

Perbandingan Arsitektur CNN dan Teknik Peningkatan Citra untuk Klasifikasi Pneumonia pada Citra *Chest X-Ray*

Jasmine Mumtaz^{1*}, Teguh Iman Hermanto², Muhammad Rafi Muttaqin³

^{1*23}Teknik Informatika, Sekolah Tinggi Teknologi Wastukencana Purwakarta

*Email: jasminemumtaz34@gmail.com

Abstract

Pneumonia is an acute respiratory disease that contributes to a high global mortality rate. Clinical identification typically utilizes chest x-rays; however, manual analysis by medical practitioners is prone to subjectivity and is time-consuming. Therefore, this research designs an automatic classification system to recognize pneumonia utilizing a Convolutional Neural Network (CNN) while assessing the effectiveness of various image enhancement methods. During the initial preprocessing phase, data balancing was conducted through the random undersampling technique to reduce the majority class, generating a balanced dataset of 1,000 training images for each category. Image pixel quality was then improved using logarithmic transformation, Histogram Equalization (HE), Contrast Limited Adaptive Histogram Equalization (CLAHE), and gamma correction. The model performance evaluation involved Base CNN, VGG-16, and EfficientNet-V2 architectures, which were validated using a stratified 5-fold cross-validation scheme. The experimental findings indicated that combining the VGG-16 architecture with the HE technique yielded the most optimal detection performance. This model achieved an accuracy of 93.44%, precision of 97.30%, recall of 89.40%, F1-score of 93.17%, and an Area Under the Curve (AUC) score of 0.9845. On the other hand, the EfficientNet-V2 model showed suboptimal performance on this dataset, which is indicated to be caused by the implementation of frozen layers during the training process.

Keywords: Chest x-ray, convolutional neural network, image enhancement, pneumonia, VGG-16

Abstrak

Pneumonia merupakan penyakit pernapasan akut yang memicu tingkat mortalitas global yang tinggi. Identifikasi klinis pada umumnya menggunakan radiografi dada (*chest x-ray*), namun proses analisis manual oleh praktisi medis rentan terhadap subjektivitas dan membutuhkan waktu yang lama. Oleh karena itu, riset ini merancang sebuah sistem klasifikasi otomatis untuk mengenali pneumonia dengan memanfaatkan *Convolutional Neural Network* (CNN) sekaligus mengkaji efektivitas metode peningkatan kualitas citra (*image enhancement*). Pada fase awal prapemrosesan, penyeimbangan data dilakukan melalui teknik *random undersampling* untuk mereduksi kelas dominan, sehingga diperoleh dataset seimbang sebanyak 1.000 citra latih untuk setiap kategori. Kualitas piksel citra kemudian diperbaiki menggunakan transformasi logaritmik, *Histogram Equalization* (HE), *Contrast Limited Adaptive Histogram Equalization* (CLAHE), serta koreksi gamma. Pengujian kinerja model melibatkan arsitektur Base CNN, VGG-16, dan EfficientNet-V2 yang divalidasi menggunakan skema *stratified 5-fold cross validation*. Temuan eksperimen mengindikasikan bahwa perpaduan arsitektur VGG-16 dan teknik HE memberikan kinerja deteksi yang paling optimal. Model tersebut mencapai tingkat akurasi 93,44%, presisi 97,30%, *recall* 89,40%, *F1-score* 93,17%, serta luasan *Area Under the Curve* (AUC) 0,9845. Di sisi lain, model EfficientNet-V2 menunjukkan kinerja yang kurang optimal pada dataset ini, yang diindikasikan akibat penerapan skema pembekuan lapisan (*frozen layers*) selama proses pelatihan.

Kata kunci: Chest x-ray, convolutional neural network, image enhancement, pneumonia, VGG-16

1. Pendahuluan

Infeksi saluran pernapasan akut, khususnya pneumonia, tercatat sebagai salah satu pemicu utama tingginya angka kematian global, terutama pada populasi balita di bawah usia lima tahun [1]. Secara medis, tahapan identifikasi awal penyakit infeksi ini umumnya sangat bertumpu pada pengamatan klinis terhadap hasil pencitraan radiografi dada (*Chest X-Ray / CXR*) [2]. Walaupun teknik pencitraan tersebut dinilai praktis dan tidak invasif, analisis yang dilakukan secara manual oleh para tenaga kesehatan kerap membutuhkan durasi panjang dan berisiko tinggi dipengaruhi oleh bias subjektivitas pengamat [3]. Hambatan terbesar dalam pembacaan rontgen secara konvensional terletak pada tipisnya perbedaan corak visual antara jaringan paru yang sehat dengan area yang terpapar infeksi (infiltrat). Kesulitan ini makin bertambah apabila citra yang dihasilkan mesin memiliki resolusi pencahayaan serta tingkat kontras yang buruk, sehingga memperbesar peluang terjadinya kesalahan atau keterlambatan penanganan pasien yang berakibat fatal.

Sebagai upaya untuk meminimalisir kelemahan diagnosis manual, teknologi kecerdasan buatan melalui pendekatan *Deep Learning*, secara khusus pemodelan *Convolutional Neural Network* (CNN), kini mulai diandalkan karena kemampuannya mendeteksi dan mengekstraksi pola citra medis secara otomatis dengan tingkat akurasi yang menjanjikan [4]. Kendati demikian, kinerja algoritma jaringan saraf ini sangat rentan terhadap kualitas data yang diproses. Citra CXR mentah umumnya mengandung derau (*noise*) dan distribusi cahaya asimetris yang berpotensi menyedatkan filter konvolusi komputer saat tahap pelatihan [5]. Di samping itu, dataset klinis di dunia nyata sering kali dihantui oleh fenomena ketimpangan jumlah kelas (*class imbalance*), yang dapat membuat probabilitas tebakan model cenderung berpihak pada kelas mayoritas [6].

Berbagai riset terdahulu telah berupaya menanggulangi kendala ini dengan menguji ragam arsitektur; misalnya, model klasik VGG-

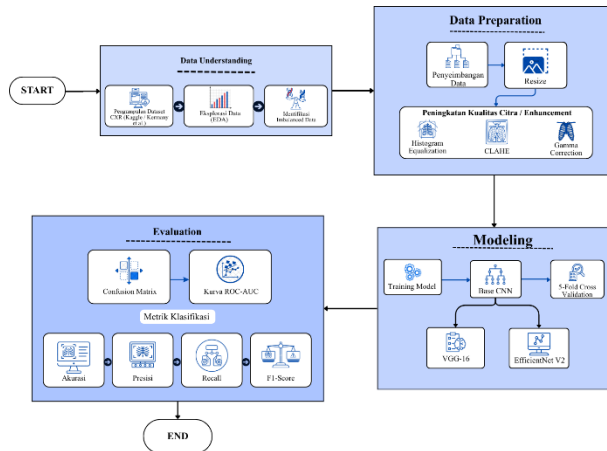
16 yang dikenal sangat solid untuk mengekstraksi fitur fundamental [7], hingga pemanfaatan arsitektur mutakhir EfficientNet-V2 yang menawarkan efisiensi komputasi jauh lebih ringan [8]. Literatur lain juga mengonfirmasi bahwa penerapan metode perbaikan kualitas visual, seperti *Contrast Limited Adaptive Histogram Equalization* (CLAHE), sukses mendongkrak presisi pengklasifikasian model secara drastis [9].

Meskipun penjelajahan terhadap arsitektur jaringan telah banyak dilakukan, riset komparatif yang mengintegrasikan teknik penyeimbangan data (*random undersampling*) dengan evaluasi simultan terhadap berbagai metode prapemrosesan—seperti *Histogram Equalization* (HE), CLAHE, dan koreksi Gamma—dalam satu kerangka pengujian masih tergolong minim [10]. Berangkat dari celah penelitian (*research gap*) tersebut, studi ini menawarkan kebaruan dengan mengevaluasi dan membandingkan kinerja arsitektur *baseline* Base CNN, VGG-16, dan arsitektur mutakhir EfficientNet-V2 pada dataset yang telah diselaraskan. Dengan menerapkan skema pengujian *stratified 5-fold cross-validation* yang terstruktur, riset ini berkontribusi merumuskan perpaduan paling optimal antara metode peningkatan citra dan arsitektur *Deep Learning*. Lebih dari itu, sebagai kontribusi praktis, model komputasi terbaik dari penelitian ini turut didemonstrasikan ke dalam sebuah prototipe *dashboard* interaktif berbasis web yang berpotensi direalisasikan sebagai sistem cerdas pendukung keputusan klinis terotomatisasi guna mempercepat penapisan penyakit pneumonia.

2. Metoda Penelitian

Metoda yang dipakai dalam penelitian ini berupa langkah kerja serta rangkaian kegiatan yang mengimplementasikan alur kerja *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Pendekatan struktural ini secara khusus ditekankan pada penyempurnaan efisiensi komputasi di fase prapemrosesan data citra medis, serta optimalisasi perancangan arsitektur model klasifikasinya. Alur kerja

komprehensif ini divisualisasikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1. Pemahaman dan Penyeimbangan Data

Objek observasi yang dianalisis berupa koleksi data sekunder citra Chest X-Ray (Pneumonia) yang diekstraksi dari pangkalan data terbuka Kaggle [11]. Tinjauan awal terhadap 5.863 citra mentah menemukan adanya asimetri distribusi kelas yang ekstrem, di mana jumlah citra paru-paru terinfeksi pneumonia (3.876 sampel) jauh melampaui citra paru-paru sehat (1.342 sampel). Untuk mencegah kecenderungan algoritma memihak pada kelas dominan (*bias heuristik*), diterapkanlah teknik *random undersampling* [12]. Prosedur ini memangkas populasi kelas mayoritas secara acak sehingga tercipta titik ekuilibrium absolut, yakni 1.000 citra untuk masing-masing kategori (menghasilkan total 2.000 citra latih yang seimbang). Seluruh gambar tersebut kemudian diseragamkan resolusi spasialnya menjadi matriks 224×224 piksel agar kompatibel dengan standar masukan arsitektur transfer learning.

2.2. Transformasi dan Peningkatan Kualitas Citra

Tahapan optimalisasi kualitas piksel dieksekusi secara eksklusif memanfaatkan fungsi transformasi logaritmik untuk meregangkan rentang kontras pada intensitas cahaya rendah tanpa mendistorsi pola klinis

paru-paru, yang diformulasikan pada Persamaan 1:

$$I_{out} = c \cdot \ln(I_{in} + 1) \dots \dots \dots (1)$$

Di mana I_{out} dan I_{in} adalah intensitas piksel keluaran dan masukan, serta c merupakan konstanta penskalaan. Selanjutnya, kualitas visual gambar direkayasa dan dibandingkan menggunakan tiga algoritma pemrosesan citra:

1. Histogram Equalization (HE):

Bertugas meratakan kepadatan probabilitas piksel secara global menggunakan distribusi kumulatif [13]. Perhitungan probabilitas (*Probability Density Function* / PDF) dan distribusi kumulatif (*Cumulative Distribution Function* / CDF) ditunjukkan pada Persamaan 2:

$$P(X) = \frac{n_k}{n}$$

$$C(a) = \sum_{j=0}^k P(x_j) \dots \dots \dots (2)$$

2. Contrast Limited Adaptive Histogram Equalization (CLAHE):

Memecah gambar ke dalam grid 8×8 piksel guna menajamkan detail anatomi secara lokal sekaligus membatasi amplifikasi derau [14]. Batas pemotongan (β) serta distribusi ulang nilai piksel berlebih (R) dihitung menggunakan Persamaan 3:

$$\beta = \frac{MN}{G} \left(1 + \frac{\alpha}{100} (AS_{max} - 1) \right) \quad (3)$$

$$R = \frac{N_{clip}}{L} \dots \dots \dots (2)$$

3. Koreksi Gamma: Mengaplikasikan pemetaan non-linear guna menyeimbangkan iluminasi citra rontgen yang gelap [15], sesuai dengan Persamaan 4:

$$I_{out} = c \cdot I_{in}^Y \dots \dots \dots (4)$$

2.3. Pemodelan Arsitektur Jaringan

Eksperimen komputasi ini membandingkan kinerja model Base CNN konvensional dengan dua arsitektur transfer learning mutakhir, yaitu VGG-16 dan EfficientNet-V2. Pada VGG-16, dimensi keluaran spasial (*output feature map*) dari setiap operasi konvolusinya dikalkulasi menggunakan Persamaan 5 [16]:

$$O = 1 + \frac{N+2P-F}{S} \dots \dots \dots (5)$$

Pada tahap inialisasi, VGG-16 dan EfficientNet-V2 memanfaatkan skema pembekuan lapisan latih (*frozen layers*) dari bobot *ImageNet* untuk menjaga kemampuan fundamental dalam mengekstraksi fitur. Matriks fitur yang dihasilkan kemudian diratakan menggunakan *GlobalAveragePooling2D*, dinormalisasi melalui *BatchNormalization*, serta dilewatkan pada dua *hidden layer* berkapasitas 256 dan 128 neuron dengan fungsi aktivasi *Rectified Linear Unit* (ReLU). Guna mencegah terjadinya *overfitting*, disematkan teknik regularisasi *Dropout* dengan rasio 0,5 dan 0,3. Pengklasifikasian biner akhir diputuskan oleh satu neuron tunggal beraktivasi *Sigmoid*.

Untuk menjamin keterulangan eksperimen (*reproducibility*), proses pelatihan model dieksekusi menggunakan pengoptimal (*optimizer*) Adam dengan tingkat pembelajaran (*learning rate*) sebesar 0,001 dan fungsi kerugian *binary crossentropy*. Seluruh skenario pemodelan dilatih menggunakan batasan 15 *epoch* dengan ukuran angkatan (*batch size*) sebesar 32. Guna mencegah terjadinya *overfitting* selama pelatihan, diterapkan metode *Early Stopping* dengan parameter *patience* 3, yang akan menghentikan proses iterasi secara otomatis apabila tidak terdapat penurunan nilai *validation loss*.

2.4. Validasi dan Evaluasi Kinerja

Pengujian ketahanan dan objektivitas model divalidasi secara ketat menggunakan metode pengujian silang *Stratified 5-Fold Cross Validation*. Tolok ukur keberhasilan deteksi dievaluasi secara diskrit menggunakan matriks kebingungan (*confusion matrix*) yang memetakan nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN).

		Predicted Class		
		Positive	Negative	
Actual Class	Positive	True Positive (TP) Type II error	False Negative (FN) Type I error	Recall $\frac{TP}{TP + FN}$
	Negative	False Positive (FP) Type I error	True Negative (TN)	Specificity $\frac{TN}{TN + FP}$
		Precision $\frac{TP}{TP + FP}$	Negative predictive value $\frac{TN}{TN + FN}$	Accuracy $\frac{TP + TN}{TP + TN + FP + FN}$

Gambar 2. *Confusion Matrix*

Berdasarkan keempat parameter absolut tersebut, ekstraksi metrik performa statistik dievaluasi menggunakan kumpulan rumus berikut [17]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$F1 - Score = \frac{Precision \times Recall}{Precision + Recall}$$

Di samping komputasi metrik di atas, kemampuan probabilitas algoritma dalam memisahkan kelas anatomi paru-paru sehat dan sakit turut divalidasi melalui perhitungan

luasannya *Area Under the Curve* (AUC) pada grafik *Receiver Operating Characteristic* (ROC).

Proses pembagian data latih dan data uji dieksekusi secara acak menggunakan *Stratified 5-Fold Cross Validation* guna memastikan keseimbangan proporsi kelas (rasio 1:1) tetap terjaga pada setiap iterasi pengujian. Terkait dengan pemisahan data berdasarkan identitas pasien (*patient-level split*), dataset sekunder yang diunduh dari repositori publik tidak menyertakan metadata rekam medis individu secara terstruktur. Oleh karena itu, pemisahan data secara ketat berbasis pasien tidak dapat diimplementasikan. Meskipun demikian, risiko kebocoran data (*data leakage*) akibat sampel dari pasien yang sama berada di partisi latih dan uji telah ditekan semaksimal mungkin melalui skema *cross-validation* yang ketat serta penerapan *data augmentation* secara *real-time* untuk mencegah model sekadar menghafal karakteristik spesifik citra.

3. Hasil Penelitian

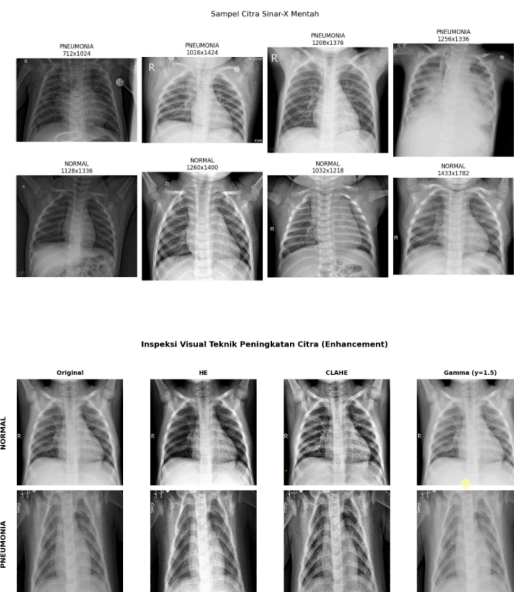
Bagian ini memaparkan secara komprehensif temuan eksperimen, meliputi dampak dari rekayasa kualitas citra, evaluasi metrik performa terhadap 12 skenario pengujian silang, serta penentuan arsitektur terbaik dengan bersandar pada justifikasi klinis melalui matriks kebingungan dan kurva ROC.

3.1. Prapemrosesan dan Rekayasa Visual Citra

Tinjauan awal terhadap koleksi data mentah mengindikasikan adanya asimetri proporsi kelas yang sangat tajam, dengan perbandingan 3.876 sampel Pneumonia melawan 1.342 sampel Normal. Tantangan ketimpangan ini sukses diselesaikan melalui implementasi *Random undersampling*, yang secara konsisten memangkas sampel kelas mayoritas hingga tercapai titik ekuilibrium absolut sebanyak 1.000 citra latih untuk tiap-tiap kelas.

Pada aspek kualitas iluminasi, pencahayaan citra yang sub-optimal direkayasa

melalui tiga teknik manipulasi piksel. Peninjauan secara visual menunjukkan bahwa metode *Histogram Equalization* (HE) sangat agresif dalam memunculkan tekstur bercak putih (infiltrat paru) melalui penyebaran nilai probabilitas secara global, meskipun berisiko memunculkan efek pencerahan berlebih (*over-enhancement*). Di sisi lain, metode CLAHE terbukti mampu mempertegas batas anatomi tulang rusuk secara adaptif tanpa mengamplifikasi derau. Sementara itu, teknik koreksi Gamma bekerja secara non-linear untuk menggelapkan latar belakang, sehingga kabut peradangan dapat terpisah secara halus dari struktur tulang seperti pada gambar 3.



Gambar 3. Citra X-Ray Mentah dan Inpeksi Visual Teknik *Image Enhancement*

3.2. Komparasi Kinerja Metrik Klasifikasi

Proses pengujian kinerja model mencakup 12 skenario eksperimen yang mengombinasikan arsitektur jaringan (*Base CNN*, *VGG-16*, *EfficientNet-V2*) dengan modifikasi data masukan (*Baseline*, *HE*, *CLAHE*, *Gamma*). Seluruh pengujian ini divalidasi menggunakan skema *Stratified 5-Fold Cross Validation* untuk memastikan stabilitas dan objektivitas varians prediksi. Rekapitulasi hasil komputasi dari pengujian tersebut dirangkum pada Tabel 1.

Tabel 1. Metrik Klasifikasi 12 Skenario Pemodelan

No	Arsitektur	Metode Preprocessing	Accuracy	Precision	Recall	F1-Score	AUC
1	Base CNN	Baseline	0.743098	0.743049	1.00000	0.852585	0.792177
2	Base CNN	HE	0.829429	0.894051	0.75300	0.811744	0.916706
3	Base CNN	CLAHE	0.752945	0.960840	0.54100	0.610786	0.917442
4	Base CNN	Gamma	0.718779	0.816199	0.67300	0.639650	0.859621
5	VGG-16	Baseline	0.742906	0.742906	1.00000	0.852491	0.883166
6	VGG-16	HE	0.934472	0.973057	0.89400	0.931752	0.984512
7	VGG-16	CLAHE	0.925471	0.984075	0.86500	0.920637	0.986890
8	VGG-16	Gamma	0.913462	0.961232	0.86200	0.908855	0.974700
9	EfficientNet	Baseline	0.757479	0.790972	0.91663	0.848834	0.785551
10	EfficientNet	HE	0.727352	0.774804	0.67000	0.703967	0.791529
11	EfficientNet	CLAHE	0.711347	0.765178	0.66600	0.689477	0.803805
12	EfficientNet	Gamma	0.713356	0.782770	0.61700	0.678642	0.775156

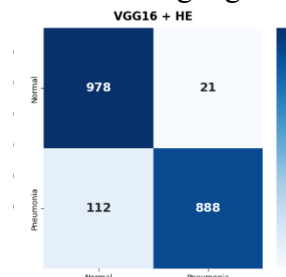
Merujuk pada Tabel 1, ditemukan tiga temuan analitis yang sangat krusial dari agregasi nilai rata-rata *5-Fold Cross Validation*. Pertama, model yang dilatih menggunakan data mentah (Baseline) mengalami fenomena bias heuristik; di mana VGG-16 dan Base CNN meraih nilai Recall sempurna (1,000) namun Precision-nya sangat rendah (sekitar 0,743). Hal ini membuktikan bahwa algoritma gagal mengenali fitur spasial dan hanya memprediksi secara membabi buta ke arah kelas dominan.

Kedua, penerapan arsitektur *transfer learning* VGG-16 yang dipasangkan dengan metode HE mencatatkan lonjakan metrik yang optimal, dengan tingkat akurasi mencapai 93,44% dan F1-Score 93,17%. Keberhasilan VGG-16 dalam mengekstraksi fitur dasar paru-paru ini sejalan dengan penelitian terdahulu pada [7] yang juga mengonfirmasi ketangguhan arsitektur VGG-16 untuk citra Chest X-Ray. Terakhir, arsitektur mutakhir EfficientNet-V2 justru menunjukkan kinerja yang kurang optimal (akurasi berkisar 71%-75%). Merujuk pada karakteristik arsitektur yang sangat dalam (*deep*), stagnansi performa ini diindikasikan terjadi karena skema pembekuan lapisan (*frozen base*) dari bobot ImageNet kurang adaptif terhadap representasi citra medis berskala abu-abu (*grayscale*) tanpa didampingi proses penyesuaian bobot (*fine-tuning*) lanjutan, sebuah fenomena yang umum terjadi pada penerapan transfer learning citra medis kompleks.

Dalam proses eksekusi komputasional pada *environment cloud*, ditemukan adanya anomali teknis berupa reduksi 1 sampel pada kelas Normal, sehingga total data uji tervalidasi pada kelas tersebut menjadi 999 citra. Fenomena eksklusi otomatis ini merupakan bentuk penyaringan tingkat tinggi (*data sanitization*) oleh *pipeline TensorFlow*, di mana algoritma secara mandiri menggugurkan data yang teridentifikasi sebagai artefak direktori tersembunyi (*hidden system files*) atau citra yang mengalami kerusakan struktur *header (corrupted file)* selama proses penarikan data mentah. Pengguguran 1 sampel ini ditoleransi secara statistik karena tidak memengaruhi rasio ekuilibrium secara signifikan maupun mengubah probabilitas distribusi kelas.

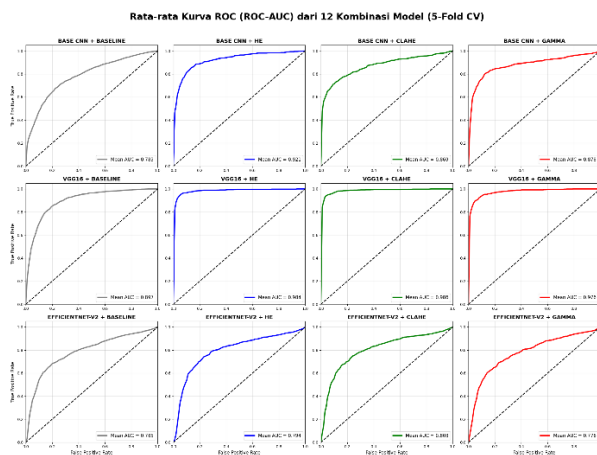
3.3. Evaluasi Keandalan Klinis dan Separabilitas Probabilitas

Dalam domain diagnostik medis, tingkat akurasi tunggal tidak dapat dijadikan pedoman absolut tanpa mengevaluasi sebaran kegagalan prediksi pada matriks kebingungan.



Gambar 4. Confusion Matrix VGG-16 + HE

Observasi terhadap Matriks Kebingungan mengonfirmasi bahwa perpaduan VGG-16 dengan prapemrosesan HE menawarkan tingkat keamanan klinis tertinggi. Model ini berhasil mendeteksi 888 sampel positif secara presisi dan secara drastis menekan angka kegagalan diagnosis (*False Negative*) hingga menyisakan 112 citra. Meminimalkan kesalahan Tipe II (*Type II Error*) ini adalah keharusan mutlak dalam prosedur medis agar tidak ada pasien terinfeksi yang terabaikan dari penanganan.

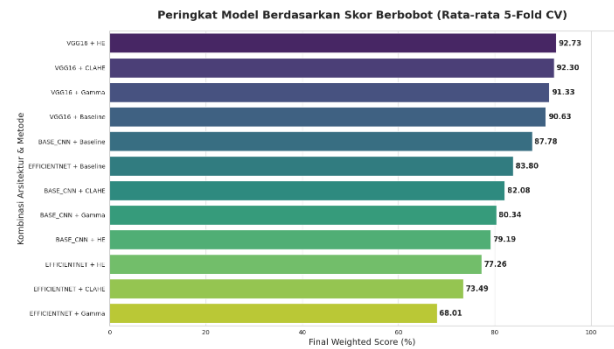


Gambar 5. Komparasi Kurva ROC dan Nilai AUC (5-Fold CV)

Untuk memvalidasi kapasitas probabilitas model dalam memisahkan pasien berparu-paru sehat dan sakit, dievaluasilah luasan *Area Under the Curve* (AUC). Kurva ROC menunjukkan bahwa VGG-16 berbasis HE memiliki batas separasi yang sangat tangguh dengan nilai AUC 0,9845 (mendekati probabilitas sempurna 1,0). Skor istimewa ini menegaskan bahwa algoritma tersebut memiliki keandalan diagnostik tingkat tinggi (kategori *excellent*).

Mengingat ketatnya margin evaluasi antar model, justifikasi arsitektur terbaik diputuskan melalui perhitungan Skor Berbobot (*Weighted Score*). Skema ini mendistribusikan bobot secara proporsional berdasarkan urgensi klinis, dengan prioritas tertinggi diberikan pada metrik *Recall* (40%) guna menekan risiko kegagalan deteksi (*False Negative*). Selanjutnya, parameter *Area Under the Curve* (AUC) diberi bobot 25% sebagai indikator konsistensi

pemisahan kelas, *F1-Score* sebesar 20% untuk menjaga harmonisasi presisi, dan *Accuracy* sebesar 15%. Pembobotan terstruktur ini diterapkan untuk memastikan model yang terpilih tidak hanya akurat secara matematis, tetapi juga benar-benar peka dalam mendeteksi kasus positif pneumonia secara medis.



Gambar 6. Visualisasi Peringkatan Model Terbaik

Berdasarkan komputasi pembobotan, arsitektur VGG-16 yang diintegrasikan dengan *Histogram Equalization* (HE) sukses menempati peringkat pertama dengan skor absolut sebesar 92,73%. Pencapaian komprehensif ini mengunci kesimpulan bahwa VGG-16 termodifikasi HE merupakan sistem klasifikasi yang paling presisi, tangguh, dan aman untuk dikembangkan menjadi instrumen pendukung keputusan klinis (*Clinical Decision Support System*).

4. Kesimpulan

Berdasarkan serangkaian pengujian komputasi dan analisis metrik yang telah dieksekusi, penelitian ini menghasilkan beberapa kesimpulan utama. Pertama, penerapan teknik *random undersampling* terbukti sangat efektif dalam mereduksi ketimpangan jumlah kelas pada dataset *Chest X-Ray*, sehingga algoritma terhindar dari fenomena bias heuristik terhadap kelas dominan saat fase pelatihan berlangsung. Kedua, dalam hal rekayasa kualitas visual, metode *Histogram Equalization* (HE) menunjukkan hasil ekstraksi fitur yang paling optimal dibandingkan CLAHE dan koreksi Gamma. Distribusi probabilitas piksel secara global yang dihasilkan oleh HE sukses

mempertegas kontras area infiltrat pada paru-paru yang terinfeksi, sehingga pola peradangan tersebut jauh lebih mudah dipetakan oleh matriks konvolusi.

Ketiga, Ketiga, komparasi arsitektur mengonfirmasi bahwa integrasi *transfer learning* VGG-16 dengan prapemrosesan HE menghasilkan sistem deteksi pneumonia yang paling tangguh. Model spesifik ini mencapai akurasi tertinggi sebesar 93,44%, presisi 97,30%, sensitivitas (*recall*) 89,40%, *F1-Score* 93,17%, serta keandalan pemisahan probabilitas (AUC) di angka 0,9845. Tingginya capaian metrik tersebut, khususnya kemampuan menekan angka *False Negative*, membuktikan bahwa model ini memiliki potensi yang menjanjikan sebagai alat bantu pendukung keputusan klinis (*Clinical Decision Support System*) tahap awal. Meskipun demikian, validasi menggunakan kumpulan data klinis lintas institusi berskala lebih besar tetap diperlukan sebelum model ini dapat diterapkan secara operasional penuh di rumah sakit. Terakhir, penelitian ini menemukan bahwa arsitektur EfficientNet-V2 menunjukkan kinerja yang kurang optimal ketika memproses citra medis berskala abu-abu, yang utamanya dipicu oleh pembekuan bobot lapisan (*frozen layers*) sehingga menghambat adaptasi fitur radiografi.

5. Saran

Guna menyempurnakan keandalan instrumen diagnostik di masa mendatang, terdapat beberapa poin saran yang direkomendasikan. Pertama, disarankan untuk mengimplementasikan skema pencairan lapisan latih (*unfreezing layers*) atau *fine-tuning* secara bertahap pada arsitektur berhierarki dalam seperti EfficientNet-V2, sehingga jaringan tersebut mampu mempelajari representasi anatomi secara lebih mendalam. Kedua, ruang lingkup pelatihan perlu diperluas dengan memanfaatkan dataset klinis multi-sumber guna mendongkrak kemampuan generalisasi algoritma terhadap variasi resolusi dari berbagai mesin rontgen rumah sakit. Terakhir, peneliti selanjutnya direkomendasikan untuk

mengeksplorasi algoritma lokalisasi objek (seperti YOLO atau *Mask R-CNN*) agar sistem tidak hanya mengklasifikasikan status pneumonia, tetapi juga mampu memberikan kotak pembatas (*bounding box*) secara visual pada area paru-paru yang mengalami infeksi kritis.

6. Daftar Pustaka

- [1] UNICEF, World Health Organization, World Bank, and United Nations, "Levels and Trends in Child Mortality: Report 2023 – Estimates Developed by the United Nations Inter-agency Group for Child Mortality Estimation," Washington, D.C., 2023. Accessed: Mar. 12, 2026. [Online]. Available: <http://documents.worldbank.org/curated/en/099606103132489570>
- [2] C. A. Gomez *et al.*, "Clinical and microbiological characteristics of pediatric patients hospitalized for pneumococcal pneumonia before and after the introduction of pneumococcal conjugate vaccines," *Rev. Peru. Med. Exp. Salud Publica*, vol. 42, pp. 63–69, 2025.
- [3] Y. Jin, W. Chang, and B. Ko, "Generating Chest X-Ray Progression of Pneumonia Using Conditional Cycle Generative Adversarial Networks," *IEEE Access*, vol. 11, 2023, doi: 10.1109/ACCESS.2023.3305994.
- [4] R. Siddiqi and S. Javaid, "Deep Learning for Pneumonia Detection in Chest X-ray Images: A Comprehensive Survey," 2024. doi: 10.3390/jimaging10080176.
- [5] M. Ali *et al.*, "Pneumonia Detection Using Chest Radiographs with Novel EfficientNetV2L Model," *IEEE Access*, vol. 12, pp. 34691–34707, 2024, doi: 10.1109/ACCESS.2024.3372588.
- [6] W. Khan, N. Zaki, and L. Ali, "Intelligent Pneumonia Identification from Chest X-Rays: A Systematic Literature Review," 2021. doi: 10.1109/ACCESS.2021.3069937.
- [7] A. Rasyid and L. Heryawan, "Klasifikasi Penyakit Tuberculosis (TB) Organ Paru Manusia Berdasarkan Citra Rontgen Thorax Menggunakan Metode Convolutional Neural Network (CNN)," *Jurnal Manajemen Informasi Kesehatan Indonesia (JMIKI)*, vol.

- 11, no. 1, 2023, doi: 10.33560/jmiki.v11i1.484.
- [8] K. Guo, J. Cheng, K. Li, L. Wang, Y. Lv, and D. Cao, "Diagnosis and detection of pneumonia using weak-label based on X-ray images: a multi-center study," *BMC Med. Imaging*, vol. 23, no. 1, Dec. 2023, doi: 10.1186/s12880-023-01174-4.
- [9] N. E. P. N. E. Putri, R. R. R. Rahmadewi, and S. M. P. S. M. Putri, "Studi Komparasi HE dan CLAHE pada Image Enhancement untuk Segmentasi Citra dalam Deteksi Tumor pada Otak," *Telekontran: Jurnal Ilmiah Telekomunikasi, Kendali dan Elektronika Terapan*, vol. 13, no. 2, pp. 298–307, 2025.
- [10] J. A. Prakash, V. Ravi, V. Sowmya, and K. P. Soman, "Stacked ensemble learning based on deep convolutional neural networks for pediatric pneumonia diagnosis using chest X-ray images," *Neural Comput. Appl.*, vol. 35, no. 11, pp. 8259–8279, Apr. 2023, doi: 10.1007/s00521-022-08099-z.
- [11] P. Mooney, "Chest X-Ray Images (Pneumonia)," Kaggle. Accessed: May 06, 2026. [Online]. Available: <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>
- [12] V. I. Yani, A. Aradea, and H. Mubarak, "Optimasi Prakiraan Cuaca Menggunakan Metode Ensemble pada Naïve Bayes dan C4.5," *Jurnal Teknik Informatika Dan Sistem Informasi*, vol. 8, no. 3, pp. 607–619, 2022.
- [13] A. Mansour, M. Gaheen, and م. ع. ر. عماشة, "Enhancement of radiographs using the HE and CLAHE techniques," vol. 2024, no. 10, 2024, doi: 10.21608/sjeud.2024.290584.1038.
- [14] S. Saiwaeo, S. Arwatchananukul, L. Mungmai, W. Preedalikit, and N. Aunsri, "Human skin type classification using image processing and deep learning approaches," *Heliyon*, vol. 9, no. 11, 2023, doi: 10.1016/j.heliyon.2023.e21176.
- [15] E. Mahamud, N. Fahad, M. Assaduzzaman, S. M. Zain, K. O. M. Goh, and M. K. Morol, "An explainable artificial intelligence model for multiple lung diseases classification from chest X-ray images using fine-tuned transfer learning," *Decision Analytics Journal*, vol. 12, 2024, doi: 10.1016/j.dajour.2024.100499.
- [16] M. M. Taye, "Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions," 2023. doi: 10.3390/computation11030052.
- [17] E. Helmud, E. Helmud, F. Fitriyani, and P. Romadiana, "Classification Comparison Performance of Supervised Machine Learning Random Forest and Decision Tree Algorithms Using Confusion Matrix," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 13, no. 1, 2024, doi: 10.32736/sisfokom.v13i1.1985.