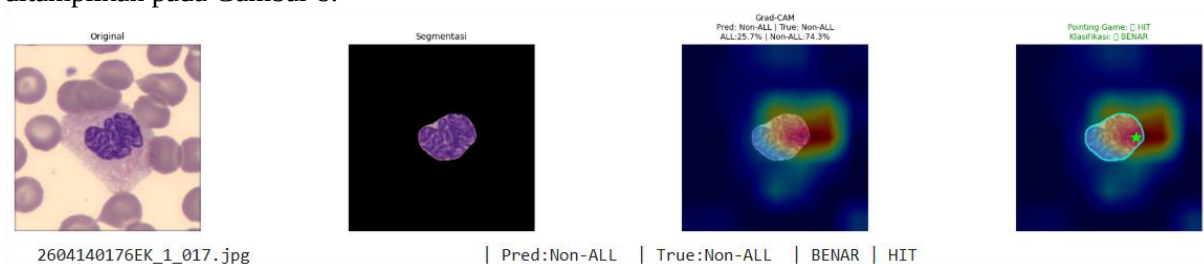
Tabel 4. Perbandingan Performa Model *Baseline* dan Model Usulan

Model	Accuracy	AUC-ROC	Macro F1	FP	FN
ResNet-50V2 tanpa SE Block	0,86	0,91	0,85	198	45
ResNet-50V2 + SE Block	0,89	0,95	0,89	128	57

3.2.2 Pointing Game Accuracy

Evaluasi interpretabilitas model dilakukan menggunakan metode *Pointing Game Accuracy* (PGA) melalui dua skenario pengujian, yaitu sampel purposif dan sampel acak. Pada pengujian pertama, digunakan 20 sampel citra yang dipilih secara purposif dan representatif, terdiri dari 10 citra kelas ALL dan 10 citra kelas Non-ALL yang mencakup variasi morfologi sel dari kedua sumber dataset. Pemilihan secara purposif ini bertujuan untuk memastikan keterwakilan karakteristik morfologi yang beragam, seperti sel ALL dengan bentuk inti bulat dan elips serta sel Non-ALL dari berbagai kondisi hematologi. Hasil evaluasi pada sampel ini menunjukkan nilai PGA sebesar 100%, di mana puncak energi atensi Grad-CAM berada pada area sel darah putih yang tersegmentasi (HIT).

Selanjutnya, untuk memperoleh gambaran performa yang lebih umum, dilakukan pengujian kedua menggunakan 100 sampel citra yang dipilih secara acak, terdiri dari 50 citra kelas ALL dan 50 citra kelas Non-ALL. Hasil pengujian ini menunjukkan nilai PGA sebesar 65%, yang mengindikasikan bahwa pada sebagian citra, atensi model masih terdistraksi oleh latar belakang. Oleh karena itu, nilai PGA sebesar 100% pada sampel purposif menunjukkan kesesuaian fokus atensi model pada sampel yang dipilih, sedangkan nilai PGA sebesar 65% pada sampel acak menunjukkan bahwa kesesuaian fokus atensi masih bervariasi pada citra yang diuji. Hasil ini memberikan gambaran mengenai kesesuaian area atensi Grad-CAM terhadap area sel yang tersegmentasi, tetapi belum dapat dianggap sebagai validasi klinis oleh ahli patologi. Contoh citra yang digunakan dalam evaluasi *Pointing Game Accuracy* ditampilkan pada Gambar 8.



Gambar 8. Contoh Citra *Pointing Game Accuracy*

3.3. Pembahasan

Hasil penelitian menunjukkan bahwa model ResNet-50V2 yang terintegrasi dengan *Squeeze-and-Excitation* (SE) Block dan Grad-CAM mampu melakukan klasifikasi ALL dan Non-ALL pada dataset multi-sumber dengan akurasi sebesar 0,89 dan AUC-ROC sebesar 0,95. Performa tersebut diperoleh pada dataset yang menggabungkan citra dari beberapa sumber dengan karakteristik data yang beragam, sehingga kondisi pengujian tidak sepenuhnya homogen seperti pada penelitian yang menggunakan satu dataset publik. Variasi sumber data dapat mencakup perbedaan kualitas citra, kondisi preparasi, perangkat akuisisi, serta karakteristik morfologi sel. Selain pengujian pada dataset utama, pengujian sistem juga dilakukan menggunakan 100 citra RS Kanker Dharmais yang dipisahkan dari proses pelatihan, validasi, dan pengujian model pada dataset utama. Kondisi data riil yang bervariasi ini sejalan dengan tinjauan sistematis oleh Acevedo dkk. [24], yang menegaskan bahwa performa model *deep learning* pada domain hematologi sering kali mengalami penurunan drastis saat dihadapkan pada data klinis nyata akibat tingginya artefak visual dan variasi preparasi apusan darah. Tantangan ini dipertegas oleh Matek dkk [25], melalui konsep variabilitas antar-institusi (*inter-site variability*). Mereka menyoroti bahwa perbedaan protokol pewarnaan dan jenis pemindai (*scanner*) antar-laboratorium memicu pergeseran distribusi data (*domain shift*) yang menjadi hambatan utama dalam generalisasi model secara universal.

Kondisi tersebut dapat menjelaskan mengapa akurasi 0,89 pada penelitian ini lebih rendah dibandingkan beberapa penelitian terdahulu yang melaporkan akurasi hingga 0,99. Perbedaan tersebut perlu dipertimbangkan dengan melihat kondisi eksperimen, terutama karakteristik dan sumber dataset yang digunakan. Dataset tunggal yang homogen cenderung memiliki distribusi citra yang lebih seragam, sedangkan dataset multi-sumber dalam penelitian ini menghadirkan variasi yang lebih besar. Oleh

karena itu, perbedaan nilai akurasi tidak dapat dibandingkan secara langsung tanpa mempertimbangkan perbedaan dataset, jumlah data, preprocessing, arsitektur, dan skema pengujian.

Terkait 128 sampel *False Positive* (FP) dan 57 sampel *False Negative* (FN) yang tercatat, penelitian ini belum melakukan pemeriksaan manual terhadap setiap citra yang mengalami kesalahan prediksi. Oleh karena itu, penyebab spesifik dari masing-masing kesalahan tidak dapat ditentukan secara pasti. Secara umum, *False Positive*, yaitu citra Non-ALL yang diprediksi sebagai ALL, dapat terjadi ketika karakteristik visual citra Non-ALL memiliki kemiripan dengan pola morfologi yang dipelajari model sebagai karakteristik ALL. Sebaliknya, *False Negative*, yaitu citra ALL yang diprediksi sebagai Non-ALL, dapat terjadi ketika variasi morfologi sel ALL menyebabkan pola yang relevan lebih sulit dikenali oleh model. Analisis terhadap contoh citra FP dan FN secara individual dengan mempertahankan informasi sumber data serta melibatkan ahli patologi klinik diperlukan pada penelitian selanjutnya untuk mengidentifikasi karakteristik kesalahan model secara lebih spesifik.

Berdasarkan hasil pengujian, model memperoleh akurasi sebesar 0,89 dan AUC-ROC sebesar 0,95 pada data uji multi-sumber. Untuk menganalisis pengaruh SE Block, dilakukan perbandingan dengan model baseline ResNet-50V2 tanpa SE pada dataset dan skema pengujian yang sama. Model *baseline* memperoleh akurasi sebesar 0,86 dan AUC-ROC sebesar 0,91, sedangkan model dengan SE Block memperoleh akurasi sebesar 0,89 dan AUC-ROC sebesar 0,95. Nilai *macro average F1-score* juga meningkat dari 0,85 menjadi 0,89. Selain itu, penggunaan SE Block menurunkan jumlah *False Positive* dari 198 menjadi 128 sampel dan meningkatkan *recall* kelas Non-ALL dari 0,76 menjadi 0,85. Namun, jumlah *False Negative* meningkat dari 45 menjadi 57 sampel dan *recall* kelas ALL sedikit menurun dari 0,95 menjadi 0,93. Hasil tersebut menunjukkan bahwa integrasi SE Block memberikan peningkatan performa secara keseluruhan, terutama pada kemampuan model dalam mengidentifikasi kelas Non-ALL, meskipun terdapat peningkatan kesalahan *False Negative* pada kelas ALL. Hasil pada penelitian ini berbeda dengan beberapa penelitian yang melaporkan akurasi lebih tinggi, tetapi menggunakan test set dari satu sumber dengan kondisi data yang lebih terkontrol [26]. Perbedaan karakteristik dataset tersebut menunjukkan bahwa nilai akurasi perlu dipertimbangkan berdasarkan kondisi eksperimen yang digunakan. Selanjutnya, hasil klasifikasi divisualisasikan melalui modul Grad-CAM untuk memberikan gambaran mengenai area citra yang berkontribusi terhadap keputusan model. Sejalan dengan penelitian terkait explainable AI pada bidang medis [27], peta Grad-CAM pada beberapa hasil pengujian menunjukkan aktivasi pada area sel, termasuk struktur nukleus dan sitoplasma. Visualisasi tersebut dapat digunakan sebagai informasi pendukung untuk memahami fokus model dalam menghasilkan prediksi.

3.4. Keterbatasan Penelitian

Penelitian ini telah menghasilkan sistem klasifikasi leukemia yang fungsional dan menunjukkan performa yang kompetitif, namun terdapat beberapa keterbatasan yang perlu diakui secara transparan untuk memberikan konteks yang tepat terhadap penerapannya secara klinis. Pertama, sistem yang dikembangkan dirancang dan divalidasi sebagai alat bantu skrining awal, bukan sebagai instrumen diagnosis definitif. Keputusan diagnosis ALL tetap merupakan kewenangan eksklusif dokter spesialis patologi klinik atau hematologi yang mempertimbangkan temuan klinis, hasil laboratorium lengkap, dan pemeriksaan lanjutan seperti imunofenotipe dan analisis sitogenetik. Sistem ini berperan sebagai lapis pertama untuk memflagasi citra yang berpotensi ALL guna memprioritaskan pemeriksaan lebih lanjut, khususnya di fasilitas layanan kesehatan tingkat pertama yang memiliki keterbatasan akses terhadap tenaga ahli patologi. Kedua, validasi interpretabilitas melalui *Pointing Game Accuracy* dilakukan secara otomatis menggunakan mask segmentasi HSV sebagai *proxy ground truth*, dan belum divalidasi melalui anotasi langsung oleh ahli patologi. Validasi berbasis ahli patologi yang menandai area diagnostik secara manual pada sampel representatif merupakan langkah penting yang perlu ditempuh pada penelitian lanjutan untuk mengkonfirmasi kesesuaian fokus atensi model dengan penilaian klinis aktual. Ketiga, pengujian sistem saat ini masih dilakukan dalam lingkungan komputasi terkontrol. Pengujian prospektif di fasilitas kesehatan nyata dengan kondisi akuisisi citra yang bervariasi termasuk perbedaan jenis mikroskop, kondisi pewarnaan, dan operator belum dilakukan. Penelitian lanjutan direncanakan untuk melakukan uji klinis prospektif di fasilitas kesehatan mitra guna mengukur performa sistem secara langsung dalam alur kerja diagnostik aktual. Keempat, analisis terhadap pola kesalahan model khususnya pada 128 sampel *False Positive* dibatasi oleh tidak adanya metadata sumber per sampel pada

dataset gabungan yang digunakan, sehingga identifikasi subtype sel Non-ALL yang paling sering salah diklasifikasikan tidak dapat dilakukan secara kuantitatif dalam penelitian ini.

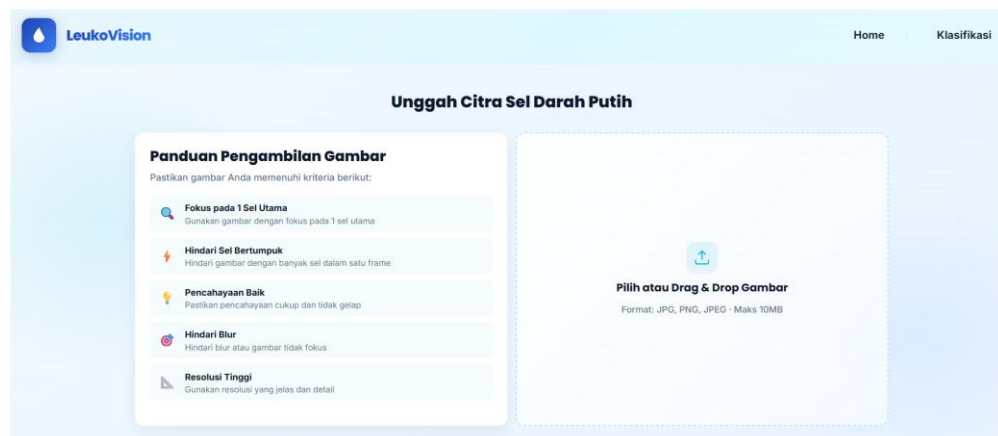
3.5. Tampilan Antarmuka Sistem

Implementasi antarmuka pengguna (*user interface*) pada sistem klasifikasi leukemia yang dikembangkan bertujuan untuk menjembatani model komputasi *deep learning* pada sisi backend agar dapat diakses secara mudah, interaktif, dan informatif oleh pengguna.



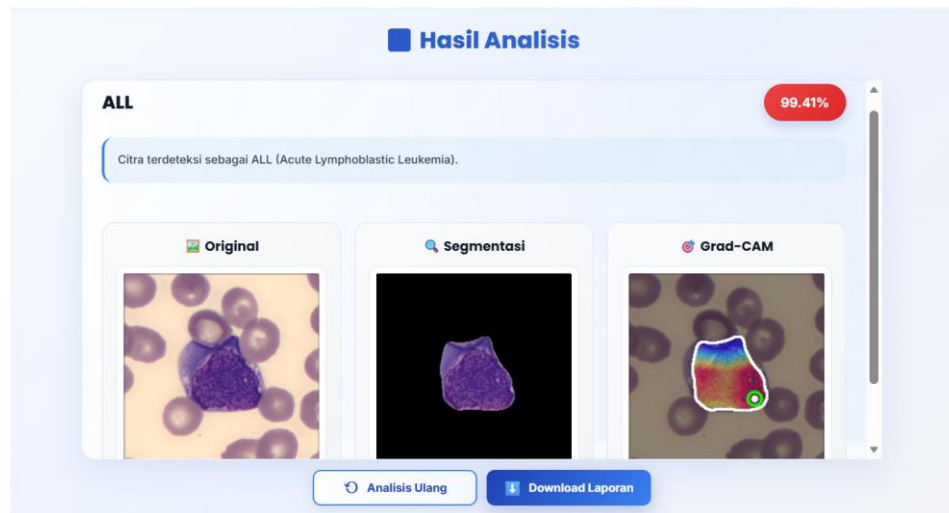
Gambar 9. Tampilan Halaman Awal Sistem

Halaman awal (*home*) dari sistem klasifikasi ini dapat dilihat pada Gambar 9. Berdasarkan tampilan antarmuka pada gambar tersebut, halaman awal menyajikan informasi pengantar mengenai platform yang diberi nama LeukoVision. Antarmuka ini dilengkapi dengan menu navigasi utama di pojok kanan atas, yaitu menu *Home* dan *Klasifikasi*, serta tombol aksi utama *Mulai Analisis* untuk mengarahkan pengguna ke fitur inti sistem.



Gambar 10. Tampilan Halaman *Upload* Gambar

Proses input data pada sistem difasilitasi melalui halaman unggah citra sel darah putih yang disajikan pada Gambar 10. Berdasarkan visualisasi antarmuka tersebut, halaman dirancang secara ergonomis dengan membagi ruang menjadi dua fungsi utama yang saling mendukung. Pada sisi kanan, terdapat fitur interaktif *Pilih atau Drag & Drop Gambar*. Sementara itu, sisi kiri halaman dilengkapi dengan panel *Panduan Pengambilan Gambar*.



Gambar 11. Tampilan Halaman Hasil Klasifikasi

Implementasi antarmuka halaman hasil diagnosis sistem disajikan pada Gambar 11. Berdasarkan visualisasi tersebut, halaman ini mengintegrasikan luaran model komputasi ke dalam komponen informasi yang komprehensif bagi pengguna. Sisi atas menampilkan hasil klasifikasi final berupa label prediksi disertai tingkat keyakinan (*confidence score*). Untuk menjamin transparansi diagnosis, bagian inti halaman menyajikan tiga panel visualisasi citra secara berdampingan: citra asli (*Original*), citra hasil isolasi objek (*Segmentasi*) dengan latar belakang hitam, serta peta aktivasi kelas (*Grad-CAM*) untuk menunjukkan area fitur visual yang menjadi fokus keputusan model. Fungsionalitas halaman ini ditutup dengan tombol aksi *Analisis Ulang* dan *Download Laporan* untuk mempermudah manajemen data klinis pengguna.

LeukoVision - Laporan Analisis

Tanggal: 1/6/2026, 02.41.36

Hasil Klasifikasi:

Klasifikasi: ALL

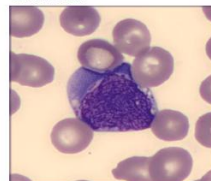
Akurasi: 99.41%

Keterangan:

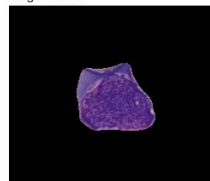
Citra terdeteksi sebagai ALL (Acute Lymphoblastic Leukemia).

Visualisasi Analisis:

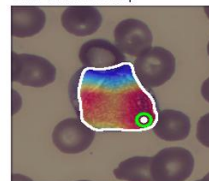
Gambar Asli



Segmentasi



Grad-CAM Heatmap



Gambar 12. Tampilan Unduh Hasil Klasifikasi

Tampilan luaran dokumen laporan dalam format PDF yang diunduh oleh pengguna disajikan pada Gambar 12. Dokumen cetak ini berfungsi sebagai rekam medis digital resmi dari *platform* LeukoVision yang memuat seluruh riwayat pemeriksaan secara ringkas. Pada bagian *header*, dokumen mencantumkan cap waktu (*timestamp*) deteksi untuk validasi data klinis.

4. KESIMPULAN

Penelitian ini menerapkan arsitektur ResNet-50V2 dengan *Squeeze-and-Excitation* (SE) Block dan Grad-CAM untuk klasifikasi citra sel darah putih menjadi *Acute Lymphoblastic Leukemia* (ALL) dan Non-ALL. Model dilatih menggunakan strategi *transfer learning* dua fase dan dievaluasi pada data

uji sebanyak 1.682 citra. Hasil pengujian menunjukkan bahwa model mencapai akurasi sebesar 0,89, AUC-ROC sebesar 0,95, dan macro F1-score sebesar 0,89. Dibandingkan model ResNet-50V2 tanpa SE Block, model dengan SE Block meningkatkan akurasi dari 0,86 menjadi 0,89, AUC-ROC dari 0,91 menjadi 0,95, serta menurunkan jumlah False Positive dari 198 menjadi 128, meskipun jumlah *False Negative* meningkat dari 45 menjadi 57.

Evaluasi interpretabilitas menggunakan Grad-CAM melalui *Pointing Game Accuracy* (PGA) memperoleh nilai 100% pada 20 sampel purposif dan 65% pada 100 sampel acak. Hasil tersebut menunjukkan bahwa kesesuaian fokus atensi model belum konsisten pada seluruh citra, sehingga visualisasi Grad-CAM perlu digunakan sebagai informasi pendukung dalam memahami prediksi model. Secara fungsional, model telah diimplementasikan ke dalam aplikasi berbasis web menggunakan *framework Flask* untuk menampilkan hasil klasifikasi dan visualisasi Grad-CAM. Sistem yang dikembangkan ditujukan sebagai alat bantu skrining awal ALL dan bukan sebagai alat diagnosis definitif, sehingga hasil prediksi tetap memerlukan interpretasi dan pemeriksaan oleh tenaga medis.

5. DAFTAR PUSTAKA

- [1] S. Levianto dan A. A. Soebroto, "Sistem Pakar Diagnosis Penyakit Leukimia dengan Metode Fuzzy Tsukamoto," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 12, hlm. 5329–5338, 2021, [Daring]. Tersedia pada: <http://j-ptiik.ub.ac.id>
- [2] W. Liu, J. Fang, M. Zhu, J. Zhou, dan C. Yuan, "Global, regional, and National burden of childhood leukemia from 1990 to 2021," *BMC Pediatr.*, vol. 25, no. 1, Des 2025, doi: 10.1186/s12887-025-05847-7.
- [3] J. Huang dkk., "Disease Burden, Risk Factors, and Trends of Leukaemia: A Global Analysis," *Front. Oncol.*, vol. 12, Jul 2022, doi: 10.3389/fonc.2022.904292.
- [4] International Agency for Research on Cancer, Global Cancer Observatory: Cancer Today, World Health Organization, 2020. [Online]. Available: <https://gco.iarc.fr/today>.
- [5] M. Z. Ullah dkk., "An attention-based convolutional neural network for acute lymphoblastic leukemia classification," *Applied Sciences (Switzerland)*, vol. 11, no. 22, Nov 2021, doi: 10.3390/app112210662.
- [6] A. N. Hasanah, O. Al Sas, Y. D. Kosen, dan J. Kesehatan Budi, "34 Application of Artificial Intelligence in White Blood Cell Classification Application Of Artificial Intelligence In White Blood Cell Classification Based On Microscopic Images: A Scoping Review," *Jurnal Kesehatan Budi Luhur*, vol. 18, no. 2, 2025.
- [7] M. N. Khasanah, A. Harjoko, dan I. Candradewi, "Klasifikasi Sel Darah Putih Berdasarkan Ciri Warna dan Bentuk dengan Metode K-Nearest Neighbor (K-NN)," *IJEIS*, vol. 6, no. 2, hlm. 151–162, 2016.
- [8] A. N. Liyantoko, I. Candradewi, dan A. Harjoko, "Klasifikasi Sel Darah Putih dan Sel Limfoblas Menggunakan Metode Multilayer Perceptron Backpropagation," *IJEIS (Indonesian Journal of Electronics and Instrumentation Systems)*, vol. 9, no. 2, hlm. 173, Okt 2019, doi: 10.22146/ijeis.49943.
- [9] J. L. Diaz Resendiz, V. Ponomaryov, R. Reyes Reyes, dan S. Sadovnychiy, "Explainable CAD System for Classification of Acute Lymphoblastic Leukemia Based on a Robust White Blood Cell Segmentation," *Cancers (Basel)*, vol. 15, no. 13, hlm. 3376, Jun 2023, doi: 10.3390/cancers15133376.
- [10] M. Jawahar, L. J. Anbarasi, S. Narayanan, dan A. H. Gandomi, "An attention-based deep learning for acute lymphoblastic leukemia classification," *Sci. Rep.*, vol. 14, no. 1, Des 2024, doi: 10.1038/s41598-024-67826-9.
- [11] M. Bukhari, S. Yasmin, S. Sammad, dan A. A. Abd El-Latif, "A Deep Learning Framework for Leukemia Cancer Detection in Microscopic Blood Samples Using Squeeze and Excitation Learning," *Math. Probl. Eng.*, vol. 2022, hlm. 1–18, Jan 2022, doi: 10.1155/2022/2801227.
- [12] O. Islam, M. Assaduzzaman, dan M. Z. Hasan, "An explainable AI-based blood cell classification using optimized convolutional neural network," *J. Pathol. Inform.*, vol. 15, Des 2024, doi: 10.1016/j.jpi.2024.100389.

- [13] H. Harnelia, "Analisis Sentimen Review Skincare Skintific Dengan Algoritma Support Vector Machine (SVM)," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, Apr 2024, doi: 10.23960/jitet.v12i2.4095.
- [14] A. Gupta and R. Gupta, "ALL Challenge dataset of ISBI 2019," The Cancer Imaging Archive, 2019. [Online]. Available: <https://doi.org/10.7937/tcia.2019.dc64i46r>.
- [15] C. Matek, S. Schwarz, C. Marr, and K. Spiekermann, "A single-cell morphological dataset of leukocytes from AML patients and non-malignant controls," The Cancer Imaging Archive, 2019. [Online]. Available: <https://doi.org/10.7937/tcia.2019.36f5o9ld>.
- [16] H. Harnelia, S. B. Saudi, F. Agsaria, dan R. A. Saputra, "Analisis Kualitas dan Klasifikasi Jenis Tanah Berbasis Pengolahan Citra: Teknik Image Sharpening dan CNN ResNet untuk pemetaan pemanfaatan Daerah Pesisir," *Telematika*, vol. 22, no. 2, Nov 2025, doi: 10.31315/telematika.v22i2.14516.
- [17] F. Ramadhani, C. Paramita, dan E. R. Subhiyakto, "Disparitas Efektivitas CLAHE pada Berbagai Arsitektur Deep Learning untuk Klasifikasi Katarak Berbasis Citra Fundus," *Technology and Science (BITS)*, vol. 7, no. 3, hlm. 1785–1796, 2025, doi: 10.47065/bits.v7i3.8725.
- [18] M. R. Kusnaldi, T. Gulo, dan S. Aripin, "Penerapan Normalisasi Data Dalam Mengelompokkan Data Mahasiswa Dengan Menggunakan Metode K-Means Untuk Menentukan Prioritas Bantuan Uang Kuliah Tunggal," *Journal of Computer System and Informatics (JoSYC)*, vol. 3, no. 4, hlm. 330–338, Sep 2022, doi: 10.47065/josyc.v3i4.2112.
- [19] U. UNGKAWA dan G. AL HAKIM, "Klasifikasi Warna pada Kematangan Buah Kopi Kuning menggunakan Metode CNN Inception V3," *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, vol. 11, no. 3, hlm. 731, Jul 2023, doi: 10.26760/elkomika.v11i3.731.
- [20] M. Lehnert, J. Krejčí, M. Janura, dan M. De Ste Croix, "Age-Related Changes in Landing Mechanics in Elite Male Youth Soccer Players: A Longitudinal Study," *Applied Sciences*, vol. 12, no. 11, hlm. 5324, Mei 2022, doi: 10.3390/app12115324.
- [21] Dede Kurniadi, Rifky Muhammad Shidiq, dan Asri Mulyani, "Perbandingan Penggunaan Optimizer dalam Klasifikasi Sel Darah Putih Menggunakan Convolutional Neural Network," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 14, no. 1, hlm. 77–86, Mar 2025, doi: 10.22146/jnteti.v14i1.17162.
- [22] M. Reyad, A. M. Sarhan, dan M. Arafa, "A modified Adam algorithm for deep neural network optimization," *Neural Comput. Appl.*, vol. 35, no. 23, hlm. 17095–17112, Agu 2023, doi: 10.1007/s00521-023-08568-z.
- [23] F. Muhammad dkk., "Deteksi Anomali Tekstural pada Gambar Generatif AI menggunakan Convolutional Neural Network (CNN) dengan Arsitektur EfficientNetB2 under a Creative Commons Attribution-NonCommercial ShareAlike 4.0 International (CC BY-NC-SA 4.0)," *Jurnal Instek Informasi Sains dan Teknologi*, vol. 11, no. 1, 2026, [Daring]. Tersedia pada: <https://journal.uin-alauddin.ac.id/index.php/instek>
- [24] A. Acevedo, A. Merino, S. Alférez, A. Molina, L. Boldú, dan J. Rodellar, "A dataset of microscopic peripheral blood cell images for development of automatic recognition systems," *Data Brief*, vol. 30, Jun 2020, doi: 10.1016/j.dib.2020.105474.
- [25] C. Marr, "Expert-annotated datasets for deep-learning based classification methods for morphologic Leukaemia diagnostics Christian Matek," *Short Paper – MIDL 2024 submission*, vol. 30, Jun 2020, doi: 10.1016/j.dib.2020.105474.
- [26] B. Elsayed dkk., "Deep learning enhances acute lymphoblastic leukemia diagnosis and classification using bone marrow images," *Front. Oncol.*, vol. 13, Des 2023, doi: 10.3389/fonc.2023.1330977.
- [27] M. Deutges, A. Sadafi, N. Navab, dan C. Marr, "Neural Cellular Automata for Lightweight, Robust and Explainable Classification of White Blood Cell Images," *MICCAI (Medical Image Computing and Computer Assisted Intervention)*, 2024, [Daring]. Tersedia pada: <https://github.com/marrlab/WBC-NCA>.