

Optimasi Model CNN Berbasis *Transfer Learning* Untuk Klasifikasi Pneumonia pada Citra X-Ray Dada

Rasikh Khalil Pasha¹
Kholiq Budiman²

^{1,2}Ilmu Komputer, FMIPA, Universitas Negeri Semarang, Sekaran, Kec. Gn. Pati, Kota Semarang, Jawa Tengah 50229, Indonesia

¹rasikhkopi@students.unnes.ac.id, ²kholiq.budiman@mail.unnes.ac.id

***Penulis Korespondensi:**
Rasikh Khalil Pasha
rasikhkopi@students.unnes.ac.id

Abstrak

Pneumonia merupakan salah satu penyebab utama kematian anak di dunia dan diagnosisnya bergantung pada interpretasi citra radiografi dada (X-ray) yang rentan terhadap kesalahan manusia. Penelitian ini bertujuan mengoptimalkan model *Convolutional Neural Network* (CNN) berbasis *transfer learning* dengan arsitektur DenseNet-121 untuk klasifikasi pneumonia pada citra X-ray dada. Model dikembangkan menggunakan dataset dari Kaggle yang terdiri atas dua kelas: Normal dan Pneumonia. Data mengalami pra-pemrosesan berupa augmentasi dan penyeimbangan kelas. Lima strategi *fine-tuning* diuji, mulai dari hanya melatih *classifier* hingga membuka seluruh lapisan *pretrained*. Evaluasi dilakukan menggunakan metrik akurasi, presisi, recall, F1-score, dan ROC-AUC. Hasil menunjukkan bahwa strategi *fine-tuning* dengan membuka blok 3 sampai 4 (*Unfreeze Block 3-4*) menghasilkan performa terbaik dengan akurasi 94,39%, F1-score 95,61%, dan ROC-AUC 98,04%. Studi ini menunjukkan bahwa strategi *fine-tuning* selektif mampu meningkatkan performa klasifikasi secara signifikan dibanding melatih hanya *classifier* atau seluruh jaringan.

Kata Kunci: DenseNet-121; Fine-Tuning; Pneumonia; Transfer Learning

Abstract

Pneumonia is a leading cause of child mortality worldwide, and its diagnosis often relies on chest X-ray interpretation, which is prone to human error. This study aims to optimize a *Convolutional Neural Network* (CNN) model based on transfer learning using the DenseNet-121 architecture for pneumonia classification in chest X-ray images. The model was trained on a Kaggle dataset consisting of two classes: Normal and Pneumonia. Preprocessing included class balancing and data augmentation. Five fine-tuning strategies were tested, ranging from training only the classifier to unfreezing the entire pretrained layers. Evaluation metrics included accuracy, precision, recall, F1-score, and ROC-AUC. Results showed that the strategy of unfreezing Block 3-4 yielded the best performance with 94.39% accuracy, 95.61% F1-score, and 98.04% ROC-AUC. This study demonstrates that selective fine-tuning strategies significantly improve classification performance compared to training only the classifier or the entire network.

Keywords: DenseNet-121; Fine-Tuning; Pneumonia; Transfer Learning

1. Pendahuluan

Pneumonia merupakan penyakit pernapasan serius dan menjadi salah satu masalah kesehatan global maupun nasional. Data Global Burden of Disease (GBD) 2021 menunjukkan bahwa pneumonia menyumbang sekitar 2,1 juta kematian di seluruh dunia [1]. Data tersebut juga menunjukkan bahwa kelompok usia di bawah 5 tahun sangat rentan terhadap pneumonia, yang menyebabkan sekitar 500.000 kematian [1]. Setiap hari, hampir 2.200 anak balita meninggal karena pneumonia, setara dengan satu anak setiap 39 detik [2]. Di Indonesia, pneumonia juga menjadi beban kesehatan utama anak-anak. Pneumonia menyumbang 16% dari keseluruhan kematian balita pada 2018 dan merupakan penyebab kematian anak usia <5 tahun terbesar kedua pada 2017. Pada tahun 2018 diperkirakan sekitar 19.000 anak di bawah lima tahun meninggal akibat pneumonia di Indonesia [3]. Fakta-fakta ini menegaskan urgensi penanganan pneumonia sebagai prioritas kesehatan publik di tingkat global maupun nasional.

Diagnosis pneumonia hampir selalu memerlukan konfirmasi radiologis. Standar emas untuk menegakkan diagnosis pneumonia menurut pedoman internasional adalah radiografi dada (sinar-X) [4]. Namun, interpretasi radiografi dada memerlukan waktu dan sangat bergantung pada keahlian radiolog. Dalam praktik klinis, pembacaan CXR dapat menimbulkan variabilitas antar-pemeriksa dan potensi kesalahan interpretasi [5]. Tantangan ini diperparah oleh ketersediaan tenaga radiologi yang terbatas, terutama di daerah dengan sumber daya kesehatan rendah [6]. Kondisi ini menimbulkan kebutuhan akan sistem diagnosis berbasis citra otomatis yang dapat membantu atau menggantikan pembacaan manual oleh dokter, guna mempercepat deteksi pneumonia dan mengurangi beban kerja radiolog.

Dalam beberapa tahun terakhir, metode pembelajaran mendalam (deep learning) berbasis convolutional neural networks (CNN) telah merevolusi klasifikasi citra, termasuk citra medis [7], [8], [9], [10]. CNN secara konsisten menjadi pendekatan utama dalam kompetisi dan penelitian analisis citra medis [11]. Namun, pelatihan CNN dari awal membutuhkan data yang sangat banyak, sedangkan data medis sering kali terbatas. Untuk mengatasi hal ini, pendekatan transfer learning banyak digunakan, yaitu dengan memanfaatkan model CNN pralatih pada dataset besar seperti ImageNet, lalu menyesuaikannya dengan tugas klasifikasi khusus seperti pneumonia [12]. Strategi ini terbukti mampu meningkatkan performa dan efisiensi pelatihan model meskipun jumlah data medis relatif kecil [13].

Berbagai arsitektur CNN telah dikembangkan untuk menangani kompleksitas data citra, mulai dari arsitektur sederhana seperti AlexNet hingga yang lebih kompleks seperti VGG, ResNet, Inception, dan DenseNet. Masing-masing arsitektur memiliki karakteristik unik dalam hal kedalaman jaringan, cara propagasi fitur, dan efisiensi parameter. Dalam konteks klasifikasi citra medis, arsitektur yang mampu mempertahankan fitur penting dan mengatasi degradasi gradien pada jaringan dalam sangatlah penting.

Salah satu arsitektur yang menonjol dalam hal ini adalah DenseNet (Densely Connected Convolutional Networks). Sejumlah penelitian telah menunjukkan efektivitas CNN dalam klasifikasi pneumonia dari citra X-ray dada. Salehi et al. (2021) [14], misalnya, mengevaluasi sejumlah model CNN pralatih seperti VGG19, DenseNet121, Xception, ResNet50, dengan pendekatan transfer learning. DenseNet-121 menghasilkan performa terbaik dengan akurasi sebesar 86.8%. Studi lain oleh Chowdhury et al. (2020) [15] menggunakan DenseNet-201 dan mencatat akurasi hingga 99,7% untuk klasifikasi biner, menunjukkan potensi tinggi dari DenseNet dalam klasifikasi citra medis. Alhudhaif et al. (2021) [16] juga menemukan bahwa DenseNet-201 mengungguli model lain dalam membedakan pneumonia COVID-19 dan non-COVID, dengan akurasi 94,96% dan F1-score 92,11%.

Keunggulan DenseNet dalam klasifikasi citra medis tidak lepas dari arsitekturnya yang unik. DenseNet menghubungkan setiap lapisan ke semua lapisan sebelumnya, memungkinkan aliran gradien dan reuse fitur yang lebih baik. Struktur ini mendorong efisiensi parameter dan memberikan efek regularisasi implisit yang membantu mencegah overfitting, terutama ketika data terbatas. Dengan semua kelebihan ini, DenseNet menjadi arsitektur yang sangat menjanjikan untuk klasifikasi pneumonia pada citra X-ray dada [17].

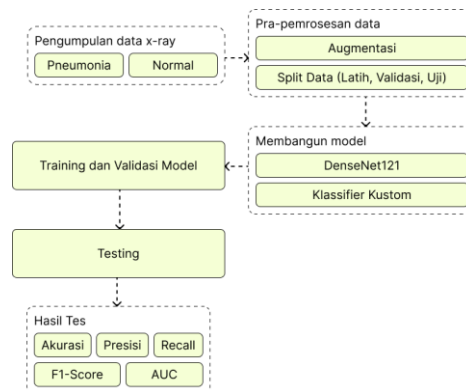
Meski begitu, tinjauan terhadap penelitian-penelitian sebelumnya mengungkapkan bahwa banyak studi hanya memanfaatkan DenseNet sebagai *feature extractor* atau model pralatih tanpa mengeksplorasi lebih dalam strategi *fine-tuning* yang digunakan. Sebagian besar pendekatan *fine-tuning* dilakukan secara statis atau arbitrer—misalnya, membekukan sebagian besar *layer* awal dan hanya melatih *layer* akhir—tanpa analisis sistematis terhadap variasi konfigurasi *fine-tuning* yang mungkin berdampak signifikan pada performa model. Padahal, studi terbaru seperti oleh Davila et al. (2024) [18] menunjukkan bahwa pemilihan strategi *fine-tuning* yang tepat dapat memberikan peningkatan performa yang substansial, khususnya pada model DenseNet. Dalam

konteks klasifikasi pneumonia pada citra X-ray dada, belum banyak studi yang secara sistematis membandingkan pengaruh kedalaman fine-tuning pada arsitektur DenseNet, sehingga masih terbuka ruang eksplorasi terhadap konfigurasi optimal yang dapat menghasilkan generalisasi model yang lebih baik.

Oleh karena itu, penelitian ini bertujuan untuk mengisi kesenjangan tersebut dengan mengkaji dan mengoptimalkan strategi fine-tuning DenseNet secara sistematis untuk klasifikasi pneumonia pada citra X-ray dada. Dengan mengevaluasi berbagai konfigurasi fine-tuning dari model pralatih DenseNet, diharapkan dapat ditemukan strategi yang paling efektif dalam meningkatkan akurasi dan efisiensi sistem klasifikasi otomatis. Pendekatan ini tidak hanya membangun di atas keberhasilan penelitian terdahulu, tetapi juga memberikan kontribusi baru dalam bentuk panduan praktis bagi pengembangan sistem deteksi pneumonia berbasis deep learning yang lebih presisi dan dapat diandalkan.

2. Metode Penelitian

Penelitian ini memanfaatkan arsitektur Convolutional Neural Network (CNN) dengan menggunakan *pretrained model* DenseNet-121. Pengembangan dan pelatihan model dilakukan menggunakan bahasa pemrograman Python 3 serta library PyTorch. Seluruh proses pengembangan dilaksanakan di platform Google Colab.



Gambar 1. Alur Metodologi Penelitian

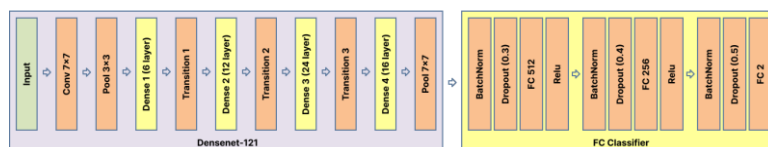
Alur metodologi penelitian secara keseluruhan digambarkan pada Gambar 1. Proses dimulai dengan pengumpulan data citra X-ray dada yang diklasifikasikan ke dalam dua kategori, yaitu pneumonia dan normal. Selanjutnya, data menjalani tahap pra-pemrosesan yang mencakup augmentasi untuk meningkatkan keragaman data serta pembagian dataset ke dalam data latih, validasi, dan uji. Setelah itu, dilakukan pembangunan model dengan memanfaatkan arsitektur DenseNet-121, diikuti dengan penyesuaian klasifier kustom. Model kemudian dilatih dan divalidasi menggunakan data yang telah disiapkan. Setelah pelatihan selesai, model dievaluasi pada data uji dan performanya diukur menggunakan metrik evaluasi seperti akurasi, presisi, recall, F1-score, dan AUC.

Dataset yang digunakan dalam penelitian ini adalah dataset Chest X-Ray Images yang berasal dari Kaggle dan bersumber dari Pusat Medis Wanita dan Anak Guangzhou [19]. Dataset ini terdiri dari 5856 gambar X-ray dada yang dikategorikan ke dalam dua kelas, yaitu Normal dan Pneumonia. Dataset ini telah dibagi ke dalam tiga subset utama, yaitu pelatihan, validasi, dan pengujian. Pada subset pelatihan, terdapat 3875 gambar Pneumonia dan 1341 gambar Normal. Sementara itu, subset validasi terdiri dari 8 gambar untuk masing-masing kelas. Pada subset pengujian, terdapat 390 gambar Pneumonia dan 234 gambar Normal.

Dataset yang digunakan memiliki ketidakseimbangan kelas, khususnya pada subset pengujian, di mana sampel Pneumonia hampir tiga kali lebih banyak dari kelas Normal. Untuk mengatasinya,

dilakukan oversampling pada subset pelatihan dengan mereplikasi sampel kelas Normal hingga seimbang dengan kelas Pneumonia. Selain itu, diterapkan augmentasi data secara eksklusif pada subset pelatihan untuk meningkatkan generalisasi model [20].

Proses augmentasi meliputi penyesuaian ukuran gambar (sisi terpendek menjadi 256 piksel), center cropping menjadi 224×224 piksel, serta transformasi seperti *flip* horizontal dan vertikal, rotasi hingga $\pm 20^\circ$, skala 0,7–1,3, dan *color jitter* (kecerahan 0,5–2,0). Setelah itu, gambar dikonversi ke tensor dan dinormalisasi. Sementara itu, pada subset validasi dan pengujian, preprocessing lebih sederhana, hanya mencakup penyesuaian ukuran, center cropping, konversi ke tensor, dan normalisasi.



Gambar 2. Arsitektur Model yang Digunakan

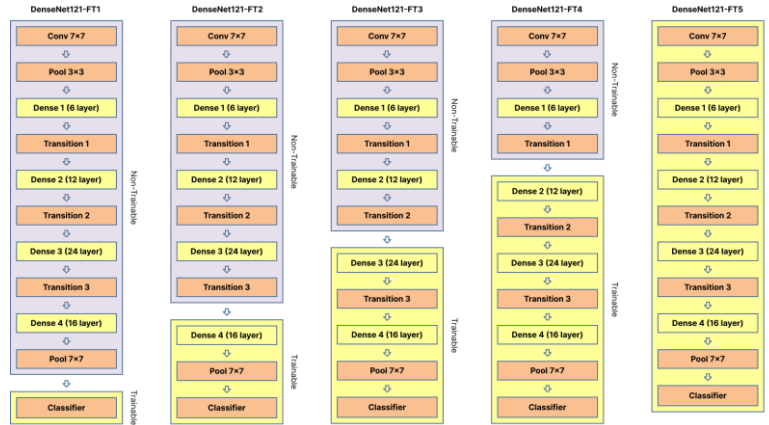
Gambar 2 memperlihatkan arsitektur model yang digunakan, yaitu DenseNet-121 yang telah dilatih sebelumnya pada ImageNet sebagai *backbone* untuk klasifikasi pneumonia pada citra X-ray dada. DenseNet-121 dipilih karena arsitekturnya yang efisien, dengan koneksi padat antar lapisan yang memungkinkan aliran informasi dan gradien secara optimal, mengurangi risiko *vanishing gradient* dan mendorong *reuse* fitur [17].

Setiap *layer* dalam DenseNet menerima input dari semua layer sebelumnya dalam blok yang sama, sehingga mendukung ekstraksi fitur yang kuat tanpa menambah kompleksitas parameter secara signifikan. Arsitektur ini terbukti efektif untuk klasifikasi citra medis, termasuk pneumonia, dengan kinerja tinggi dan efisiensi yang baik [17].

Untuk menyesuaikan DenseNet-121 dengan klasifikasi biner, bagian classifier standarnya diganti dengan *classifier* kustom. Modifikasi ini mencakup penambahan *batch normalization* dan *dropout* untuk mengurangi *overfitting*, serta tiga lapisan *fully connected* berturut-turut (512, 256, dan 2 neuron) dengan aktivasi ReLU pada dua lapisan pertama dan sigmoid pada output untuk menghasilkan prediksi dua kelas, yaitu Normal dan Pneumonia.

Selama tahap pelatihan, model dilatih menggunakan *batch size* 32 dan *weight decay* 0.00001 untuk menjaga efisiensi komputasi dan mengurangi risiko *overfitting*. *Transfer learning* menjadi pendekatan efektif saat data pelatihan terbatas, karena memungkinkan pemanfaatan fitur dari model yang telah dilatih sebelumnya, seperti DenseNet-121 pada ImageNet. Meskipun citra X-ray berbeda dengan gambar alami, fitur dasar dari pelatihan awal tetap relevan dan dapat disesuaikan untuk klasifikasi pneumonia [13].

Dalam studi ini, *fine-tuning* dilakukan dengan membuka (*unfreeze*) sebagian lapisan DenseNet-121 secara bertahap, mulai dari hanya melatih bagian *classifier* hingga seluruh *dense block*. Arsitektur DenseNet-121 yang terdiri dari empat *dense block* digunakan sebagai *backbone*, dan proses *fine-tuning* dibagi menjadi lima level untuk mengevaluasi pengaruh kedalaman pelatihan terhadap performa model.



Gambar 3. Skema Pembagian Tingkat *Fine-Tuning* Untuk Eksperimen

Gambar 3 menunjukkan skema lima konfigurasi fine-tuning yang diterapkan, yaitu hanya melatih classifier (Baseline), membuka dan melatih Blok 4, Blok 3–4, Blok 2–4, hingga seluruh layer pretrained (Full Fine-Tuning). Setiap peningkatan kedalaman fine-tuning menghasilkan jumlah trainable parameters yang lebih besar, yakni sekitar 660 ribu parameter pada Baseline, 2.82 juta pada Blok 4, 6.18 juta pada Blok 3–4, 7.24 juta pada Blok 2–4, dan 7.61 juta parameter saat seluruh layer dilibatkan. Perbedaan jumlah parameter mencerminkan peningkatan kompleksitas dan beban komputasi, baik dari sisi waktu pelatihan maupun penggunaan memori GPU.

Evaluasi performa model dilakukan menggunakan metrik presisi, recall, F1 score, akurasi, dan AUC. Nilai-nilai ini dihitung berdasarkan empat komponen utama: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Adapun rumus evaluasi yang digunakan adalah:

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2)$$

$$\text{F1Score} = \frac{2 \times (\text{Recall} \times \text{Presisi})}{\text{Recall} + \text{Presisi}} \quad (3)$$

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Sedangkan AUC (*Area Under the ROC Curve*) mengukur performa model pada berbagai ambang klasifikasi dengan memplot *True Positive Rate* terhadap *False Positive Rate*.

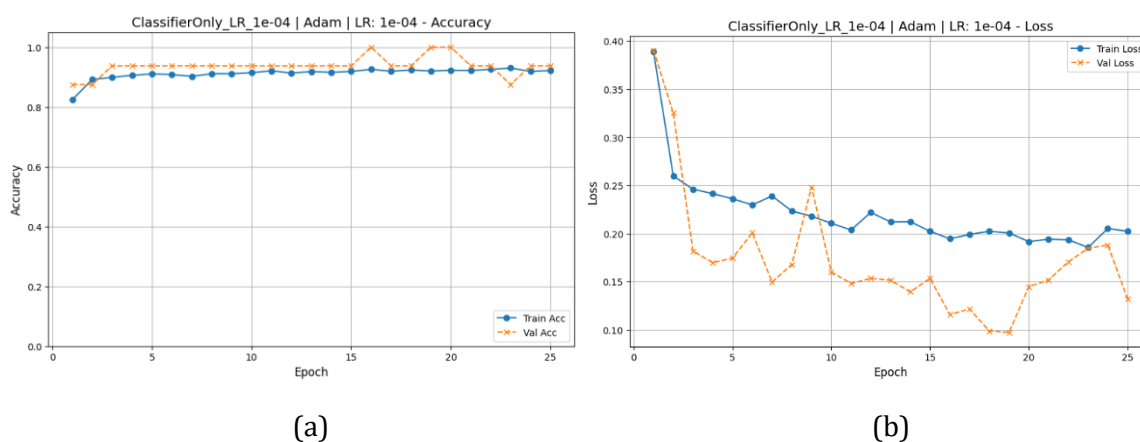
3. Hasil

Penelitian ini berfokus pada optimasi model CNN berbasis *transfer learning* untuk klasifikasi pneumonia pada citra X-Ray dada, dengan memanfaatkan arsitektur DenseNet dan menerapkan berbagai strategi *fine-tuning*. Pendekatan dilakukan dengan melatih ulang (*fine-tune*) bagian tertentu dari model *pretrained* secara bertahap. Setiap strategi tidak hanya melatih *layer* klasifikasi, tetapi juga satu atau lebih blok dari *layer pretrained* yang dibuka (*unfreeze*). Strategi yang diuji mencakup: hanya melatih *classifier* (Baseline), membuka dan melatih Blok 4, Blok 3–4, Blok 2–4, hingga seluruh *layer pretrained* (Full Fine-Tuning). Setiap pendekatan dievaluasi menggunakan *optimizer* Adam dengan tiga nilai *learning rate* (0.01, 0.001, dan 0.0001), selama 25 *epoch*, menggunakan metrik akurasi, presisi, recall, F1-score, dan ROC-AUC.

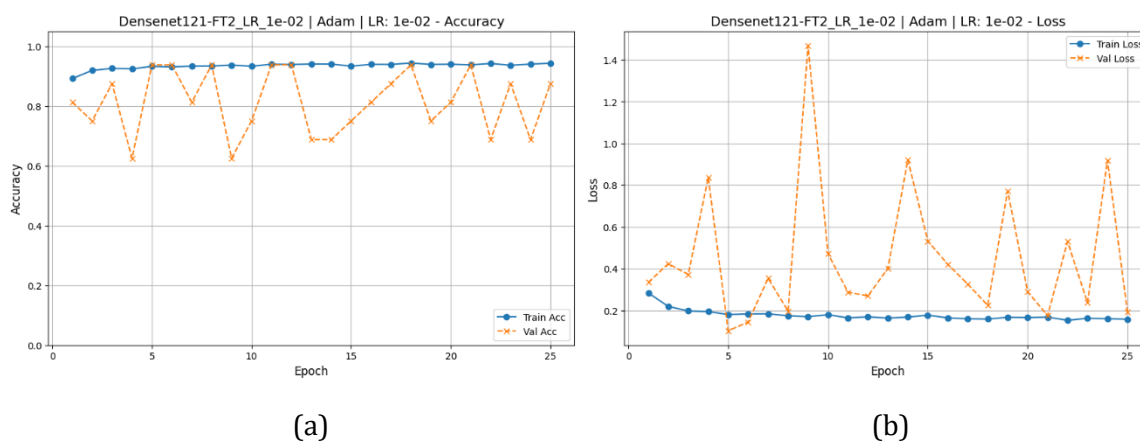
Tabel 1. Hasil Evaluasi Terbaik Tiap Strategi

Strategi	Learning Rate	Akurasi	Presisi	Recall	F1-Score	ROC-AUC
<i>Baseline</i>	0.0001	90,38%	90,24%	94,87%	92,50%	95,64%
<i>Unfreze Block 4</i>	0.01	91,67%	91,63%	95,38%	93,47%	96,18%
<i>Unfreze Block 3 - 4</i>	0.001	94,39%	93,61%	97,69%	95,61%	98,04%
<i>Unfreze Block 2 - 4</i>	0.0001	93,91%	93,56%	96,92%	95,21%	97,38%
<i>Full Fine-Tuning</i>	0.0001	92,79%	95,51%	92,82%	94,15%	97,41%

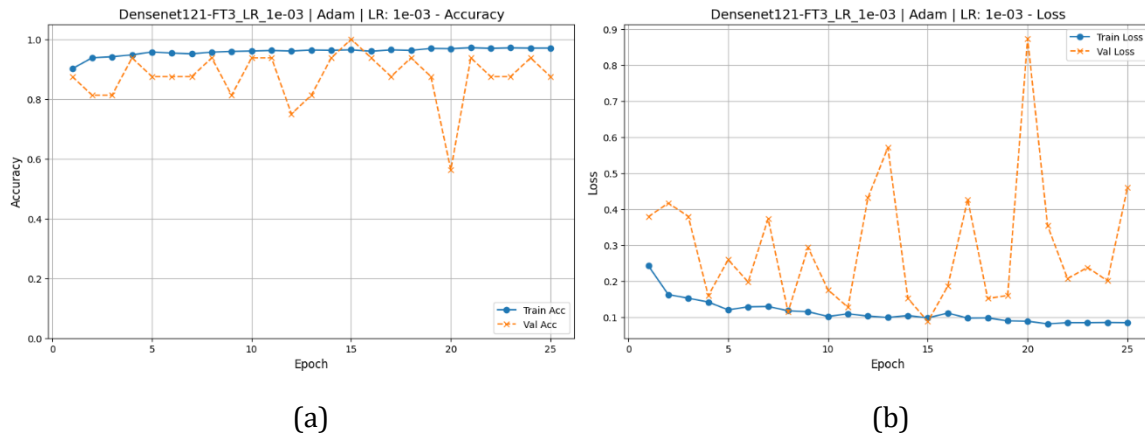
Tabel 1 menunjukkan bahwa strategi *Unfreeze Block 3–4* memberikan performa terbaik dengan akurasi 94,39%, F1-score 95,61%, dan ROC-AUC 98,04%. Strategi *Unfreeze Block 2–4* menyusul dengan hasil yang juga tinggi, yakni akurasi 93,91% dan ROC-AUC 97,38%. *Full Fine-Tuning* mencatat akurasi 92,79% dan ROC-AUC 97,41%, menunjukkan efektivitas meski sedikit di bawah dua strategi sebelumnya. Sementara itu, *Unfreeze Block 4* dan *Baseline* menunjukkan performa yang lebih rendah, masing-masing dengan akurasi 91,67% dan 90,38%. Hasil ini menunjukkan bahwa melibatkan lebih banyak blok *pretrained* dalam proses pelatihan cenderung meningkatkan performa model, dengan hasil optimal pada pelatihan ulang blok 3-4 bersama *classifier*.

**Gambar 4.** Akurasi dan loss saat training dan validasi untuk strategi *Baseline*

Gambar 4 menunjukkan performa strategi *Baseline* dengan *learning rate* 0.0001. Akurasi pelatihan meningkat dari sekitar 82% menjadi 92%, sementara akurasi validasi cukup stabil di kisaran 91–100%. *Loss* pelatihan menurun dari 0,39 ke 0,20, dan *loss* validasi tetap rendah di kisaran 0,10–0,20, menandakan pelatihan yang stabil dan tanpa gejala *overfitting*.

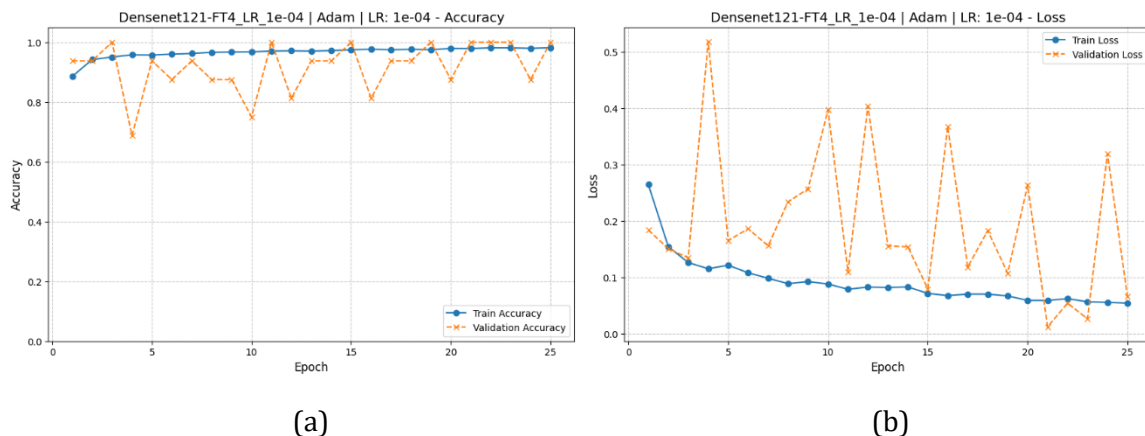
**Gambar 5.** Akurasi dan loss saat training dan validasi untuk strategi *Unfreeze Block 4*

Pada strategi *Unfreeze Block 4* dengan *learning rate* 0.01 pada Gambar 5, akurasi pelatihan mencapai lebih dari 95%, namun akurasi validasi berfluktuasi dari 62% hingga 95%. *Loss* pelatihan konsisten rendah, sementara *loss* validasi mengalami fluktuasi tajam hingga sempat menyentuh 1.4. Meskipun demikian, fluktuasi ini tidak menunjukkan tren peningkatan yang terus-menerus, sehingga *overfitting* tidak terdeteksi secara signifikan. Fluktuasi tersebut lebih mencerminkan variasi antar *batch* validasi, dan model tetap menunjukkan kemampuan generalisasi yang cukup baik.



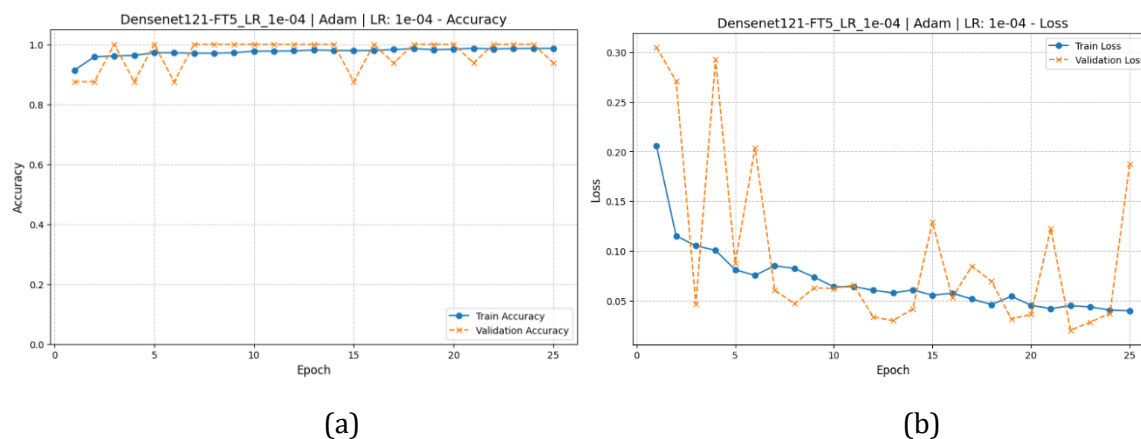
Gambar 6. Akurasi dan loss saat training dan validasi untuk strategi *Unfreeze Block 3 - 4*

Kemudian, Gambar 6 menggambarkan performa strategi *Unfreeze Block 3-4* dengan *learning rate* 0.001. Akurasi pelatihan stabil di atas 95%, dan akurasi validasi berfluktuasi tetapi tetap menyentuh angka di atas 90% pada beberapa *epoch* dan sekali menyentuh 100%. *Loss* pelatihan rendah dan stabil di bawah 0.2, sedangkan *loss* validasi menunjukkan variasi yang lebih tinggi, tetapi tanpa tren kenaikan yang progresif. Hal ini menunjukkan bahwa model tetap mampu melakukan generalisasi dengan baik, didukung oleh hasil evaluasi pada data uji.



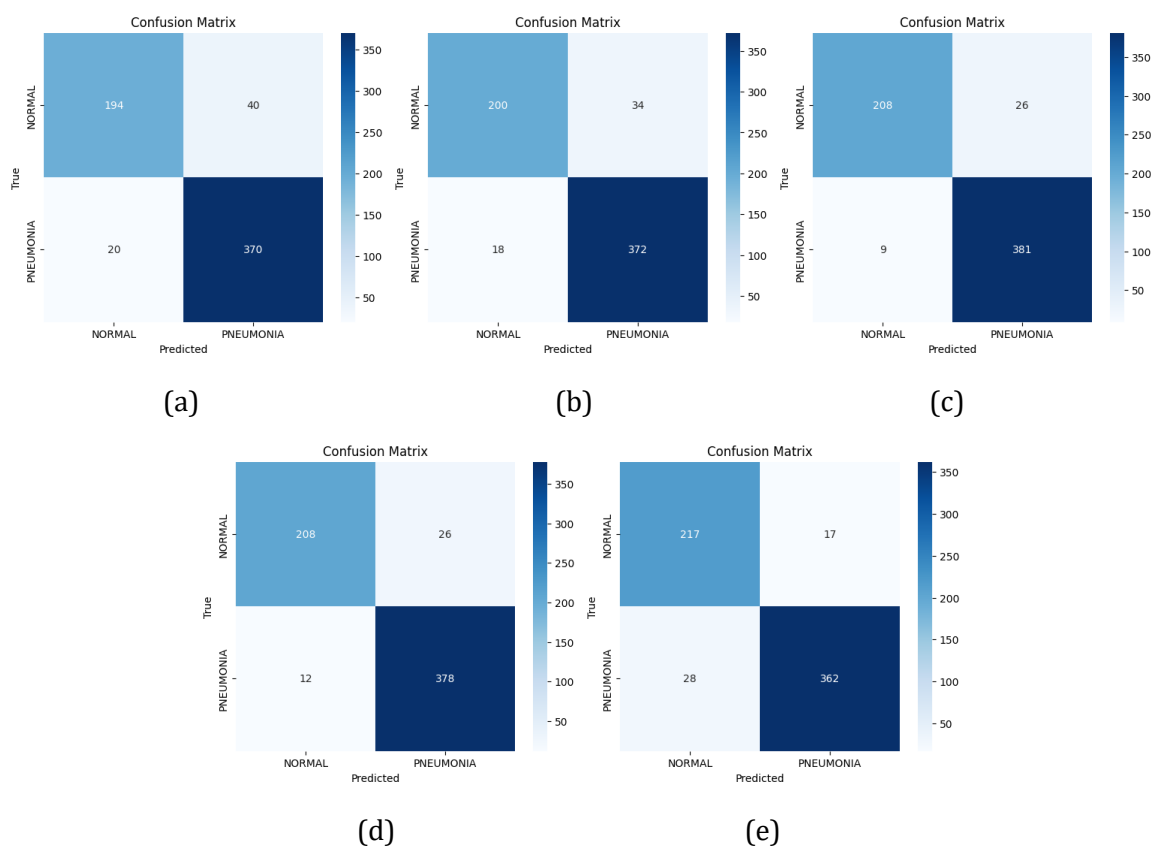
Gambar 7. Akurasi dan loss saat training dan validasi untuk strategi *Unfreeze Block 2 - 4*

Strategi *Unfreeze Block 2-4* dengan *learning rate* 0.0001 yang divisualisasikan pada Gambar 7 memperlihatkan akurasi pelatihan yang sangat tinggi, melampaui 98%. Akurasi validasi berfluktuasi antara 70% hingga beberapa kali menyentuh 100%. *Loss* pelatihan turun hingga di bawah 0.1, dan *loss* validasi menunjukkan variasi tajam namun tidak menunjukkan pola peningkatan, yang berarti tidak terdapat *overfitting* yang signifikan.



Gambar 8. Akurasi dan loss saat training dan validasi untuk strategi Full Fine-Tuning

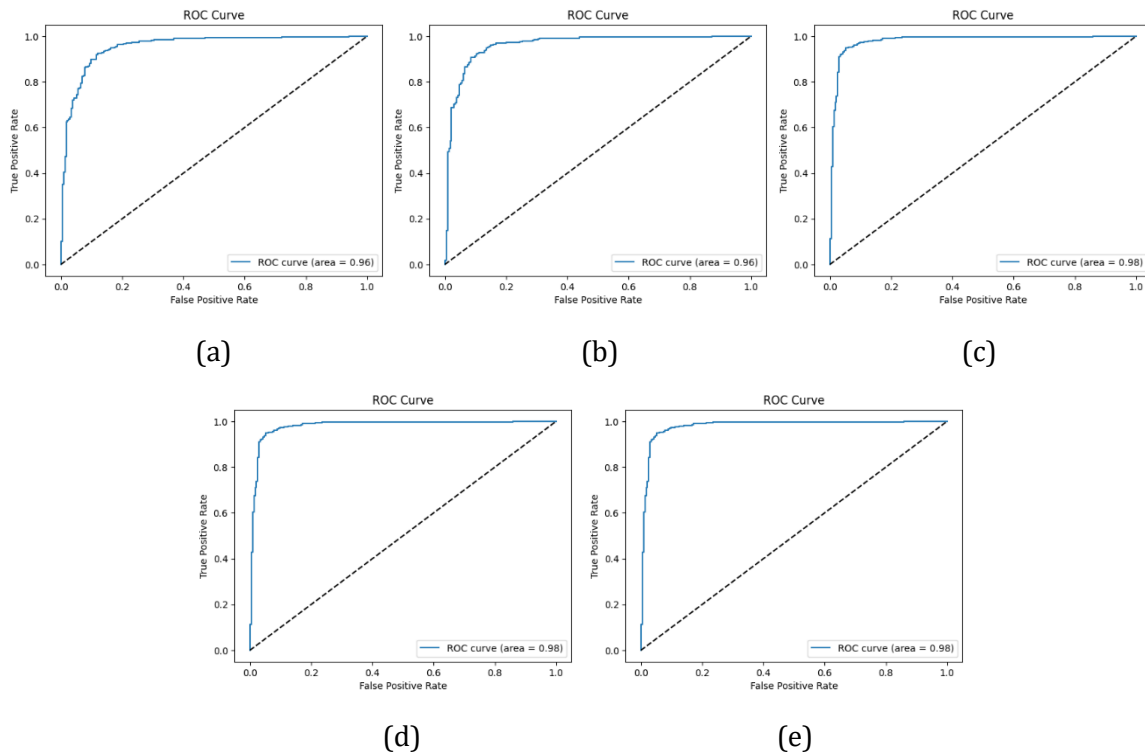
Terakhir, Pada strategi *Full Fine-Tuning* dengan *learning rate* 0.0001 pada Gambar 10 memperlihatkan akurasi pelatihan yang sangat stabil dan mendekati 100%, sementara akurasi validasi berfluktuasi tetapi sebagian besar tetap berada di atas 90% dan sering menyentuh 100%. *Loss* pelatihan tetap rendah di bawah 0.1, dan *loss* validasi menunjukkan variasi tanpa pola kenaikan. Hal ini mengindikasikan pelatihan yang stabil dan kemampuan generalisasi model yang baik.



Gambar 9. Hasil Confusion Matrix Untuk Baseline (a), Unfreeze Block 4 (b), Unfreeze Block 3-4 (c), Unfreeze Block 2-4 (d), dan Full-Fine Tuning (e)

Gambar 9 menunjukkan *confusion matrix* dari lima strategi *fine-tuning* pada model DenseNet-121. Model *baseline* (a) menghasilkan 194 prediksi benar untuk kelas NORMAL dan 370 untuk PNEUMONIA, dengan kesalahan 40 dan 20 masing-masing. Pada strategi *Unfreeze Block 4* (b),

performa meningkat menjadi 200 benar NORMAL dan 372 PNEUMONIA. Strategi *Unfreeze Block 3-4* (c) lebih baik lagi, dengan 208 NORMAL dan 381 PNEUMONIA. *Unfreeze Block 2-4* (d) mempertahankan 208 NORMAL dan sedikit menurun di PNEUMONIA menjadi 378. Sementara itu, *Full-Fine Tuning* (e) mencapai prediksi benar tertinggi untuk NORMAL sebanyak 217, namun PNEUMONIA turun ke 362.



Gambar 10. Hasil kurva ROC untuk Untuk Baseline (a), Unfreeze Block 4 (b), Unfreeze Block 3-4 (c), Unfreeze Block 2-4 (d), dan Full-Fine Tuning (e)

Gambar 10 menyajikan kurva *Receiver Operating Characteristic* (ROC) dari lima strategi pelatihan yang berbeda, yaitu *Baseline* (a), *Unfreeze Block 4* (b), *Unfreeze Block 3-4* (c), *Unfreeze Block 2-4* (d), dan *Full Fine-Tuning* (e). Setiap sub-gambar menampilkan *True Positive Rate* (TPR) terhadap *False Positive Rate* (FPR) dengan garis diagonal putus-putus sebagai garis referensi klasifikasi acak. Area di bawah kurva (AUC) merupakan indikator kinerja model, di mana nilai AUC mendekati 1 menunjukkan kinerja klasifikasi yang sangat baik. Hasil visual menunjukkan bahwa strategi *Baseline* (a) dan *Unfreeze Block 4* (b) mencapai AUC sebesar 0.96. Sementara itu, strategi *Unfreeze Block 3-4* (c), *Unfreeze Block 2-4* (d) dan *Full Fine-Tuning* (e) menunjukkan kinerja yang sedikit lebih unggul dengan AUC sebesar 0.98. Konsistensi nilai AUC yang tinggi di seluruh strategi mengindikasikan bahwa model secara umum memiliki kemampuan diskriminasi yang kuat antara kelas positif dan negatif, meskipun ada sedikit peningkatan performa dengan strategi *fine-tuning* yang lebih komprehensif.

4. Pembahasan

Berdasarkan hasil evaluasi pada Tabel 1, strategi *Unfreeze Block 3-4* menunjukkan performa terbaik dengan akurasi mencapai 94,39 %, F1-score 95,61 %, dan ROC-AUC sebesar 98,04 %. Kenaikan performa ini mengindikasikan bahwa membuka lapisan dari blok 3 sampai 4 pada pretrained DenseNet-121 memungkinkan model untuk menyesuaikan fitur tingkat menengah hingga tinggi secara lebih optimal terhadap karakteristik citra pneumonia pada X-Ray dada. Sebaliknya, strategi *Baseline* (hanya melatih *classifier*) memiliki akurasi terendah (90,38 %), yang

menegaskan bahwa melatih *classifier* saja kurang memadai untuk mentransfer pengetahuan dari dataset ImageNet ke domain medis dengan spektrum fitur berbeda.

Visualisasi akurasi dan loss pada Gambar 4–8 memperkuat temuan ini. Pada strategi *Unfreeze Block 3–4*, akurasi pelatihan dan validasi relatif stabil di atas 90% tanpa tren kenaikan loss validasi—menunjukkan tidak adanya *overfitting* signifikan meski sejumlah lapisan dibuka. Sementara itu, pada *Unfreeze Block 4* dan *Unfreeze Block 2–4*, fluktuasi akurasi validasi dan loss validasi yang lebih besar (hingga 1.4) menandakan bahwa membuka terlalu sedikit (hanya blok 4) atau terlalu banyak lapisan (blok 2–4) dapat menyebabkan ketidakstabilan pada *batch* tertentu, meski tidak selalu mempengaruhi kemampuan generalisasi akhir.

Analisis *confusion matrix* pada Gambar 9 menggarisbawahi bahwa *Unfreeze Block 3–4* meningkatkan prediksi benar untuk kedua kelas—NORMAL dan PNEUMONIA—lebih baik dibanding strategi lainnya. Hal ini berarti model mampu menangkap pola tekstur dan kerapatan yang membedakan paru sehat dan yang terinfeksi pneumonia. Meski *Full Fine-Tuning* mencapai prediksi benar NORMAL tertinggi (217), penurunan pada prediksi PNEUMONIA (362) mengindikasikan trade-off antara spesifisitas dan sensitivitas ketika semua lapisan dioptimasi ulang.

Kurva ROC pada Gambar 10 menunjukkan konsistensi kinerja diskriminasi yang tinggi di semua strategi, tetapi sedikit lebih unggul pada *Unfreeze Block 3–4*, *Unfreeze Block 2–4*, dan *Full Fine-Tuning* dengan AUC 0.98. Perbedaan marginal ini menegaskan bahwa fine-tuning lapisan menengah hingga akhir (blok 3–4) sudah cukup untuk mendapatkan representasi yang kaya tanpa harus melatih ulang seluruh jaringan, sehingga lebih efisien dari segi komputasi.

Hasil penelitian ini menunjukkan performa yang lebih unggul dibandingkan studi sebelumnya. Model DenseNet-121 berhasil mencapai akurasi sebesar 94,39%, melampaui studi [21] yang menggunakan CNN dengan akurasi 87,82%, serta studi [22] yang menggunakan ResNet-50 dengan akurasi 94% setelah melalui 100 epoch. Sebaliknya, akurasi pada penelitian ini dicapai hanya dalam 25 epoch, yang menunjukkan efisiensi komputasi yang lebih tinggi tanpa mengorbankan performa model.

Namun demikian, penelitian ini memiliki beberapa keterbatasan. Model hanya diuji menggunakan data dari satu sumber, yaitu Pusat Medis Guangzhou, sehingga performanya belum dapat dipastikan jika diterapkan pada data dari rumah sakit atau populasi berbeda. Selain itu, validasi belum mencakup *external test set*, yang penting untuk mengevaluasi generalisasi model di luar data yang digunakan dalam pelatihan dan validasi internal.

5. Penutup

Penelitian ini menunjukkan bahwa penerapan strategi fine-tuning selektif pada arsitektur DenseNet-121 dapat meningkatkan performa klasifikasi pneumonia secara signifikan pada citra X-Ray dada. Strategi *Unfreeze Block 3–4* memberikan keseimbangan optimal antara stabilitas pelatihan, kemampuan generalisasi, dan efisiensi komputasi, dengan capaian akurasi 94,39%, F1-score 95,61%, dan ROC-AUC 98,04%. Temuan ini menegaskan bahwa membuka sejumlah blok pretrained secara terarah lebih efektif dibandingkan hanya melatih *classifier* atau melakukan full fine-tuning.

Kontribusi utama studi ini adalah menunjukkan bahwa fine-tuning pada blok 3–4 DenseNet-121 menghasilkan kinerja optimal tanpa memerlukan pelatihan ulang seluruh jaringan, sehingga menawarkan solusi efisien untuk klasifikasi medis berbasis deep learning dengan keterbatasan sumber daya. Untuk penelitian selanjutnya, disarankan mengeksplorasi teknik regularisasi lanjutan, seperti dynamic dropout atau adaptive L2 regularization, serta memperluas cakupan dataset dengan variasi populasi pasien guna mengevaluasi ketahanan model dalam konteks klinis yang lebih beragam.

Referensi

- [1] R. G. Bender *et al.*, "Global, regional, and national incidence and mortality burden of non-COVID-19 lower respiratory infections and aetiologies, 1990–2021: a systematic analysis from the Global Burden of Disease Study 2021," *Lancet Infect Dis*, vol. 24, no. 9, pp. 974–1002, Sep. 2024, doi: 10.1016/S1473-3099(24)00176-2.
- [2] F. M. Qaimkhani, M. Hussain, Y. Shiren, and J. Xingfang, "Pneumonia Detection Using Deep Learning Methods," *International Journal Of Scientific Advances*, vol. 3, no. 3, 2022, doi: 10.51542/ijscia.v3i3.32.
- [3] M. Husna, F. Dewi Pertiwi, and A. Saputra Nasution, "FAKTOR-FAKTOR YANG BERHUBUNGAN DENGAN KEJADIAN PNEUMONIA PADA BALITA DI PUSKESMAS SEMPLAK KOTA BOGOR 2020," *PROMOTOR*, vol. 5, no. 3, pp. 273–280, May 2022, doi: 10.32832/pro.v5i3.6168.
- [4] S. Showkat and S. Qureshi, "Efficacy of Transfer Learning-based ResNet models in Chest X-ray image classification for detecting COVID-19 Pneumonia," *Chemometrics and Intelligent Laboratory Systems*, vol. 224, p. 104534, May 2022, doi: 10.1016/j.chemolab.2022.104534.
- [5] W. B. Gefter, B. A. Post, and H. Hatabu, "Commonly Missed Findings on Chest Radiographs," *Chest*, vol. 163, no. 3, pp. 650–661, Mar. 2023, doi: 10.1016/j.chest.2022.10.039.
- [6] D. J. Mollura *et al.*, "Artificial Intelligence in Low- and Middle-Income Countries: Innovating Global Health Radiology," *Radiology*, vol. 297, no. 3, pp. 513–520, Dec. 2020, doi: 10.1148/radiol.2020201434.
- [7] M. M. Zulfa and C. Sri Kusuma Aditya, "CATARACT CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORK (CNN) INCEPTION RESNETV2," *Jurnal Teknik Informatika (Jutif)*, vol. 5, no. 5, pp. 1299–1307, Oct. 2024, doi: 10.52436/1.jutif.2024.5.5.2340.
- [8] E. C. Yaurentius, T. R. D. Saputri, E. Tanuwijaya, and R. E. Sutanto, "COMPARATIVE STUDY OF CNN-BASED ARCHITECTURES ON EYE DISEASES CLASSIFICATION USING FUNDUS IMAGES TO AID OPHTHALMOLOGIST," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 1, pp. 249–257, Feb. 2025, doi: 10.52436/1.jutif.2025.6.1.3699.
- [9] T. Rahman *et al.*, "Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray," *Applied Sciences*, vol. 10, no. 9, p. 3233, May 2020, doi: 10.3390/app10093233.
- [10] N. Nurhaeni, S. E. Prastya, A. Hidayat, and F. N. Anisa, "Pemodelan Sistem Deteksi Parasit Malaria pada Citra Mikroskopis Sel Darah Menggunakan Metode Deep Learning," *SMATIKA JURNAL*, vol. 14, no. 02, pp. 409–416, Dec. 2024, doi: 10.32664/smatika.v14i02.1475.
- [11] A. W. Salehi *et al.*, "A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope," *Sustainability*, vol. 15, no. 7, p. 5930, Mar. 2023, doi: 10.3390/su15075930.
- [12] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, "Transfer learning for medical image classification: a literature review," *BMC Med Imaging*, vol. 22, no. 1, p. 69, Dec. 2022, doi: 10.1186/s12880-022-00793-7.
- [13] Z. Zhao, L. Alzubaidi, J. Zhang, Y. Duan, and Y. Gu, "A comparison review of transfer learning and self-supervised learning: Definitions, applications, advantages and limitations," *Expert Syst Appl*, vol. 242, p. 122807, May 2024, doi: 10.1016/j.eswa.2023.122807.

- [14] M. Salehi, R. Mohammadi, H. Ghaffari, N. Sadighi, and R. Reiazi, "Automated detection of pneumonia cases using deep transfer learning with paediatric chest X-ray images," *Br J Radiol*, vol. 94, no. 1121, May 2021, doi: 10.1259/bjr.20201263.
- [15] M. E. H. Chowdhury *et al.*, "Can AI Help in Screening Viral and COVID-19 Pneumonia?," *IEEE Access*, vol. 8, pp. 132665–132676, 2020, doi: 10.1109/ACCESS.2020.3010287.
- [16] A. Alhudhaif, K. Polat, and O. Karaman, "Determination of COVID-19 pneumonia based on generalized convolutional neural network model from chest X-ray images," *Expert Syst Appl*, vol. 180, p. 115141, Oct. 2021, doi: 10.1016/j.eswa.2021.115141.
- [17] T. Zhou, X. Ye, H. Lu, X. Zheng, S. Qiu, and Y. Liu, "Dense Convolutional Network and Its Application in Medical Image Analysis," *Biomed Res Int*, vol. 2022, no. 1, Jan. 2022, doi: 10.1155/2022/2384830.
- [18] A. Davila, J. Colan, and Y. Hasegawa, "Comparison of fine-tuning strategies for transfer learning in medical image classification," *Image Vis Comput*, vol. 146, p. 105012, Jun. 2024, doi: 10.1016/j.imavis.2024.105012.
- [19] Paul Mooney, "Chest X-Ray Images (Pneumonia)," Kaggle. Accessed: Mar. 24, 2025. [Online]. Available: <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>
- [20] A. A. Handoko, M. A. Rosid, and U. Indahyanti, "Implementasi Convolutional Neural Network (CNN) Untuk Pengenalan Tulisan Tangan Aksara Bima," *SMATIKA JURNAL*, vol. 14, no. 01, pp. 96–110, Jul. 2024, doi: 10.32664/smatika.v14i01.1196.
- [21] V. Praskatama, C. A. Sari, E. H. Rachmawanto, and N. Mohd Yaacob, "PNEUMONIA PREDICTION USING CONVOLUTIONAL NEURAL NETWORK," *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 5, pp. 1217–1226, Oct. 2023, doi: 10.52436/1.jutif.2023.4.5.1353.
- [22] Y. Azhar, W. Priyo Wicaksono, and Z. Sari, "PNEUMONIA DIAGNOSIS THROUGH DEEP LEARNING: RESNET50V2 MODEL IMPLEMENTATION," vol. 13, no. 2, 2024, doi: 10.23887/v13i2.72068.