

**IMPLEMENTASI EFFICIENTNETV2M DENGAN WEIGHTED
ENSEMBLE NETWORKS UNTUK KLASIFIKASI CITRA
ACUTE LYMPHOBLASTIC LEUKEMIA**

PROPOSAL SKRIPSI



Oleh:

Glendy Hernandez Putra Mahardika Gunantoro

220411100076

Dosen Pembimbing 1 : Prof. Dr. Arif Muntasa, S.Si., M.T.

Dosen Pembimbing 2 : Husni, S.Kom., M.T.

PROGRAM STUDI S1 TEKNIK INFORMATIKA

JURUSAN TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS TRUNOJOYO MADURA

BANGKALAN

2025

ABSTRAK

Diagnosis Acute Lymphoblastic Leukemia (ALL) secara konvensional melalui pemeriksaan mikroskopis oleh ahli patologi merupakan proses yang memakan waktu, sangat bergantung pada keahlian, dan rentan terhadap subjektivitas antar pengamat. Seiring perkembangan teknologi, deep learning menawarkan solusi menjanjikan untuk diagnosis berbantuan komputer, namun kebanyakan penelitian membahas performa arsitektur tunggal dan hanya sedikit yang membahas tentang *ensemble learning*. Penelitian ini mengusulkan pendekatan dua tahap: pertama, investigasi sistematis terhadap tiga skenario LR (0.001, 0.0001, dan 0.00001) untuk melatih model EfficientNetV2-M; kedua, implementasi metode Weighted Ensemble of Networks (WEN). Kebaruan utama penelitian ini terletak pada investigasi dua skenario WEN: WEN Inter-LR (menggabungkan model-model dari 5-Fold di dalam satu skenario LR) dan WEN Intra-LR (menggabungkan model-model "juara" antar skenario LR). Untuk setiap skenario WEN, penelitian ini akan menguji dan membandingkan tiga metrik berbeda sebagai dasar penentuan bobot: Akurasi, F1-Score, dan Cohen's Kappa, yang dieksplorasi karena keandalannya menangani ketidakseimbangan kelas. Data yang digunakan adalah dataset publik C-NMC 2019, dan seluruh model dilatih menggunakan 5-Fold Cross Validation untuk evaluasi yang robust. Penelitian ini berfokus pada analisis komparatif untuk mengidentifikasi LR optimal serta skenario ensemble dan skema pembobotan WEN yang diharapkan mampu menghasilkan performa klasifikasi ALL paling optimal, melampaui kinerja model tunggal terbaik.

Kata Kunci: *Acute Lymphoblastic Leukemia, EfficientNetV2, Weighted Ensemble of Networks, Klasifikasi Citra*

DAFTAR ISI

ABSTRAK	i
DAFTAR ISI	ii
DAFTAR GAMBAR	iv
DAFTAR TABLE	v
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.2.1 Permasalahan.....	5
1.2.2 Solusi Permasalahan	5
1.2.3 Pertanyaan Penelitian	6
1.3 Tujuan dan Manfaat	7
1.4 Batasan Masalah	8
1.5 Metode Usulan.....	9
1.6 Sistematika Penulisan.....	11
BAB II KAJIAN PUSTAKA	13
2.1 Penelitian Terdahulu	13
2.2 Acute Lymphoblastic Leukemia (ALL).....	17
2.3 Klasifikasi	18
2.4 Convolutional Neural Network (CNN)	18
2.4.1 Convolution Layer.....	19
2.4.2 Non-Linear Activation Functions (Swish)	21
2.4.3 Pooling Layer	22
2.4.4 Fully Connected Layer	23
2.5 EfficientNet.....	26
2.6 Optimizer	33
2.7 Area Under Curve.....	34

2.7.1	Kurva Receiver Operating Characteristic (ROC)	35
2.8	Weighted Ensemble of Network	36
2.8.1	Pembobotan Berdasarkan F1-Score.....	38
2.8.2	Pembobotan Berdasarkan Akurasi.....	38
2.8.3	Pembobotan Berdasarkan Cohen's Kappa	38
2.9	Cross Validation.....	39
2.10	Evaluasi Kinerja Sistem.....	40
2.10.1	Confusion Matrix.....	41
2.10.2	Accuracy	41
2.10.3	Precision	42
2.10.4	Recall	42
2.10.5	F1-Score.....	42
2.10.6	Cohen's Kappa	43
2.11	Learning Rate	44
2.12	Class Weight	45
BAB III METODE USULAN.....		48
3.1	Dataset	48
3.2	Rancangan Arsitektur Penelitian	50
3.3	Rancangan Sistem Pelatihan Model.....	51
3.4	Rancangan Weighted Ensemble of Networks (WEN).....	54
3.5	Skenario Uji Coba	57
DAFTAR PUSTAKA		60

DAFTAR GAMBAR

Gambar 2. 1 Convolutional Neural Network.....	19
Gambar 2. 2 Proses Konvolusi	19
Gambar 2. 3 Grafik fungsi aktivasi swish	22
Gambar 2. 4 Pooling Layer.....	23
Gambar 2. 5 Fully Connected Layer	24
Gambar 2. 6 Grafik penskalaan depth dan resolution terhadap akurasi dan kompleksitas.....	29
Gambar 2. 7 Arsitektur EfficientNet.....	31
Gambar 2. 8 Grafik Performa EfficientNet	32
Gambar 2. 9 K-Fold Cross Validation	40
Gambar 3. 1 File Gambar ALL.....	49
Gambar 3. 2 Rancangan Arsitektur Penelitian.....	50
Gambar 3. 3 Rancangan Sistem Pelatihan Model	52
Gambar 3. 4 Rancangan Weighted Ensemble of Networks (Intra-LR)	56

DAFTAR TABLE

Table 2. 1 Penelitian Terkait	16
Table 2. 2 Confusion Matrix.....	41
Table 3. 1 Komposisi Struktur Asli Dataset	48
Table 3. 2 Skenario Uji Coba Pelatihan.....	58
Table 3. 3 Skenario Emsemble Pembobotan	58
Table 3. 1 Komposisi Struktur Asli Dataset	42
Table 3. 2 Skenario Uji Coba Pelatihan.....	51
Table 3. 3 Skenario Emsemble Pembobotan	51

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Kanker darah merupakan salah satu tantangan terbesar dalam dunia medis global, dengan *Acute Lymphoblastic Leukemia* (ALL) menjadi jenis kanker darah yang paling umum didiagnosis pada anak-anak. Patologi penyakit ini ditandai oleh proliferasi abnormal, atau pertumbuhan tak terkendali, dari sel limfoblas (sel darah putih yang belum matang) di dalam sumsum tulang[1]. Akumulasi sel-sel abnormal ini mendesak sel-sel darah sehat—termasuk sel darah merah, trombosit, dan sel darah putih matang—yang pada akhirnya dapat menyebar secara progresif ke organ vital lain seperti hati, limpa, dan sistem saraf pusat. Diagnosis dini menjadi faktor yang sangat krusial; keberhasilan terapi dan angka harapan hidup pasien berkorelasi langsung dengan seberapa cepat penyakit ini terdeteksi dan ditangani[2].

Meskipun urgensinya sangat tinggi, standar untuk diagnosis ALL secara konvensional masih mengandalkan metode yang padat karya dan memiliki keterbatasan signifikan. Metode ini adalah pemeriksaan mikroskopis apusan darah tepi atau sumsum tulang oleh ahli patologi. Proses ini secara inheren sangat bergantung pada faktor manusia: keahlian, pengalaman, dan bahkan kondisi fisik serta kelelahan (*fatigue*) pemeriksa. Ketergantungan ini melahirkan risiko subjektivitas dan variabilitas hasil antar-pengamat (*inter-observer variability*)[3]. Selain itu, proses manual ini memakan waktu yang berharga. Dalam skenario di mana laboratorium menerima ratusan sampel, keterbatasan ini dapat menyebabkan keterlambatan diagnosis. Keterbatasan inilah yang mendorong perlunya inovasi teknologi untuk mengembangkan sistem diagnosis yang lebih objektif, konsisten, cepat, dan akurat.

Perkembangan pesat dalam bidang kecerdasan buatan (Artificial Intelligence), khususnya deep learning, menawarkan solusi yang sangat menjanjikan untuk mengatasi tantangan tersebut. Sistem diagnosis berbantuan komputer (*Computer-Aided Diagnosis - CAD*) yang ditenagai oleh *deep learning* telah menunjukkan kapasitas yang luar biasa dalam menganalisis citra medis[4]. Arsitektur Convolutional

Neural Network (CNN) secara khusus telah terbukti sangat efektif dalam tugas klasifikasi citra medis, karena kemampuannya untuk secara otomatis mempelajari fitur-fitur hierarkis dari data visual, dari tepi sederhana hingga pola selular yang kompleks, tanpa perlu rekayasa fitur (*feature engineering*) manual[5].

Dalam beberapa tahun terakhir, berbagai penelitian telah berupaya menerapkan beragam arsitektur CNN untuk klasifikasi citra ALL. Banyak peneliti fokus pada penggunaan model tunggal yang dioptimalkan. Sebagai contoh, penelitian oleh Amreen Batool[6], menunjukkan hasil yang sangat impresif dengan menggunakan arsitektur EfficientNet-B3, yang dimodifikasi dengan depthwise separable convolutions untuk efisiensi, dan berhasil mencapai akurasi 99.31% pada dataset C-NMC 2019. Demikian pula, Muhammad Zakir Ullah[7], mengintegrasikan modul Efficient Channel Attention (ECA) ke dalam arsitektur VGG16 untuk membantu model fokus pada fitur relevan, dan berhasil mencapai akurasi rata-rata 91.1%.

Meskipun model tunggal ini menunjukkan akurasi yang tinggi, ketergantungan pada satu arsitektur memiliki kelemahan, yaitu rentan terhadap overfitting dan bias terhadap set fitur spesifik yang dipelajari oleh arsitektur tersebut. Untuk mengatasi ini, pendekatan lain adalah dengan melakukan optimasi hyperparameter secara ekstensif atau memodifikasi arsitektur dasarnya. Penelitian oleh Kevyn Alifian Hernanda Wibowo[5], membuktikan hal ini dengan menerapkan Optimasi Bayesian pada EfficientNetV2M, yang berhasil meningkatkan akurasi dari 87.61% menjadi 91.37%. Pendekatan yang lebih radikal dilakukan oleh Arif Muntasar[8], yang memodifikasi EfficientNetV2-S dengan Hybrid Kernel dan Enkapsulasi Jaringan, dan berhasil mencapai F1-Score 97.26%.

Pendekatan-pendekatan tersebut, meskipun sangat kuat, menunjukkan bahwa performa model tunggal masih dapat ditingkatkan, seringkali dengan menambah kompleksitas pada arsitektur atau proses tuning. Pendekatan ketiga, yang juga sangat kompleks, adalah metode hibrida yang menggabungkan *feature fusion* dari beberapa model CNN sebelum dimasukkan ke classifier machine learning tradisional. Ibrahim Abdulrab Ahmed[9], menerapkan ini dengan menggabungkan fitur dari DenseNet121,

ResNet50, dan MobileNet, yang kemudian diklasifikasikan oleh Random Forest, dan berhasil mencapai akurasi 98.8% pada dataset C-NMC 2019. Kelemahan dari metode fusi fitur semacam ini adalah kompleksitas dalam implementasi dan tingginya kebutuhan komputasi untuk mengekstraksi fitur dari berbagai model secara bersamaan.

Keterbatasan dan kompleksitas dari berbagai pendekatan ini—baik mengandalkan model tunggal yang dioptimalkan maupun fusi fitur yang rumit—menegaskan sebuah fakta: mengandalkan satu model tunggal, terlepas dari seberapa canggih arsitekturnya, seringkali belum cukup untuk menangkap seluruh kompleksitas dan variasi fitur pada citra sel leukemia yang heterogen[10]. Setiap arsitektur (seperti VGG, ResNet, atau EfficientNet) memiliki bias dan cara pandangnya sendiri dalam mengekstraksi pola. Di sinilah filosofi ensemble learning menawarkan alternatif yang elegan dan kuat.

Teknik ensemble learning digunakan untuk menggabungkan prediksi dari beberapa model, sehingga menghasilkan keputusan akhir yang lebih robust dan akurat. Daripada bertaruh pada satu ahli saja, ensemble learning bertindak seperti panel ahli di mana keputusan kolektif lebih dipercaya[11]. Salah satu pendekatan ensemble yang efektif adalah Weighted Ensemble of Networks (WEN), yang memberikan bobot kontribusi yang berbeda untuk setiap model berdasarkan kinerjanya. Dengan cara ini, model yang lebih unggul akan memiliki pengaruh lebih besar terhadap hasil klasifikasi akhir[12].

Penelitian ini mengusulkan implementasi arsitektur EfficientNetV2 sebagai base model. Arsitektur ini dipilih karena merupakan evolusi dari EfficientNet yang revolusioner, yang tidak hanya berfokus pada akurasi tetapi juga pada efisiensi komputasi [5]. EfficientNetV2 menggunakan compound scaling untuk menyeimbangkan kedalaman, lebar, dan resolusi jaringan secara cerdas, serta memanfaatkan blok MBConv yang efisien. Secara spesifik, penelitian ini akan berfokus pada satu varian, yaitu EfficientNetV2-M (*Medium*), yang dipilih karena menawarkan keseimbangan optimal antara kedalaman arsitektur dan efisiensi komputasi.

Namun, untuk mencapai performa puncak dari arsitektur tersebut, optimasi hyperparameter menjadi aspek yang krusial. Arsitektur yang kuat saja tidak cukup; performa model deep learning sangat bergantung pada konfigurasi pelatihannya. Salah satu tantangan utama dalam pelatihan adalah overfitting (di mana model terlalu menghafal data latih dan gagal pada data baru), yang pengelolaannya sangat dipengaruhi oleh pemilihan hyperparameter yang tepat.

Selanjutnya, penelitian ini akan melakukan investigasi sistematis terhadap hyperparameter yang paling berpengaruh, yaitu *Learning Rate* (LR). Model EfficientNetV2-M akan dilatih dan dievaluasi dalam tiga skenario LR yang berbeda: 0.001, 0.0001, dan 0.00001, untuk menemukan laju pembelajaran yang paling optimal.

Setelah mengeksplorasi performa model tunggal, penelitian ini akan melangkah lebih jauh dengan menerapkan *Weighted Ensemble of Networks* (WEN). Di sinilah letak kebaruan (novelty) utama dari penelitian ini. Berbeda dari penelitian sebelumnya yang menggabungkan arsitektur berbeda[12], penelitian ini mengusulkan strategi ensemble yang didasarkan pada hasil optimasi *hyperparameter*. Penelitian ini akan mengeksplorasi dua skenario ensemble utama: pertama, Ensemble Inter-LR, yang menggabungkan model-model dari 5-Fold Cross Validation yang dilatih di dalam satu skema LR yang sama (misalnya, 5 model dari LR 0.001 digabungkan); dan kedua, Ensemble Intra-LR, yang menggabungkan model-model terbaik yang mewakili setiap skema LR (model terbaik dari LR 0.001, 0.0001, dan 0.00001 digabungkan).

Pendekatan WEN ini, yang divalidasi oleh Chayan Mondal [12], akan diinvestigasi lebih lanjut. Temuan Mondal akan dijadikan titik awal untuk menguji kembali skema pembobotan mana yang paling efektif. Oleh karena itu, kebaruan penelitian ini adalah menerapkan dan memvalidasi strategi pembobotan WEN (Akurasi, F1-Score, dan Cohen's Kappa) secara spesifik pada ensemble model EfficientNetV2-M yang dihasilkan dari skenario pelatihan *hyperparameter* (LR) yang berbeda.

Hipotesis penelitian ini adalah: (1) Akan terdapat perbedaan kinerja yang signifikan antara model tunggal yang dilatih dengan LR 0.001, 0.0001, dan 0.00001.

(2) Model Weighted Ensemble (baik Inter-LR maupun Intra-LR) akan memiliki performa yang lebih tinggi daripada performa model tunggal terbaik sekalipun. (3) Strategi ensemble Intra-LR berpotensi menjadi model paling unggul karena menggabungkan "keahlian" dari model-model yang dilatih dengan cara yang fundamental berbeda. (4) Akan terdapat perbedaan kinerja yang signifikan di antara ketiga skema pembobotan (Akurasi, F1, Kappa), yang akan mengidentifikasi strategi ensemble paling optimal.

Melalui penerapan metode ini pada dataset publik C-NMC 2019, penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam pengembangan sistem CAD untuk diagnosis leukemia. Hasil dari penelitian ini berpotensi meningkatkan akurasi, objektivitas, dan efisiensi diagnosis. Lebih jauh lagi, penelitian ini dapat menjadi referensi penting bagi pengembangan metode ensemble learning yang dikombinasikan dengan strategi optimasi hyperparameter untuk aplikasi klasifikasi citra medis lainnya.

1.2 Rumusan Masalah

1.2.1 Permasalahan

Bagaimana merancang dan mengimplementasikan model deep learning yang efektif dan optimal sebagai metode diagnosis Acute Lymphoblastic Leukemia (ALL) pada citra mikroskopis sel darah. Tantangan utamanya terletak pada penentuan konfigurasi hyperparameter yang tepat, seperti Learning Rate, untuk melatih arsitektur tunggal (EfficientNetV2-M) secara optimal. Selain itu, terdapat permasalahan lebih lanjut apakah penggabungan beberapa model melalui metode ensemble dapat memberikan peningkatan performa yang lebih signifikan dibandingkan model tunggal terbaik sekalipun.

1.2.2 Solusi Permasalahan

Solusi yang diusulkan dalam penelitian ini adalah mengimplementasikan sistem klasifikasi deep learning melalui pendekatan investigasi multi-tahap untuk menemukan konfigurasi paling optimal. Pendekatan ini terbagi dalam dua tahapan utama:

1. Pelatihan Model Tunggal dan Optimasi Hyperparameter: Tahap pertama adalah melatih dan mengevaluasi arsitektur Convolutional Neural Network (CNN) EfficientNetV2-M. Pada tahap ini, dilakukan investigasi sistematis terhadap hyperparameter *Learning Rate* (LR) dengan menguji tiga skenario berbeda (0.001, 0.0001, dan 0.00001) untuk menemukan performa model tunggal terbaik.
2. Implementasi Weighted Ensemble of Networks (WEN): Tahap kedua adalah menggabungkan model-model yang telah dilatih pada tahap pertama. Metode WEN akan diinvestigasi dalam dua skenario (Inter-LR dan Intra-LR) dan diuji menggunakan tiga skema pembobotan berbeda (berdasarkan Akurasi, F1-Score, dan Cohen's Kappa) untuk menentukan strategi ensemble yang paling unggul.

Dengan membandingkan performa dari berbagai skenario hyperparameter (LR) pada model tunggal dan berbagai skenario pada model ensemble (WEN), penelitian ini bertujuan mengidentifikasi arsitektur dan strategi pelatihan yang paling akurat serta andal untuk diagnosis ALL.

1.2.3 Pertanyaan Penelitian

Berdasarkan permasalahan yang telah dijelaskan di atas, dapat dirumuskan pertanyaan dalam penelitian sebagai berikut:

1. Bagaimana pengaruh perbedaan Learning Rate (0.001, 0.0001, dan 0.00001) terhadap performa arsitektur tunggal EfficientNetV2-M dalam melakukan klasifikasi citra Acute Lymphoblastic Leukemia?
2. Bagaimana performa dari penerapan metode Weighted Ensemble of Networks (WEN), baik skenario Inter-LR (gabungan 5-Fold dalam satu LR) maupun skenario Intra-LR (gabungan model terbaik antar LR) dalam klasifikasi citra ALL?
3. Manakah di antara skema pembobotan berdasarkan metrik Akurasi, F1-Score, dan Cohen's Kappa yang menghasilkan performa paling optimal pada model

Weighted Ensemble of Networks (Inter-LR dan Intra-LR) untuk kasus klasifikasi citra ALL?

1.3 Tujuan dan Manfaat

A. Tujuan Penelitian

Sejalan dengan rumusan masalah yang telah dipaparkan, tujuan dari penelitian ini adalah sebagai berikut:

1. Mengevaluasi dan membandingkan performa arsitektur tunggal EfficientNetV2-M yang dilatih menggunakan tiga skema *Learning Rate* (0.001, 0.0001, dan 0.00001) untuk klasifikasi citra *Acute Lymphoblastic Leukemia*.
2. Menganalisis performa yang dihasilkan oleh metode *Weighted Ensemble of Networks* (WEN), baik skenario Inter-LR (gabungan 5-Fold dalam satu LR) maupun skenario Intra-LR (gabungan model terbaik antar LR), dibandingkan dengan performa arsitektur tunggal terbaik.
3. Mengidentifikasi skema pembobotan yang paling optimal untuk model *Weighted Ensemble of Networks* (Inter-LR dan Intra-LR) dengan membandingkan penggunaan metrik Akurasi, F1-Score, dan Cohen's Kappa sebagai dasar penentuan bobot.

B. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Bagi Bidang Medis: Memberikan kontribusi terhadap pengembangan sistem diagnosis komputer yang akurat dan objektif untuk *Acute Lymphoblastic Leukemia*, sehingga dapat membantu ahli patologi dalam mempercepat dan meningkatkan keandalan proses diagnosis.
2. Bagi Ilmu Pengetahuan: Memperkaya khazanah penelitian di bidang deep learning, khususnya terkait penerapan arsitektur EfficientNetV2-M dan

efektivitas metode *Weighted Ensemble of Networks* dengan berbagai skema pembobotan untuk klasifikasi citra medis.

3. Bagi Peneliti Selanjutnya: Hasil penelitian ini dapat menjadi referensi yang kuat dan landasan bagi pengembangan metode deteksi penyakit berbasis citra medis yang lebih lanjut, baik dalam domain hematologi maupun bidang medis lainnya.
4. Bagi Penulis: Meningkatkan kompetensi dan pengalaman praktis dalam merancang, mengimplementasikan, dan mengevaluasi sistem cerdas berbasis deep learning untuk menyelesaikan permasalahan nyata di sektor kesehatan.

1.4 Batasan Masalah

Agar penelitian ini lebih terarah dan fokus pada tujuan yang telah ditetapkan, maka diberikan batasan-batasan masalah sebagai berikut:

1. Objek penelitian ini terbatas pada klasifikasi citra untuk penyakit *Acute Lymphoblastic Leukemia* (ALL). Jenis leukemia lain atau subtipe dari ALL tidak termasuk dalam lingkup kajian penelitian ini.
2. Dataset yang digunakan dalam penelitian adalah dataset publik C-NMC 2019. Penelitian ini tidak melakukan proses pengumpulan atau pengambilan data primer dari rumah sakit maupun laboratorium.
3. Arsitektur model yang digunakan dibatasi pada satu varian dari EfficientNetV2, yaitu Medium (M).
4. Metode ensemble yang diterapkan terbatas pada *Weighted Ensemble of Networks* (WEN), dengan skema pembobotan berdasarkan hasil validasi metrik Akurasi, F1-Score, dan Cohen's Kappa.
5. Proses pelatihan model menggunakan pendekatan transfer learning dengan memanfaatkan bobot pra-terlatih (*pre-trained weights*) dari dataset ImageNet.

6. Evaluasi performa model dibatasi pada metrik kuantitatif yang mencakup akurasi, presisi, recall, F1-score, Cohen's Kappa, serta visualisasi kurva *Receiver Operating Characteristic* (ROC) dan nilai *Area Under Curve* (AUC). Aspek lain seperti kecepatan komputasi dan penggunaan memori tidak dievaluasi.
7. Penelitian ini berfokus pada domain teknis klasifikasi citra dan tidak membahas aspek klinis yang lebih dalam, seperti penentuan stadium penyakit atau prognosis pasien. Tahapan implementasi atau deployment model ke dalam sistem klinis juga tidak termasuk dalam cakupan penelitian.
8. Eksperimen hyperparameter terbatas pada investigasi tiga skema *Learning Rate* (0.001, 0.0001, dan 0.00001). *Hyperparameter* lain seperti optimizer (Adam), jumlah epoch, dan batch size diatur sebagai nilai konstan selama proses pelatihan.

1.5 Metode Usulan

Penelitian ini dilaksanakan secara sistematis melalui beberapa tahapan utama yang dirancang untuk mencapai tujuan yang telah ditetapkan. Adapun tahapan metodologi penelitian yang dilakukan adalah sebagai berikut:

1. Studi Literatur

Tahap awal adalah melakukan kajian mendalam terhadap literatur dan penelitian terdahulu yang relevan. Fokus studi mencakup teori klasifikasi citra medis, arsitektur deep learning khususnya EfficientNetV2, konsep ensemble learning, dan implementasi metode Weighted Ensemble of Networks (WEN). Tahap ini bertujuan untuk membangun landasan teoretis yang kuat dan mengidentifikasi posisi penelitian ini terhadap riset sebelumnya.

2. Pengumpulan dan Persiapan Data

Dataset yang digunakan adalah C-NMC 2019 yang berisi citra sel darah putih. Data tersebut kemudian melalui tahap persiapan (preprocessing) yang meliputi

proses normalisasi nilai piksel dan augmentasi data untuk meningkatkan variasi dan mencegah overfitting. Selanjutnya, dataset dibagi menjadi tiga bagian: data latih, data validasi, dan data uji.

3. Perancangan dan Pelatihan Model Tunggal

Merancang dan mengimplementasikan model klasifikasi menggunakan arsitektur EfficientNetV2-M. Pelatihan model akan dilakukan dalam tiga skenario hyperparameter yang berbeda berdasarkan *Learning Rate* (0.001, 0.0001, dan 0.00001). Setiap skenario model akan dilatih menggunakan pendekatan transfer learning dengan bobot pra-terlatih dari ImageNet, melalui proses fine-tuning untuk adaptasi pada tugas klasifikasi biner (*Acute Lymphoblastic Leukemia* vs. Normal)

4. Implementasi Weighted Ensemble of Networks (WEN)

Mengimplementasikan dua skenario ensemble menggunakan metode WEN. Skenario pertama adalah WEN Inter-LR, yang menggabungkan prediksi dari 5 model hasil cross-validation yang dilatih dalam satu skenario *Learning Rate* yang sama. Skenario kedua adalah WEN Intra-LR, yang menggabungkan prediksi dari tiga model terbaik (satu model terbaik mewakili setiap skenario *Learning Rate*). Pada tahap ini, dilakukan perhitungan bobot untuk setiap model berdasarkan performa individunya (Akurasi, F1-Score, dan Cohen's Kappa) yang diukur pada data evaluasi.

5. Pengujian dan Evaluasi Model

Melakukan pengujian terhadap model-model tunggal (hasil dari 3 skenario LR) dan model-model ensemble (Inter-LR dan Intra-LR) menggunakan data uji yang belum pernah dilihat sebelumnya. Kinerja model dievaluasi secara kuantitatif menggunakan metrik akurasi, presisi, recall, F1-score, Cohen's Kappa, serta visualisasi kurva Receiver Operating Characteristic (ROC) dan nilai Area Under Curve (AUC) untuk membandingkan efektivitas antara model

tunggal dan ensemble, serta untuk menentukan skema pembobotan WEN yang paling optimal

6. Analisis dan Penarikan Kesimpulan

Menganalisis hasil evaluasi secara komprehensif untuk menjawab rumusan masalah. Hasil analisis kemudian disajikan, dan ditarik kesimpulan mengenai performa model serta efektivitas metode yang diusulkan.

7. Penyusunan Laporan

Mendokumentasikan seluruh proses, hasil, dan analisis penelitian secara sistematis ke dalam format laporan skripsi, yang diakhiri dengan simpulan dan saran untuk penelitian di masa mendatang.

1.6 Sistematika Penulisan

Proposal penelitian ini disusun dengan sistematika penulisan yang terdiri dari tiga bab utama, yaitu sebagai berikut:

Bab I : Pendahuluan

Bab ini merupakan bagian pengantar yang menguraikan konteks penelitian. Isinya mencakup latar belakang masalah, rumusan masalah, tujuan dan manfaat yang ingin dicapai, batasan masalah, gambaran umum metodologi penelitian, serta sistematika penulisan proposal.

Bab II : Tinajuan Pustaka

Bab ini menyajikan landasan teoretis yang menjadi dasar penelitian. Pembahasan di dalamnya meliputi teori mengenai *Acute Lymphoblastic Leukemia* (ALL), pengolahan citra, *deep learning*, arsitektur EfficientNetV2, dan konsep *ensemble learning* dengan metode *Weighted Ensemble of Networks* (WEN). Selain itu, bab ini juga mengulas penelitian-penelitian relevan sebelumnya sebagai referensi.

Bab III : Metode Usulan

Bab ini menjelaskan secara rinci rancangan dan langkah-langkah yang akan dilakukan untuk menjawab rumusan masalah. Uraian di dalamnya mencakup deskripsi dataset yang akan digunakan, tahapan persiapan data (*preprocessing*), perancangan arsitektur model, skenario pelatihan dan pengujian, serta metrik evaluasi yang akan digunakan untuk mengukur kinerja model

BAB III

METODE USULAN

3.1 Dataset

Dataset yang digunakan dalam penelitian ini adalah C-NMC 2019 Dataset, yang merupakan bagian dari ALL Challenge – ISBI 2019 dan dipublikasikan melalui The Cancer Imaging Archive (TCIA) Public Access serta tersedia juga pada platform Kaggle. Dataset ini dirancang secara khusus untuk tujuan klasifikasi sel darah putih pada pasien dengan Acute Lymphoblastic Leukemia (ALL) terhadap sel darah normal (Healthy). Seluruh data berupa citra mikroskopis apusan darah yang telah melalui proses segmentasi sel secara manual, sehingga tiap citra berfokus pada satu sel tunggal.

Dataset ini pertama kali digunakan dalam IEEE ISBI 2019 Challenge: Classification of Normal vs Malignant Cells in B-ALL White Blood Cancer Microscopic Images, yang dapat diakses melalui laman resmi ISBI 2019 Challenge dan CodaLab Competition melalui link <https://www.cancerimagingarchive.net/collection/c-nmc-2019/>. Dataset ini juga dipublikasikan ulang di Kaggle melalui tautan: <https://www.kaggle.com/datasets/andrewmvd/leukemia-classification>.

Secara keseluruhan, dataset C-NMC 2019 terdiri atas 15.114 citra sel, yang berasal dari 118 individu subjek dengan distribusi 91 pasien ALL dan 72 pasien normal. Akan tetapi, dalam penelitian ini hanya digunakan 12.528 citra, sementara sebagian lainnya dialokasikan khusus untuk keperluan evaluasi kompetisi di CodaLab. Komposisi dataset berdasarkan kelas dan pembagiannya dapat dilihat pada table berikut:

Table 3. 1 Komposisi Struktur Asli Dataset

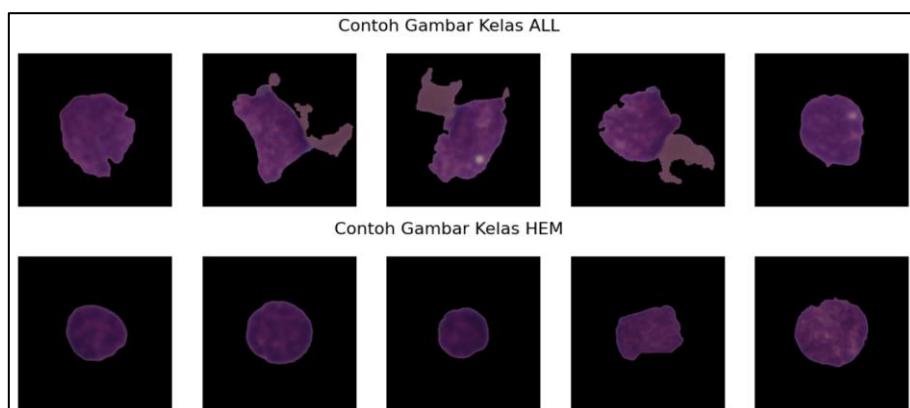
Komposisi	Sel ALL	Sel Normal	Pasien ALL	Pasien Normal
Train set composition	7272	3389	69	49

Preliminary test set composition	1219	648	13	15
Final test set composition		2586	9	8

Setiap citra disimpan dalam format BMP dengan ukuran resolusi 450×450 piksel. Label ground truth ditentukan oleh pakar onkologi sehingga memiliki validitas yang tinggi. Penamaan file mengikuti konvensi tertentu, misalnya UID_P_N_C_all untuk sel ALL (kanker) dan UID_HS_N_C_hem untuk sel normal (healthy).

Karakteristik dari dataset ini menunjukkan adanya ketidakseimbangan kelas (imbalanced dataset), karena jumlah citra sel ALL lebih banyak dibanding citra sel normal. Kondisi ini perlu ditangani dengan teknik penyeimbang data.

Gambar 3. 1 File Gambar ALL

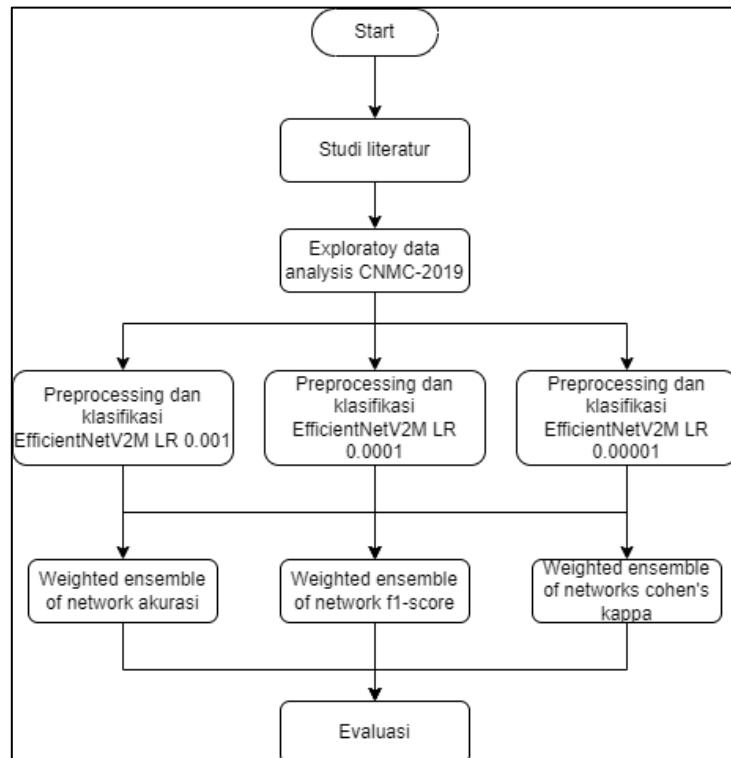


Secara morfologis dapat dilihat pada gambar diatas, sel ALL dan sel normal memiliki kemiripan yang tinggi sehingga sulit dibedakan hanya dengan pengamatan visual. Analisis laboratorium menunjukkan bahwa sel ALL cenderung memiliki ukuran bervariasi, bentuk inti sel tidak beraturan, serta terdapat nucleolus yang menonjol. Sitoplasma pada sel ALL sering kali memiliki volume besar dengan vakuola yang jelas terlihat. Perbedaan lebih nyata dapat ditemukan ketika dilakukan pewarnaan Romanowsky atau Giemsa, di mana sel ALL menampilkan warna limfoblas lebih gelap dibandingkan sel limfosit normal.

3.2 Rancangan Arsitektur Penelitian

Rancangan sistem pada penelitian ini dirancang secara sistematis untuk menjawab rumusan masalah yang telah ditetapkan. Bagian ini menjelaskan arsitektur penelitian secara keseluruhan (high-level) dari tahap awal hingga tahap evaluasi akhir. Alur kerja penelitian secara umum diilustrasikan pada Gambar 3.2.

Gambar 3. 2 Rancangan Arsitektur Penelitian



Penelitian dimulai dengan dua tahap fundamental. Pertama adalah Studi Literatur, yang bertujuan untuk mengumpulkan pengetahuan mengenai penelitian terdahulu yang sudah dilakukan dibidang Acute Lymphoblastic Leukemia serta menganalisis penggunaan metode dan hasil yang didapatkan. Tahap kedua adalah Exploratory Data Analysis (EDA) pada dataset C-NMC 2019 untuk memahami karakteristik, distribusi kelas, dan kualitas data yang akan digunakan.

Setelah tahap persiapan data selesai, penelitian masuk ke tahap inti, yaitu eksperimen pelatihan model. Seperti ditunjukkan pada Gambar 3.2, penelitian akan

menjalankan tiga alur eksperimen klasifikasi secara paralel dan independen. Ketiga alur ini menggunakan arsitektur dasar yang sama, yaitu EfficientNetV2-M, dan metodologi pelatihan yang identik (termasuk preprocessing, batch size, epoch, stage pelatihan, dan 5-Fold CV).

Perbedaan fundamental di antara ketiga alur tersebut adalah hyperparameter Learning Rate (LR) yang digunakan: 0.001, 0.0001, dan 0.00001. Setiap alur ("Preprocessing dan klasifikasi") ini akan menghasilkan serangkaian model yang telah dilatih. Proses teknis yang mendetail untuk setiap alur pelatihan paralel ini akan dijabarkan lebih lanjut pada Sub-bab 3.3.

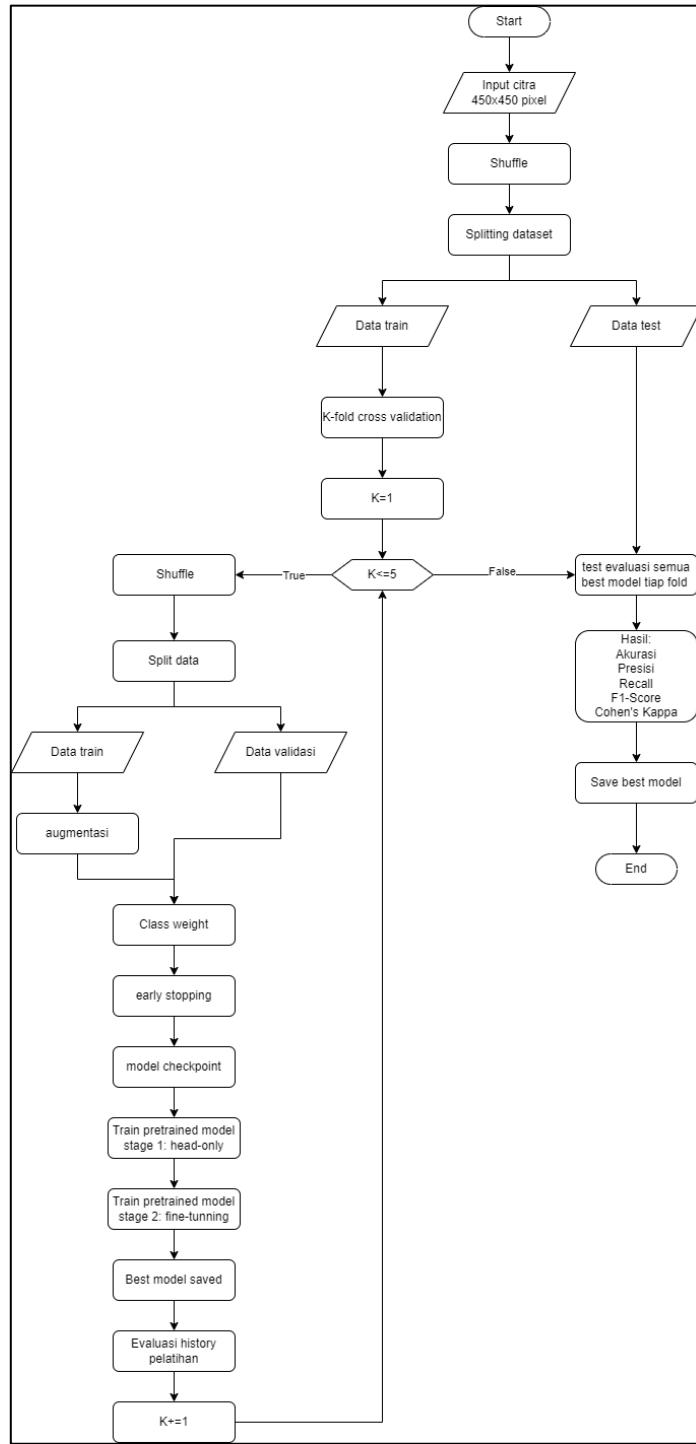
Setelah ketiga alur pelatihan menghasilkan model-model terbaiknya, tahap selanjutnya adalah pengujian Weighted Ensemble of Networks (WEN). Pada tahap ini, model-model yang telah dilatih akan digabungkan untuk menguji apakah performa gabungan dapat melampaui performa model tunggal terbaik. Seperti yang ditunjukkan pada alur, pengujian WEN akan diinvestigasi menggunakan tiga skema pembobotan yang berbeda, yaitu berdasarkan Akurasi, F1-Score, dan Cohen's Kappa. Rancangan detail untuk implementasi tahap WEN ini akan dijelaskan lebih lanjut pada Sub-bab 3.4.

Tahap terakhir dari arsitektur penelitian adalah Evaluasi. Pada tahap ini, seluruh hasil performa—baik dari model-model tunggal (hasil perbandingan 3 skenario LR) maupun dari berbagai skenario WEN—akan dianalisis dan dibandingkan secara komprehensif menggunakan data uji. Tujuannya adalah untuk mengidentifikasi hyperparameter dan strategi ensemble yang paling optimal untuk menjawab rumusan masalah.

3.3 Rancangan Sistem Pelatihan Model

Bagian ini merinci arsitektur teknis dan alur proses untuk melatih model klasifikasi. Proses yang dijelaskan di sini berlaku untuk setiap alur eksperimen Learning Rate (0.001, 0.0001, dan 0.00001) yang telah dipaparkan pada sub-bab sebelumnya. Rancangan sistem pelatihan model dapat dilihat pada Gambar 3.3.

Gambar 3. 3 Rancangan Sistem Pelatihan Model



Proses pelatihan diawali dengan Input Data Citra. Dataset yang digunakan adalah C-NMC 2019, berisi citra mikroskopis sel darah putih berukuran 450x450 piksel. Data ini kemudian melalui tahap Preprocessing Data yang meliputi:

- Resize Citra: Mempertahankan ukuran asli 450x450 piksel untuk menjaga detail citra.
- Normalisasi: Dilakukan secara otomatis melalui fungsi preprocessing bawaan EfficientNetV2, yang mengubah nilai piksel dari rentang 0-255 menjadi 0-1.
- Batching dan Prefetching: Menyiapkan data dalam batch berukuran 64 untuk mempercepat proses pelatihan.

Selanjutnya, dilakukan Shuffle dan Splitting Dataset. Data diacak lalu dibagi menjadi dua bagian: 80% data latih (training data) dan 20% data uji (testing data). Data uji ini akan disimpan dan tidak akan pernah digunakan selama proses pelatihan untuk memastikan evaluasi akhir yang objektif.

Inti dari metodologi pelatihan ini adalah penerapan 5-Fold Cross Validation ($K=5$) pada 80% data latih tersebut¹³. Seperti yang ditunjukkan pada Gambar 3.3, data latih akan masuk ke dalam loop validasi silang sebanyak 5 kali ($K=1$ hingga $K=5$). Pada setiap iterasi fold:

1. Shuffle dan Split: Data latih (80% total data) diacak dan dibagi lagi menjadi dua set: data latih baru (4 fold) dan data validasi (1 fold).
2. Augmentasi: Penerapan transformasi acak seperti flip horizontal, flip vertikal, perubahan kecerahan, dan kontras diterapkan hanya pada data latih baru untuk meningkatkan variasi data dan mencegah overfitting.
3. Penanganan Imbalance: Teknik Class Weight diterapkan. Teknik ini memberikan bobot yang lebih besar pada kelas minoritas (sel normal/HEM) selama perhitungan loss function, sehingga model tidak bias terhadap kelas mayoritas (sel ALL).

4. Callback Pelatihan: Dua callback penting diaktifkan, yaitu Early Stopping untuk menghentikan pelatihan secara otomatis jika performa pada data validasi tidak lagi meningkat, dan Model Checkpoint untuk memantau dan menyimpan bobot model terbaik berdasarkan nilai validation accuracy.
5. Pelatihan 2 Tahap: Proses pelatihan dilakukan dalam dua tahap berurutan:
 - a. Tahap 1 (Head-only Training): Seluruh lapisan dasar (base model) EfficientNetV2-M dibekukan (frozen). Lapisan head klasifikasi (yang baru ditambahkan) dilatih selama 50 epoch. Secara spesifik, head lama diganti dengan lapisan Dense yang memiliki 1 neuron dan fungsi aktivasi Sigmoid, untuk menghasilkan probabilitas dua kelas (ALL dan HEM).
 - b. Tahap 2 (Fine-tuning): Model terbaik dari Tahap 1 dimuat kembali. Sebagian lapisan atas dari base model dibuka (unfrozen), dan seluruh model dilatih kembali (di-fine-tuning) dengan learning rate yang sama selama 50 epoch.
6. Penyimpanan Model Fold: Model terbaik yang dihasilkan dari iterasi fold ini (Best model saved) disimpan.

Setelah loop K=5 selesai, akan ada 5 model terbaik (satu dari setiap fold). Langkah selanjutnya adalah mengevaluasi kelima model ini menggunakan data uji (Data test) yang telah disisihkan. Model yang menunjukkan performa tertinggi pada data uji (berdasarkan metrik Akurasi, Presisi, Recall, F1-Score, dan Cohen's Kappa) akan dipilih dan disimpan sebagai satu model "juara" (Save best model). Model "juara" inilah yang akan mewakili skenario *Learning Rate* tersebut dalam tahap ensemble di sub-bab berikutnya.

3.4 Rancangan Weighted Ensemble of Networks (WEN)

Tahap *Weighted Ensemble of Networks* (WEN) bertujuan untuk menggabungkan hasil prediksi dari beberapa model tunggal yang telah dilatih.

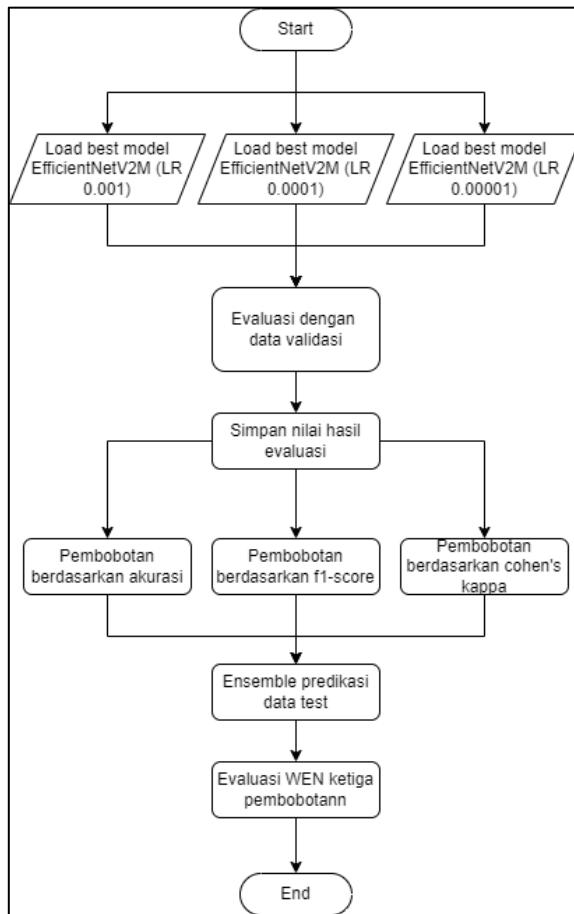
Tujuannya adalah untuk menghasilkan keputusan akhir yang lebih akurat dan stabil (robust) daripada yang dapat dicapai oleh model tunggal terbaik sekalipun.

Dalam penelitian ini, dua skenario WEN utama akan diinvestigasi:

1. WEN Inter-LR (Ensemble per Skenario LR): Skenario ini dilakukan di dalam setiap grup Learning Rate. Contohnya, 5 model terbaik yang dihasilkan dari 5-Fold Cross Validation pada skenario LR 0.001 akan digabungkan. Proses ini diulang untuk skenario LR 0.0001 dan LR 0.00001. Tujuannya adalah untuk melihat apakah gabungan dari 5 fold lebih stabil daripada satu fold "juara".
2. WEN Intra-LR (Ensemble Gabungan Juara): Skenario ini adalah inti dari pengujian ensemble penelitian ini. Di sini, tiga model "juara" (satu model terbaik yang mewakili setiap skenario LR: 0.001, 0.0001, dan 0.00001) akan digabungkan. Hipotesisnya adalah bahwa model-model yang dilatih dengan Learning Rate berbeda akan mempelajari fitur dengan cara yang berbeda, dan menggabungkan "keahlian" yang beragam ini akan menghasilkan model yang paling unggul.

Alur proses untuk skenario WEN Intra-LR (Gabungan Juara) diilustrasikan secara rinci pada Gambar 3.4.

Gambar 3. 4 Rancangan Weighted Ensemble of Networks (Intra-LR)



Tahapan prosesnya dijelaskan sebagai berikut:

1. Load Model: Sistem memanggil tiga model "juara" yang telah disimpan dari tahap 3.3, yaitu: Best model EfficientNetV2M (LR 0.001), Best model EfficientNetV2M (LR 0.0001), dan Best model EfficientNetV2M (LR 0.00001).
2. Evaluasi Awal Tiap Model: Masing-masing dari ketiga model tersebut dievaluasi menggunakan data validasi.
3. Simpan Nilai Hasil Evaluasi: Hasil dari evaluasi ini (nilai metrik Akurasi, F1-Score, dan Cohen's Kappa) disimpan variable. Nilai-nilai inilah yang akan menjadi dasar untuk perhitungan bobot.

4. Perhitungan Bobot (Tiga Skema): Proses ensemble dilakukan sebanyak tiga kali secara paralel untuk menguji tiga skema pembobotan yang berbeda:
 - a. Pembobotan berdasarkan akurasi (WEN^{acc}): Bobot untuk setiap model (LR 0.001, 0.0001, 0.00001) diambil dari nilai Akurasi yang diperoleh pada langkah 2
 - b. Pembobotan berdasarkan f1-score (WEN^{f1}): Bobot untuk setiap model diambil dari nilai F1-Score yang diperoleh pada langkah 2.
 - c. Pembobotan berdasarkan cohen's kappa (WEN^{kappa}): Bobot untuk setiap model diambil dari nilai Cohen's Kappa yang diperoleh pada langkah 2.
5. Ensemble Prediksi Data Test: Untuk setiap skema pembobotan, prediksi probabilitas dari ketiga model digabungkan menggunakan rata-rata tertimbang (weighted average) sesuai bobotnya masing-masing untuk menghasilkan prediksi akhir.
6. Evaluasi WEN Ketiga Pembobotan: Hasil akhir dari ketiga model WEN (WEN^{acc} , WEN^{f1} , dan WEN^{kappa}) dievaluasi kembali performanya menggunakan data test. Tujuannya adalah untuk membandingkan performa dari ketiga skema pembobotan tersebut dan menentukan skema mana yang menghasilkan model ensemble paling optimal, sekaligus untuk mengukur peningkatannya dibandingkan dengan performa model tunggal "juara" terbaik.

3.5 Skenario Uji Coba

Skenario uji coba dilakukan untuk memperoleh model klasifikasi citra leukemia dengan performa terbaik melalui investigasi hyperparameter yang optimal. Proses pelatihan model EfficientNetV2-M dilakukan dalam dua tahap, yaitu Head Training dan Fine-Tuning, masing-masing selama 50 epoch.

Optimizer yang digunakan adalah Adam, dengan nilai batch size sebesar 64 dan untuk menangani ketidakseimbangan kelas, digunakan teknik Class Weight. Seluruh pelatihan model tunggal dilakukan menggunakan metode 5-Fold Cross Validation.

Fokus utama dari skenario uji coba ini adalah untuk membandingkan tiga skenario pelatihan yang berbeda berdasarkan Learning Rate (LR). Detail dari ketiga skenario pelatihan model tunggal dapat dilihat pada Tabel 3.2.

Table 3. 2 Skenario Uji Coba Pelatihan

Model	Learning Rate	Epoch	Metode Validasi	Balancing Data
EfficientNetV2-M	0.001	50 (Head) + 50 (Fine-tuning)	5-Fold Cross Validation	Class Weight
	0.0001			
	0.00001			

Setelah ketiga skenario pelatihan model tunggal di atas selesai dan menghasilkan model-model terbaik, skenario uji coba berikutnya adalah menerapkan *Weighted Ensemble of Networks* (WEN). Skenario *ensemble* dibagi menjadi dua jenis investigasi:

1. WEN Inter-LR: Menggabungkan 5 model yang dihasilkan dari 5-Fold CV di dalam satu skenario LR yang sama (contoh: 5 model dari LR 0.001 digabung).
2. WEN Intra-LR: Menggabungkan 3 model "juara" (satu model terbaik yang mewakili setiap skenario LR).

Untuk setiap skenario ensemble ini, akan diuji tiga skema pembobotan (Akurasi, F1-Score, dan Cohen's Kappa). Rincian skenario ensemble dapat dilihat pada Tabel 3.3.

Table 3. 3 Skenario Ensemble Pembobotan

Skenario	Model yang digabungkan	Dasar Pembobotan
Inter-LR 0.001	5 Model Fold (dari LR 0.001)	Akurasi
		F1-Score
		Cohen's Kappa
Inter-LR 0.0001	5 Model Fold (dari LR 0.0001)	Akurasi
		F1-Score

		Cohen's Kappa
Inter-LR 0.00001	5 Model Fold (dari LR 0.00001)	Akurasi
		F1-Score
		Cohen's Kappa
Intra-LR	3 Model "Juara" (terbaik dari tiap LR)	Akurasi
		F1-Score
		Cohen's Kappa

Pada skenario Intra-LR, setiap model (yang mewakili LR-nya) memberikan output probabilitas yang kemudian dikombinasikan menggunakan bobot sesuai metrik masing-masing model pada data uji. Hasil akhir dari ensemble merupakan nilai rata-rata tertimbang dari ketiga model tersebut.

Pada skenario Inter-LR, proses yang sama dilakukan, namun model yang digabungkan adalah 5 model dari cross-validation dalam satu grup LR yang sama. Evaluasi ini akan menentukan strategi hyperparameter dan strategi ensemble yang paling optimal.

DAFTAR PUSTAKA

- [1] A. Muntasa, R. T. Wahyuningrum, Z. Tuzzahra, A. Motwakel, M. Yusuf, and W. F. Mahmudi, "A Pyramid Model of Convolutional Neural Network to Classify Acute Lymphoblastic Leukemia Images," *International Journal of Intelligent Engineering and Systems*, vol. 15, no. 6, pp. 576–588, Dec. 2022, doi: 10.22266/ijies2022.1231.51.
- [2] A. R. Revanda, C. Fatichah, and N. Suciati, "Classification of Acute Lymphoblastic Leukemia on White Blood Cell Microscopy Images Based on Instance Segmentation Using Mask R-CNN," *International Journal of Intelligent Engineering and Systems*, vol. 15, no. 5, pp. 625–637, Oct. 2022, doi: 10.22266/ijies2022.1031.54.
- [3] L. Rodrigues Moreira, A. Backes, B. Travençolo, and G. Oliveira, "Optimizing a Deep Residual Neural Network with Genetic Algorithm for Acute Lymphoblastic Leukemia Classification," *J Digit Imaging*, vol. 35, pp. 623–637, Feb. 2022, doi: 10.1007/s10278-022-00600-3.
- [4] K. Abdulameer, F. Najjar, A. A. Waad, I. Al-Kharsan, Z. N. Khudhair, and A. Salim, "Leukemia Classification using a Convolutional Neural Network of AML Images," *Malaysian Journal of Fundamental and Applied Sciences*, vol. 19, pp. 306–312, May 2023.
- [5] K. Wibowo, N. Rianto, and J. Unjung, "Improved convolutional neural network model for leukemia classification using EfficientNetV2M and bayesian optimization," *Journal of Soft Computing Exploration*, vol. 5, pp. 192–199, Jun. 2024, doi: 10.52465/josce.v5i2.378.
- [6] A. Batool and Y. Byun, "Lightweight EfficientNetB3 Model Based on Depthwise Separable Convolutions for Enhancing Classification of Leukemia White Blood Cell Images," *IEEE Access*, vol. PP, p. 1, Jan. 2023, doi: 10.1109/ACCESS.2023.3266511.
- [7] M. Zakir Ullah *et al.*, "An Attention-Based Convolutional Neural Network for Acute Lymphoblastic Leukemia Classification," *Applied Sciences*, vol. 11, p. 10662, Nov. 2021, doi: 10.3390/app112210662.
- [8] A. Muntasa *et al.*, "A Novel Hybrid Convolutional and Network Encapsulation Approach in EfficientNetV2-S Architecture for Acute Lymphoblastic Leukemia Classification," *International Journal of Intelligent Engineering and Systems*, 2024, [Online]. Available: <https://api.semanticscholar.org/CorpusID:273622375>
- [9] I. Ahmed, E. Senan, H. Shatnawi, Z. Alkhraisha, and M. Al-Azzam, "Hybrid Techniques for the Diagnosis of Acute Lymphoblastic Leukemia Based on Fusion of CNN Features," *Diagnostics*, vol. 13, p. 1026, Mar. 2023, doi: 10.3390/diagnostics13061026.
- [10] M. Ghadezadeh, M. Aria, A. Hosseini, F. Asadi, D. Bashash, and H. Abolghasemi, "A fast and efficient CNN model for B-ALL diagnosis and its subtypes classification using peripheral blood smear images," *International Journal of Intelligent Systems*, vol. 37, Nov. 2021, doi: 10.1002/int.22753.

- [11] D. Baby, S. Juliet, and M. Raj, "An efficient lymphocytic leukemia detection based on EfficientNets and ensemble voting classifier," *Int J Imaging Syst Technol*, vol. 33, p. n/a-n/a, Jul. 2022, doi: 10.1002/ima.22787.
- [12] C. Mondal *et al.*, "Ensemble of Convolutional Neural Networks to diagnose Acute Lymphoblastic Leukemia from microscopic images," *Inform Med Unlocked*, vol. 27, p. 100794, Nov. 2021, doi: 10.1016/j.imu.2021.100794.
- [13] H. Yu, L. T. Yang, Q. Zhang, D. Armstrong, and M. J. Deen, "Convolutional neural networks for medical image analysis: State-of-the-art, comparisons, improvement and perspectives," *Neurocomputing*, vol. 444, pp. 92–110, 2021, doi: <https://doi.org/10.1016/j.neucom.2020.04.157>.
- [14] I. D. Mienye, T. Swart, G. Obaido, M. Jordan, and P. Ilono, "Deep Convolutional Neural Networks in Medical Image Analysis: A Review," *Information*, p., 2025, doi: 10.3390/info16030195.
- [15] L. Chen, S. Li, Q. Bai, J. Yang, S. Jiang, and Y. Miao, "Review of Image Classification Algorithms Based on Convolutional Neural Networks," *Remote. Sens.*, vol. 13, p. 4712, 2021, doi: 10.3390/rs13224712.
- [16] Md. S. I. Khan *et al.*, "Accurate brain tumor detection using deep convolutional neural network," *Comput Struct Biotechnol J*, vol. 20, pp. 4733–4745, 2022, doi: 10.1016/j.csbj.2022.08.039.
- [17] R. Dasgupta, Y. S. Chowdhury, and S. Nanda, "Performance Comparison of Benchmark Activation Function ReLU, Swish and Mish for Facial Mask Detection Using Convolutional Neural Network," *Algorithms for Intelligent Systems*, p., 2021, doi: 10.1007/978-981-16-2248-9_34.
- [18] S. Niyas, S. Pawan, A. Kumar, and J. Rajan, "Medical image segmentation with 3D convolutional neural networks: A survey," *Neurocomputing*, vol. 493, pp. 397–413, 2021, doi: 10.1016/j.neucom.2022.04.065.
- [19] S. Puligadda, G. Karthik, U. Polina, S. Subbarao, V. P. K. Sistla, and V. K. K. Kolli, "A Novel Transfer Learning-based Efficient-Net for Visual Image Tracking," *2025 International Conference on Pervasive Computational Technologies (ICPCT)*, pp. 531–537, 2025, doi: 10.1109/icpct64145.2025.10941256.
- [20] G. Li, M. Zhang, J. Li, F. Lv, and G. Tong, "Efficient densely connected convolutional neural networks," *Pattern Recognit.*, vol. 109, p. 107610, 2021, doi: 10.1016/j.patcog.2020.107610.
- [21] R. Fu, M. Cao, D. Novák, X. Qian, and N. Alkayem, "Extended efficient convolutional neural network for concrete crack detection with illustrated merits," *Autom Constr*, p., 2023, doi: 10.1016/j.autcon.2023.105098.

- [22] C. Lin, P. Yang, Q. Wang, Z. Qiu, W. Lv, and Z. Wang, "Efficient and accurate compound scaling for convolutional neural networks," *Neural Netw*, vol. 167, pp. 787–797, 2023, doi: 10.1016/j.neunet.2023.08.053.
- [23] V.-T. Hoang, N. D. Quang, T. M. Phuong, K.-H. Jo, and V.-D. Hoang, "A compact version of EfficientNet for skin disease diagnosis application," *Neurocomputing*, vol. 620, p. 129166, 2024, doi: 10.1016/j.neucom.2024.129166.
- [24] M. Di, Z. Peng, and S. Chen, "Performance comparison and application of deep-learning-based image recognition models," vol. 13550, p. 135504, 2025, doi: 10.1111/12.3059903.
- [25] F. Mehmood, S. Ahmad, and T. Whangbo, "An Efficient Optimization Technique for Training Deep Neural Networks," *Mathematics*, p., 2023, doi: 10.3390/math11061360.
- [26] A. Carrington *et al.*, "Deep ROC Analysis and AUC as Balanced Average Accuracy, for Improved Classifier Selection, Audit and Explanation," *IEEE Trans Pattern Anal Mach Intell*, vol. 45, pp. 329–341, 2021, doi: 10.1109/tpami.2022.3145392.
- [27] X. Zhou, J. He, and C. Yang, "An ensemble learning method based on deep neural network and group decision making," *Knowl. Based Syst.*, vol. 239, p. 107801, 2021, doi: 10.1016/j.knosys.2021.107801.
- [28] L. Yates, Z. Aandahl, S. Richards, and B. Brook, "Cross validation for model selection: a review with examples from ecology," *Ecol Monogr*, p., 2022, doi: 10.1002/ecm.1557.
- [29] V. Lumumba, D. Kiprotich, M. Mpaine, N. Makena, and M. Kavita, "Comparative Analysis of Cross-Validation Techniques: LOOCV, K-folds Cross-Validation, and Repeated K-folds Cross-Validation in Machine Learning Models," *American Journal of Theoretical and Applied Statistics*, p., 2024, doi: 10.11648/j.ajtas.20241305.13.
- [30] J. Obi, "A comparative study of several classification metrics and their performances on data," *World Journal of Advanced Engineering Technology and Sciences*, p., 2023, doi: 10.30574/wjaets.2023.8.1.0054.
- [31] M. Z. R. Azhary and A. R. Ismail, "Comparative Performance of Different Convolutional Neural Network Activation Functions on Image Classification," *International Journal on Perceptive and Cognitive Computing*, p., 2024, doi: 10.31436/ijpcc.v10i2.490.
- [32] Y. Wu and L. Liu, "Selecting and Composing Learning Rate Policies for Deep Neural Networks," *ACM Trans Intell Syst Technol*, vol. 14, pp. 1–25, 2022, doi: 10.1145/3570508.
- [33] Z. Wang, H.-B. Gao, X. Wang, S. Zhao, H. Li, and X.-Q. Zhang, "Adaptive learning rate optimization algorithms with dynamic bound based on Barzilai-Borwein method," *Inf. Sci.*, vol. 634, pp. 42–54, 2023, doi: 10.1016/j.ins.2023.03.050.
- [34] S.-H. Cha and D.-S. Kim, "Efficient Training of Deep Spiking Neural Networks Using a Modified Learning Rate Scheduler," *Mathematics*, p., 2025, doi: 10.3390/math13081361.

- [35] B. Bakırarar and A. Elhan, "Class Weighting Technique to Deal with Imbalanced Class Problem in Machine Learning: Methodological Research," *Turkiye Klinikleri Journal of Biostatistics*, p., 2023, doi: 10.5336/biostatic.2022-93961.