

## EKSPLORASI PADA PEMETAAN KLASIFIKASI RADIOGRAF TORAKS PENYAKIT PARU-PARU MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK (CNN)

Andreas Rezeki Zai<sup>1</sup>, Bambang Suhardi<sup>2</sup>, Suryandika Tri Nowo<sup>3</sup>, Rika Rosnelly<sup>4</sup>, Adil Setiawan<sup>5</sup>

<sup>1, 2, 3, 4, 5</sup> Universitas Potensi Utama, Medan, Indonesia

<sup>1</sup>andreasrezekizai@gmail.com, <sup>2</sup>bambangsuwardi@uinsu.ac.id, <sup>3</sup>suryandika315@gmail.com, <sup>4</sup>rikarosnelly@gmail.com, <sup>5</sup>adio165@gmail.com

### ABSTRAK

Radiograf toraks (CXR) merupakan alat penting dalam diagnosis penyakit paru, namun interpretasinya memerlukan keahlian khusus dan berpotensi menimbulkan bias. Penelitian ini bertujuan mengeksplorasi kinerja lima arsitektur Convolutional Neural Network (CNN) berbasis transfer learning, yaitu VGG16, ResNet50, EfficientNetB0, DenseNet121, dan MobileNetV2, dalam mengklasifikasikan lima kelas penyakit paru-paru: bacterial pneumonia, COVID-19, tuberculosis, viral pneumonia, dan normal. Dataset yang digunakan dilengkapi dengan preprocessing CLAHE-RGB, augmentasi data, serta penanganan ketidakseimbangan kelas menggunakan class weighting. Evaluasi dilakukan dengan empat skenario epoch (5, 10, 15, dan 30), serta menggunakan metrik akurasi, precision, recall, F1-score, dan confusion matrix. Hasil menunjukkan bahwa model VGG16 pada epoch ke-15 memberikan performa terbaik dengan akurasi 93,95% dan F1-score 0,94. Penelitian ini menunjukkan bahwa kombinasi preprocessing yang tepat dan arsitektur CNN yang sesuai mampu meningkatkan akurasi klasifikasi penyakit paru secara signifikan.

Kata Kunci— Convolutional Neural Network, Citra CXR, VGG16, CLAHE, Penyakit Paru.

### ABSTRACT

*Chest radiography (CXR) is a vital tool in diagnosing pulmonary diseases, yet its interpretation often requires expert analysis and may involve subjectivity. This study explores the performance of five Convolutional Neural Network (CNN) architectures: VGG16, ResNet50, EfficientNetB0, DenseNet121, and MobileNetV2 for classifying five categories of lung conditions: bacterial pneumonia, COVID-19, tuberculosis, viral pneumonia, and normal. The dataset underwent preprocessing using CLAHE-RGB enhancement, data augmentation, and class balancing with class weighting. Each model was trained using four epoch scenarios (5, 10, 15, and 30) and evaluated based on accuracy, precision, recall, F1-score, and confusion matrix. The results indicate that VGG16 with 15 epochs achieved the best performance, reaching 93.95% accuracy and 0.94 F1-score. This study demonstrates that combining appropriate preprocessing techniques with suitable CNN architectures significantly enhances classification performance for pulmonary disease detection.*

Keywords— Convolutional Neural Network, CXR images, VGG16, Transfer Learning, CLAHE, Lung Disease.

## I. PENDAHULUAN

Ketepatan diagnosis penyakit paru melalui citra radiograf toraks menjadi tantangan utama dalam dunia medis, mengingat tingginya prevalensi dan variasi penyakit yang menyerang sistem pernapasan serta keterbatasan sumber daya radiolog di banyak fasilitas kesehatan. Dalam praktik klinis, diagnosis penyakit paru menuntut kecepatan dan ketepatan tinggi untuk menunjang pengambilan keputusan medis yang efektif dan efisien. Salah satu alat diagnostik yang paling umum digunakan adalah citra radiograf toraks atau *chest X-ray* (CXR), yang memiliki keunggulan dari segi ketersediaan, biaya rendah, dan kemudahan penggunaan di berbagai tingkat layanan kesehatan, termasuk di fasilitas primer.

Namun demikian, proses interpretasi citra CXR secara manual masih sangat bergantung pada keahlian dan pengalaman radiolog. Hal ini seringkali menimbulkan variabilitas subjektif yang berpotensi menyebabkan inkonsistensi dalam hasil diagnosis, terutama ketika dihadapkan pada volume data yang besar atau kasus dengan gejala yang tidak khas. Dalam beberapa tahun terakhir, pendekatan berbasis *artificial intelligence* (AI), khususnya *deep learning* dengan algoritma *Convolutional Neural Network* (CNN), mulai banyak diterapkan untuk mengotomatiskan proses identifikasi dan klasifikasi penyakit paru melalui citra radiografi. CNN dikenal memiliki kemampuan tinggi dalam mengekstraksi fitur spasial kompleks dari citra medis dan telah terbukti efektif dalam meningkatkan performa diagnostik.

Studi yang dilakukan oleh Al-Sheikh et al. (2023) menunjukkan bahwa arsitektur CNN seperti VGG16 mampu dikembangkan untuk klasifikasi multi-kelas penyakit paru dengan hasil yang menjanjikan [1]. Penelitian lain oleh Alshmrani et al. (2023) memperkuat temuan tersebut melalui penerapan VGG19 yang dikombinasikan dengan tiga blok CNN untuk mengklasifikasikan berbagai penyakit paru, termasuk COVID-19, pneumonia, tuberkulosis, kanker paru, dan *lung opacity*. Dengan menggunakan lebih dari 80.000 citra CXR yang telah melalui proses normalisasi, augmentasi, dan pengubahan ukuran (*resizing*), model mereka berhasil mencapai akurasi sebesar 96,48%, recall 93,75%, precision 97,56%, dan F1-score sebesar 95,62% [2]. Hasil ini mencerminkan efektivitas pendekatan berbasis CNN dalam menangani kompleksitas klasifikasi multi-kelas pada citra toraks.

Selain pemilihan arsitektur CNN yang sesuai, tahapan *preprocessing* citra juga berperan penting dalam meningkatkan kualitas input dan akurasi model. Salah satu teknik yang banyak digunakan adalah

*Contrast Limited Adaptive Histogram Equalization* (CLAHE), yang bertujuan untuk meningkatkan kontras lokal serta menonjolkan detail penting dalam jaringan paru. El Lel (2023) membuktikan bahwa integrasi CLAHE dengan *center cropping* dan arsitektur CNN modern seperti DenseNet dan EfficientNet mampu meningkatkan performa klasifikasi pada dataset BIMCV dan PadChest secara signifikan [3]. Sejalan dengan itu, Çiğ et al. (2023) mengembangkan pendekatan hibrida dengan menggabungkan CLAHE dan algoritma *Multi-Objective Cuckoo Search* (MOCS), dan mencatat akurasi mencapai 99,16% pada dataset COVID-19 Radiography [4]. Temuan ini menegaskan bahwa optimalisasi *preprocessing* berbasis peningkatan kontras memiliki dampak besar terhadap efektivitas sistem klasifikasi berbasis CNN.

Namun, meskipun banyak model CNN menunjukkan performa tinggi, tidak ada satu arsitektur yang secara universal unggul dalam semua jenis dataset atau kondisi penyakit. Bhatt et al. (2023) melakukan studi komparatif antara VGG16, ResNet50, dan DenseNet untuk deteksi pneumonia, dan menyimpulkan bahwa perbedaan arsitektur dan strategi pelatihan memberikan dampak signifikan terhadap hasil klasifikasi [5]. Penelitian lain oleh Jain et al. (2023) membandingkan model CNN, ResNet, dan *Vision Transformers* (ViT) untuk klasifikasi multi-label, dan menemukan bahwa meskipun ViT memiliki potensi dalam mengolah informasi global, ResNet tetap unggul dalam hal akurasi pada dataset besar seperti NIH ChestX-ray [6]. Sementara itu, Usman et al. (2023) menyoroti performa DenseNet-121 dalam mendeteksi hingga 14 jenis penyakit toraks dan menyatakan bahwa model tersebut mampu menyamai bahkan melampaui CheXNeXt dalam skenario tertentu [7]. Temuan-temuan tersebut menunjukkan pentingnya eksplorasi terhadap berbagai arsitektur CNN dalam konteks klasifikasi multi-kelas citra radiografi.

Berdasarkan latar belakang di atas, jelas bahwa keberhasilan sistem klasifikasi otomatis penyakit paru sangat dipengaruhi oleh pemilihan arsitektur model yang tepat, strategi *preprocessing* yang efektif, dan metode evaluasi yang sesuai dengan karakteristik dataset. Oleh karena itu, penelitian ini bertujuan untuk melakukan studi komparatif terhadap lima arsitektur CNN yang populer dan banyak digunakan dalam literatur, yaitu VGG16, ResNet50, EfficientNetB0, MobileNetV2, dan DenseNet121. Setiap model akan dievaluasi berdasarkan akurasi, presisi, recall, dan F1-score pada dataset citra CXR yang telah melalui tahapan *preprocessing* menggunakan CLAHE, augmentasi data, serta penyeimbangan kelas (*class*

*weighting*). Melalui pendekatan ini, diharapkan penelitian ini dapat memberikan kontribusi signifikan terhadap pemahaman kinerja relatif antar model, serta menjadi dasar dalam pengembangan sistem deteksi penyakit paru otomatis yang lebih akurat, adaptif, dan aplikatif di lingkungan klinis nyata.

## II. TINJAUAN PUSTAKA

### CNN dalam Klasifikasi Penyakit Paru-Paru

*Convolutional Neural Network* (CNN) telah menjadi pendekatan dominan dalam pengembangan sistem klasifikasi penyakit paru-paru berbasis citra radiografi, khususnya pada modalitas *chest X-ray* (CXR). Kemampuannya dalam mengekstraksi fitur spasial dari data visual secara hierarkis dan otomatis menjadikan CNN sangat efektif dalam mengidentifikasi pola patologis pada struktur paru yang sering kali sulit dikenali secara manual.

Salah satu penelitian awal yang mendemonstrasikan potensi CNN dalam klasifikasi multi-kelas penyakit paru dilakukan oleh Ashwini et al. (2021), yang merancang model CNN untuk membedakan tiga kategori utama yaitu pneumonia, fibrosis, dan kondisi normal. Melalui tahapan *preprocessing* dan augmentasi data, serta penerapan teknik *dropout* dan optimasi berbasis algoritma Adam, model tersebut berhasil mencapai akurasi sebesar 96,8%, menunjukkan efektivitas pendekatan CNN pada dataset terbatas [8].

Selanjutnya, Jain et al. (2022) mengembangkan sistem *Computer-Aided Diagnosis* (CAD) berbasis CNN untuk deteksi pneumonia. Mereka memanfaatkan *transfer learning* dengan model pralatih dan menerapkan *fine-tuning* guna menyesuaikan bobot jaringan terhadap karakteristik citra CXR. Studi ini juga mengintegrasikan teknik visualisasi *heatmap* untuk menyoroti area opasitas pada paru-paru, yang memperkuat interpretasi hasil prediksi oleh model [9].

Untuk mengatasi keterbatasan dalam klasifikasi multi-kelas berskala besar, Kabiraj et al. (2022) mengusulkan arsitektur CX-Ultraneet, yakni model CNN yang dimodifikasi berdasarkan EfficientNet dengan pendekatan *compound scaling*. Model ini berhasil mengklasifikasikan 13 kategori penyakit toraks dengan akurasi rata-rata 88%, serta mampu menangani ketidakseimbangan kelas menggunakan fungsi *loss* berbobot dan augmentasi data. Dataset besar dari NIH dan Mendeley digunakan untuk memastikan stabilitas dan generalisasi model dalam berbagai skenario klasifikasi penyakit paru [10].

Selain akurasi, aspek interpretabilitas menjadi perhatian penting dalam pengembangan sistem diagnosis otomatis berbasis CNN. Menyadari hal

tersebut, Mridha et al. (2023) membangun model klasifikasi pneumonia yang tidak hanya mencapai akurasi tinggi sebesar 99%, tetapi juga dilengkapi dengan pendekatan *Explainable AI* (XAI) seperti Grad-CAM dan Grad-CAM++. Pendekatan ini memungkinkan visualisasi bagian citra yang berkontribusi terhadap prediksi model, sehingga meningkatkan transparansi dan kepercayaan pengguna klinis terhadap sistem [11].

### Arsitektur CNN dalam Klasifikasi Citra Radiograf Toraks

Pemilihan arsitektur *Convolutional Neural Network* (CNN) berperan krusial dalam menentukan efektivitas sistem klasifikasi citra medis, termasuk radiograf toraks (*chest X-ray*). Setiap arsitektur memiliki karakteristik dan keunggulan tersendiri dalam hal kedalaman jaringan, jumlah parameter, efisiensi komputasi, serta kemampuan generalisasi terhadap berbagai jenis data.

Salah satu arsitektur yang banyak digunakan adalah ResNet-50, yaitu model *deep CNN* yang terdiri atas 50 lapisan dengan fitur khas berupa *residual connections*. Pendekatan ini pertama kali diperkenalkan untuk mengatasi masalah *degradasi akurasi* pada jaringan yang sangat dalam, di mana penambahan lapisan justru menyebabkan penurunan performa model. Residual connection memungkinkan aliran informasi secara langsung melalui *shortcut connection* tanpa harus melewati transformasi non-linear, sehingga mempercepat proses pelatihan dan menjaga kestabilan jaringan. Cahyani et al. (2023) dan Kansal et al. (2024) menunjukkan bahwa penggunaan ResNet-50 dalam klasifikasi citra CXR mampu memberikan akurasi tinggi dengan generalisasi yang baik [12][13].

Di sisi lain, EfficientNet-B0 hadir sebagai solusi untuk meningkatkan akurasi sambil tetap menjaga efisiensi komputasi. Arsitektur ini diperkenalkan dengan konsep *compound scaling*, yaitu pendekatan untuk menskalakan tiga dimensi utama model secara simultan: kedalaman jaringan, lebar layer, dan resolusi input citra. Pendekatan ini memungkinkan peningkatan performa tanpa eksplorasi arsitektur secara manual, serta menghasilkan jumlah parameter yang relatif lebih sedikit dibandingkan CNN konvensional. Kansal et al. (2024) melaporkan bahwa EfficientNet-B0 menunjukkan performa kompetitif dalam klasifikasi CXR dengan konsumsi sumber daya yang rendah, sehingga cocok digunakan pada perangkat dengan keterbatasan komputasi [13].

Pengembangan lebih lanjut dari arsitektur ini direpresentasikan oleh EfficientNetV2-L, yang diterapkan oleh Ali et al. (2024) dalam klasifikasi pneumonia berbasis citra CXR. Model ini

mengintegrasikan pendekatan *progressive learning* dan *fused convolution* untuk mempercepat pelatihan serta meningkatkan akurasi pada dataset besar. Hasil eksperimen menunjukkan bahwa EfficientNetV2-L memiliki keunggulan signifikan dibandingkan pendahulunya dalam mengklasifikasikan penyakit paru secara lebih akurat dan efisien [14].

Selain arsitektur generik, penelitian terkini juga mengarah pada pengembangan arsitektur khusus untuk penyakit tertentu. Salah satunya adalah COPDNet, yang diperkenalkan oleh Agughasi et al. (2023) untuk deteksi penyakit paru obstruktif kronik (COPD). Model ini merupakan modifikasi dari ResNet-50, dengan penambahan *attention block* dan peningkatan kapabilitas ekstraksi fitur, khususnya dalam menangkap pola spasial yang relevan dengan karakteristik COPD. Inovasi ini terbukti meningkatkan sensitivitas model dalam mengidentifikasi penyakit yang sering kali memiliki gejala tumpang tindih dengan patologi paru lainnya [15].

Pendekatan inovatif dalam pengembangan arsitektur CNN juga diperkenalkan oleh Pramanik et al. (2022) melalui model TOPCONet, yang menggabungkan CNN ringan dengan metode ensemble berbasis *Technique for Order of Preference by Similarity to Ideal Solution* (TOPSIS). Model ini terdiri dari tiga klasifikator dasar yang dilatih pada varian input berbeda—gambar RGB, kanal merah, dan citra gabungan dengan deteksi tepi Robert—untuk menangkap fitur komplementer secara maksimal. Hasil klasifikasi dari ketiga model digabung menggunakan strategi *multi-criteria decision making*, yang terbukti lebih efektif dibanding metode ensemble tradisional seperti majority voting dan soft voting. Dengan jumlah parameter yang jauh lebih kecil dibandingkan model konvensional seperti DenseNet atau VGG-19, TOPCONet menunjukkan performa sangat baik (akurasi hingga 98,78%) pada beberapa dataset CXR publik, sekaligus efisien secara komputasi dan cocok untuk implementasi di lingkungan dengan sumber daya terbatas [16].

### Transfer Learning dalam Klasifikasi Citra Radiograf Toraks (CXR)

*Transfer learning* telah menjadi pendekatan yang semakin banyak diadopsi dalam klasifikasi citra medis, termasuk radiograf toraks (*chest X-ray*), karena kemampuannya untuk memanfaatkan pengetahuan dari model pralatih yang dilatih pada dataset berskala besar seperti ImageNet. Strategi ini terbukti efektif dalam meningkatkan akurasi dan efisiensi pelatihan, terutama ketika model diterapkan pada domain medis yang cenderung memiliki keterbatasan jumlah data berlabel.

Salah satu pendekatan inovatif dalam transfer learning ditunjukkan oleh Anwar et al. (2023), yang mengembangkan metode *self-supervised learning* untuk ekstraksi representasi awal dari citra CXR tanpa memerlukan label. Representasi ini kemudian digunakan sebagai fondasi dalam proses klasifikasi penyakit paru. Hasil eksperimen mereka menunjukkan bahwa pelatihan mandiri sebelum transfer learning mampu meningkatkan performa CNN secara signifikan, terutama dalam kondisi data berlabel yang terbatas [17].

Studi lain dilakukan oleh Mirzaee dan Ghorbanzadeh (2025), yang membandingkan performa tiga arsitektur CNN populer dalam konteks transfer learning, yakni ResNet50, MobileNetV2, dan VGG16. Penelitian ini mengungkap bahwa MobileNetV2 menunjukkan efisiensi tertinggi, dengan akurasi validasi mencapai 89,21%, serta keunggulan dalam kecepatan inferensi. Hal ini menjadikan MobileNetV2 ideal untuk implementasi real-time di perangkat dengan keterbatasan sumber daya. Sebaliknya, ResNet50 memiliki kelebihan dalam hal kedalaman ekstraksi fitur namun memerlukan tuning yang lebih kompleks untuk mencapai performa optimal. Sementara itu, VGG16 menunjukkan performa yang kompeten namun membutuhkan waktu pelatihan yang lebih lama karena arsitekturnya yang lebih sederhana namun parameter-intensif [18].

Penelitian lain oleh Hamal et al. (2022) memperkuat efektivitas transfer learning dengan membandingkan performa model machine learning tradisional seperti K-Nearest Neighbors (KNN) dan Support Vector Machine (SVM) dengan pendekatan CNN berbasis transfer learning menggunakan arsitektur **InceptionV3** dan **VGG16**. Hasil eksperimen menunjukkan bahwa model CNN yang diinisialisasi dengan bobot pralatih secara signifikan melampaui model konvensional dalam mendeteksi pneumonia dari citra CXR. Akurasi tertinggi diperoleh oleh model berbasis InceptionV3 dengan nilai **90,2%**, dibandingkan SVM yang hanya mencapai **76,1%**. Penelitian ini menegaskan bahwa transfer learning tidak hanya meningkatkan akurasi klasifikasi, tetapi juga mengurangi kebutuhan waktu pelatihan serta memperluas generalisasi model pada data medis yang kompleks dan bervariasi [19].

Lebih lanjut, penelitian oleh Mirugwe et al. (2025) melakukan perbandingan menyeluruh terhadap enam arsitektur CNN, yaitu VGG16, VGG19, ResNet50, ResNet101, ResNet152, dan Inception-ResNet-V2, dalam tugas klasifikasi tuberkulosis berbasis citra CXR. Meskipun memiliki jumlah parameter yang lebih sedikit dibanding model lain, VGG16 justru menghasilkan akurasi tertinggi sebesar 99,4% dengan F1-score mencapai 98,3%. Temuan ini menunjukkan



bahwa arsitektur yang relatif ringan tetap mampu memberikan hasil sangat baik apabila strategi transfer learning dan *fine-tuning* diterapkan secara tepat. Selain itu, model kompleks seperti ResNet101 dan Inception-ResNet-V2 juga menunjukkan performa yang kompetitif, khususnya saat pelatihan difokuskan pada lapisan akhir dengan melakukan pembekuan terhadap lapisan awal [20].

Subramaniam et al. (2023) menegaskan bahwa efektivitas pendekatan *transfer learning* dalam pengembangan sistem deteksi otomatis COVID-19 berbasis citra CXR. Studi ini mencermati berbagai arsitektur CNN pralatih seperti VGG16, ResNet101, MobileNet, dan Inception-V3 yang telah diadaptasi untuk klasifikasi dua kelas (binary) maupun multi-kelas, dengan fokus pada kasus COVID-19, pneumonia, dan kondisi normal. Salah satu temuan penting dari tinjauan ini adalah bahwa penggunaan model pralatih tidak hanya mempercepat proses pelatihan dan meningkatkan akurasi, tetapi juga memungkinkan model untuk belajar dari fitur representatif yang telah dikembangkan pada dataset berskala besar seperti ImageNet. Selain itu, penelitian ini menekankan pentingnya tahapan *preprocessing* seperti normalisasi, histogram equalization, dan augmentasi berbasis rotasi serta flipping untuk meningkatkan kestabilan model dan mengurangi risiko *overfitting*. [21].

### Klasifikasi Multi-Kelas dalam Citra Radiograf Toraks (CXR)

Klasifikasi multi-kelas pada citra radiograf toraks (*chest X-ray* / CXR) merupakan tantangan tersendiri dalam bidang pengolahan citra medis, mengingat banyaknya jenis penyakit paru yang memiliki kemiripan dalam pola visual dan derajat keparahan yang bervariasi. Keberhasilan dalam tugas ini sangat bergantung pada kemampuan model untuk mengekstraksi fitur yang diskriminatif, menangani ketidakseimbangan kelas, serta menjaga akurasi prediksi untuk setiap label yang berbeda.

Nahiduzzaman et al. (2023) mengusulkan *framework* inovatif berbasis *parallel CNN-Extreme Learning Machine* (CNN-ELM) yang mampu mengklasifikasikan 17 jenis penyakit paru, termasuk COVID-19, tuberkulosis, dan pneumonia. Model ini menggabungkan CNN ringan sebagai ekstraktor fitur dan ELM sebagai pengklasifikasi utama. Pendekatan ini mampu mencapai akurasi sebesar 90,92% dan F1-score rata-rata 0,91, bahkan ketika diuji pada dataset yang bersifat tidak seimbang. Pengujian dalam berbagai skenario multi-kelas menunjukkan bahwa performa model ini melampaui sejumlah arsitektur *state-of-the-art* lainnya dalam hal kecepatan dan akurasi klasifikasi [22].

Majumder et al. (2023) mengambil pendekatan berbeda dengan mengembangkan sistem klasifikasi *multi-label* dan *multi-class* menggunakan arsitektur DenseNet untuk mendeteksi hingga 15 jenis penyakit paru. Dengan memanfaatkan strategi *data augmentation* dan *transfer learning*, serta fine-tuning terhadap bobot model pretrained, mereka berhasil mencapai F1-score sebesar 0,89 dan *Area Under Curve* (AUC) sebesar 0,95 pada dua dataset besar, yakni CheXpert dan ChestX-ray14. Studi ini menyoroti efektivitas DenseNet dalam menangani kompleksitas visual yang tumpang tindih antar penyakit paru [23].

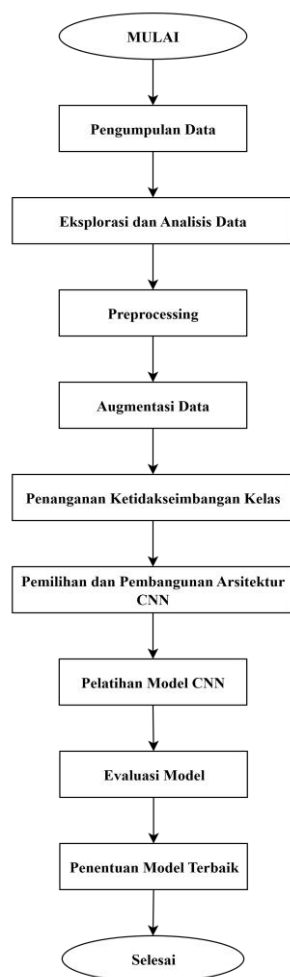
Pendekatan berbasis *ensemble learning* juga terbukti menjanjikan dalam meningkatkan akurasi klasifikasi multi-kelas. Ejaz et al. (2021) memperkenalkan model ensemble CNN yang dilatih untuk mengklasifikasikan empat kategori utama: normal, pneumonia bakteri, pneumonia virus, dan COVID-19. Mereka menggunakan model pretrained seperti ResNet50 dan InceptionV3 dalam pendekatan transfer learning, yang kemudian dikombinasikan dalam satu sistem klasifikasi. Hasil eksperimen menunjukkan akurasi tertinggi sebesar 96,78%, yang mencerminkan keunggulan pendekatan ensemble dalam mengatasi tumpang tindih antar kelas dan meningkatkan generalisasi model [24].

Analisis komparatif oleh Nallakaruppan et al. (2022) membandingkan performa berbagai arsitektur deep learning dalam klasifikasi citra CXR COVID-19, seperti ResNet, VGG16, InceptionV3, dan Xception, baik secara individual maupun dalam kombinasi. Studi ini menunjukkan bahwa InceptionV3 mencapai akurasi tertinggi sebesar 96,88% dan F1-score 0,9688, menjadikannya unggul dalam klasifikasi multi-kelas. Sementara itu, pendekatan gabungan seperti ResNet dengan VGG menunjukkan potensi tinggi untuk klasifikasi infeksi namun mengalami fluktuasi pada loss validasi. Temuan ini menekankan pentingnya pemilihan arsitektur dan stabilitas pelatihan dalam menangani kompleksitas klasifikasi multi-kelas citra CXR secara efektif [25].

### III. METODOLOGI PENELITIAN

Metodologi dalam penelitian ini mencakup tahapan sistematis yang dimulai dari pengumpulan data citra radiograf toraks (CXR) yang diunduh dari Kaggle, dilanjutkan dengan analisis distribusi data, *preprocessing* (resize, normalisasi, dan peningkatan kontras menggunakan CLAHE), serta augmentasi data untuk meningkatkan variasi dan generalisasi model. Data kemudian dibagi menjadi data latih dan uji yang digunakan untuk melatih lima arsitektur CNN pretrained, yaitu VGG16, EfficientNetB0,

DenseNet121, MobileNetV2, dan ResNet50. Masing-masing model dikonfigurasi dengan layer klasifikasi yang sama dan diuji pada empat variasi epoch (5, 10, 15, dan 30) menggunakan class weighting dan callback. Evaluasi dilakukan menggunakan metrik akurasi, precision, recall, dan f1-score, dilengkapi dengan confusion matrix dan visualisasi learning curve. Seluruh proses divisualisasikan dalam alur pada Gambar 1.



Gambar 1. Alur Penelitian

### Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari platform Kaggle dengan judul “Penyakit Paru-Paru 5 Kelas”, yang dapat diakses melalui tautan <https://www.kaggle.com/datasets/anjay007/penyakit-paru-paru-5-kelas>. Dataset ini berisi citra radiograf toraks (CXR) yang telah dikategorikan ke dalam lima kelas:

1. bacterial\_pneumonia
2. covid19
3. normal

4. tuberculosis
5. viral\_pneumonia

Dataset disusun dalam dua direktori utama, yaitu `train_dir` untuk data pelatihan dan `test_dir` untuk data pengujian, di mana masing-masing direktori berisi subfolder yang mewakili label kelas. Setelah diunduh, dataset diunggah ke Google Drive agar dapat diakses dan diolah langsung melalui *Google Colaboratory* selama proses pelatihan dan evaluasi model CNN.

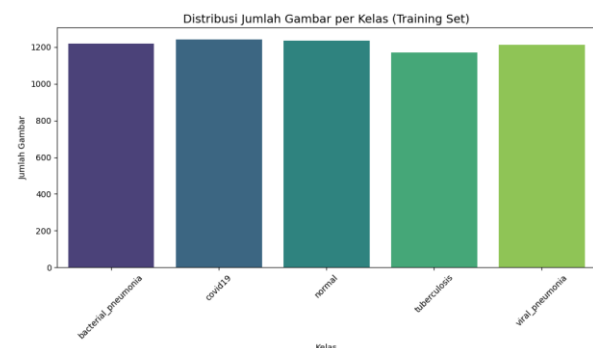
### Eksplorasi dan Analisis Data

Setelah memuat dataset, langkah pertama adalah mengeksplorasi struktur direktori dan menghitung jumlah citra pada setiap kelas di direktori pelatihan (`train_dir`). Tabel 1 menyajikan distribusi citra per kelas pada set pelatihan dan pengujian.

Tabel 1 Distribusi Kelas

Kelas	Train_dir	Test_dir
covid19	1243	248
Normal	1235	248
tuberculosis	1170	248
viral_pneumonia	1212	248
bacterial_pneumonia	1220	248
Total	6080	992

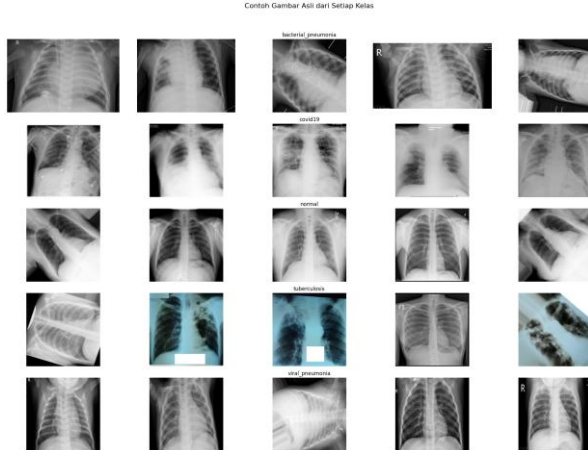
Eksplorasi ini bertujuan untuk memahami distribusi data serta memastikan proporsi sampel yang seimbang antar kelas. Berdasarkan hasil tersebut, terdapat lima kategori—*bacterial pneumonia*, *covid-19*, *normal*, *tuberculosis*, dan *viral pneumonia*—dengan jumlah citra yang relatif seimbang. Distribusi jumlah citra per kelas pada set pelatihan divisualisasikan pada Gambar 2.



Gambar 2. Distribusi Jumlah Gambar per Kelas pada Dataset Pelatihan

Selanjutnya, contoh citra asli dari masing-masing kelas disajikan untuk mengevaluasi karakteristik visual tiap kategori. Visualisasi ini diperlukan untuk memverifikasi kualitas data serta mengidentifikasi potensi variasi tekstur, kontras, atau pencahayaan yang

dapat memengaruhi proses pelatihan model. Contoh citra per kelas dapat dilihat pada Gambar 3.



Gambar 3. Contoh Citra per Kelas

### Preprocessing

Tahapan preprocessing dilakukan sebelum data digunakan dalam proses pelatihan model CNN. Tujuannya adalah untuk memastikan konsistensi ukuran, kualitas, dan format citra agar sesuai dengan kebutuhan arsitektur pretrained yang digunakan. Adapun langkah-langkah preprocessing yang diterapkan dalam penelitian ini.

#### A. Resize Citra

Pada tahap ini, seluruh citra pada dataset diubah ukurannya menjadi resolusi tetap 224×224 piksel. Proses ini dilakukan secara otomatis menggunakan parameter `target_size=(224, 224)` pada objek `ImageDataGenerator`. Tujuannya adalah untuk memastikan setiap citra memiliki dimensi yang konsisten dan sesuai dengan format input standar model CNN pretrained. Secara matematis, proses perubahan resolusi dapat dinyatakan dengan fungsi transformasi spasial sebagai berikut:

$$I_{\text{resized}}(x', y') = I\left(x \cdot \frac{W'}{W}, y \cdot \frac{H'}{H}\right)$$

Dimana:

$I(x, y)$  : intensitas piksel pada koordinat dari citra asli

$I_{\text{resized}}(x', y')$  : intensitas piksel pada koordinat dari citra hasil resize

$W, H$  : lebar dan tinggi citra asli

$W', H'$  : lebar dan tinggi citra setelah resize (dalam hal ini 224×224)

#### B. CLAHE (Contrast Limited Adaptive Histogram Equalization)

CLAHE (Contrast Limited Adaptive Histogram Equalization) adalah metode peningkatan kontras lokal yang bekerja dengan membagi gambar menjadi

beberapa blok kecil (tiles), lalu melakukan ekualisasi histogram pada masing-masing blok. Teknik ini digunakan untuk menonjolkan detail penting pada citra radiograf toraks, khususnya di area paru-paru, agar model CNN dapat mengenali pola-pola penyakit secara lebih efektif. Dalam penelitian ini, CLAHE diterapkan pada masing-masing kanal warna (R, G, B) secara independen, kemudian digabungkan kembali menjadi citra RGB.

CLAHE bekerja dengan prinsip dasar *Histogram Equalization* terbatas, menggunakan fungsi CDF (Cumulative Distribution Function) dari histogram lokal, sebagai berikut:

$$CDF(i) = \sum_{j=0}^i \frac{h(j)}{N}$$

Dimana:

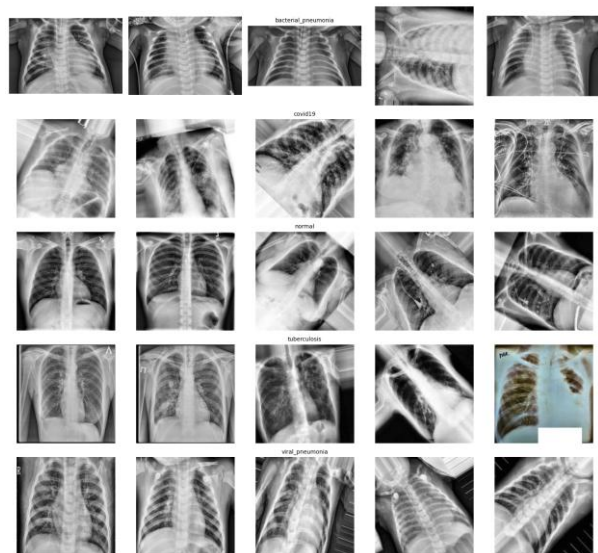
$h(j)$  : histogram nilai intensitas ke- $j$  dalam blok lokal

$N$  : jumlah total piksel dalam blok

$CDF(i)$  : probabilitas kumulatif untuk intensitas  $i$

Nilai CDF kemudian digunakan untuk menentukan nilai intensitas baru. CLAHE membatasi amplifikasi kontras melalui *clip limit*, sehingga mencegah noise diperbesar secara berlebihan. Hasil citra yang telah diterapkan CLAHE ditampilkan pada Gambar 3.

Contoh Gambar Hasil Preprocessing CLAHE (RGB)



Gambar 3. Contoh Gambar Hasil Preprocessing CLAHE (RGB)

#### C. Arsitektur CNN

Setiap arsitektur CNN pretrained memiliki fungsi preprocessing yang berbeda-beda. Fungsi ini diperlukan untuk menyesuaikan format input citra dengan distribusi data yang digunakan saat model

tersebut dilatih pertama kali pada dataset ImageNet. Oleh karena itu, penelitian ini menggunakan fungsi

Proses preprocessing yang spesifik untuk setiap arsitektur, yang secara otomatis melakukan normalisasi skala nilai piksel, pengurangan rata-rata kanal, atau transformasi lainnya.

1. VGG16 dan ResNet50 (menggunakan mean subtraction):

$$I_{\text{norm}}(x, y) = I(x, y) - \mu$$

2. MobileNetV2 dan EfficientNet (menggunakan skala [-1, 1]):

$$I_{\text{norm}}(x, y) = \frac{I(x, y)}{127.5} - 1$$

3. DenseNet121 (menggunakan skala [0, 1]):

$$I_{\text{norm}}(x, y) = \frac{I(x, y)}{255.0}$$

#### D. Normalisasi Nilai Piksel

Normalisasi nilai piksel merupakan proses penting dalam preprocessing citra sebelum dimasukkan ke dalam jaringan CNN. Tujuan dari normalisasi adalah untuk menyamakan skala nilai piksel agar sesuai dengan standar input model CNN pretrained, sehingga memudahkan proses pembelajaran dan mempercepat konvergensi selama pelatihan. Normalisasi ini tidak dilakukan secara manual, melainkan sudah termasuk dalam fungsi preprocess\_input yang digunakan pada masing-masing arsitektur CNN.

Terdapat beberapa metode normalisasi yang digunakan oleh arsitektur CNN pretrained, antara lain:

1. Skala ke [0, 1] (seperti pada DenseNet):

$$X' = \frac{x}{255}$$

2. Skala ke [-1, 1] (seperti pada MobileNetV2 dan EfficientNet):

$$X' = \frac{x}{127.5} - 1$$

3. Pengurangan rata-rata (mean subtraction) (seperti pada VGG16 dan ResNet):

$$X' = x - \mu$$

#### E. Generator Preprocessing untuk Testing

Pada tahap evaluasi, preprocessing data uji dilakukan dengan cara yang seragam dan konsisten, namun tanpa augmentasi. Hal ini penting untuk memastikan bahwa hasil evaluasi model mencerminkan kemampuan generalisasi terhadap data asli, bukan terhadap data yang telah dimodifikasi secara acak seperti pada tahap pelatihan. Oleh karena itu, digunakan objek *ImageDataGenerator* khusus untuk data testing yang hanya menerapkan fungsi preprocessing bawaan arsitektur, tanpa rotasi, pergeseran, atau flipping.

#### Augmentasi Data

Augmentasi data merupakan teknik yang digunakan untuk memperluas jumlah dan variasi data pelatihan secara artifisial tanpa menambah data baru secara manual. Tujuannya adalah untuk meningkatkan generalisasi model, mengurangi overfitting, dan membantu model mengenali objek pada berbagai kondisi. Dalam penelitian ini, augmentasi hanya diterapkan pada data pelatihan (*train\_dir*) dan tidak pada data pengujian.

Augmentasi dilakukan menggunakan kelas *ImageDataGenerator* dari pustaka Keras, yang memungkinkan transformasi citra secara real-time saat pelatihan berlangsung.

Jenis Transformasi Augmentasi yang Digunakan:

1. Rotasi Acak hingga  $\pm 20$  derajat:
2. Perpindahan Horizontal dan Vertikal (shift 10%):
3. Shear (kemiringan) hingga 10%:  
Transformasi bentuk paralelogram
4. Zoom hingga  $\pm 10\%$ :  
Pembesaran atau pengecilan area fokus citra
5. Flipping Horizontal:  
Pencerminan citra secara horizontal untuk mengubah arah tampilan
6. Fill Mode:  
Mengisi area kosong akibat transformasi dengan metode nearest (mengambil piksel terdekat)

Augmentasi diterapkan melalui parameter dalam *ImageDataGenerator*, artinya transformasi terjadi secara acak selama proses training, dan berbeda untuk setiap batch.

#### Penanganan Ketidakseimbangan Kelas

Dataset yang digunakan dalam penelitian ini memiliki distribusi jumlah citra yang tidak sepenuhnya seimbang di antara kelima kelas. Ketidakseimbangan kelas seperti ini berpotensi menyebabkan model terlalu fokus pada kelas mayoritas dan mengabaikan kelas minoritas, sehingga menghasilkan prediksi yang bias.

Untuk mengatasi hal tersebut, diterapkan metode *class weighting*, yaitu pemberian bobot pelatihan yang berbeda pada setiap kelas berdasarkan proporsi jumlah sampelnya. Kelas dengan jumlah sampel lebih sedikit akan diberikan bobot lebih besar agar kontribusinya terhadap fungsi loss setara dengan kelas mayoritas. Teknik ini diimplementasikan menggunakan fungsi *compute\_class\_weight* dari pustaka scikit-learn. Bobot kelas dihitung menggunakan rumus:

$$W_k = \frac{n}{k \cdot n_k}$$

Dimana:

$W_k$  : bobot untuk kelas ke- $k$

$n$  : total jumlah sampel

$k$  : jumlah total kelas

$n_k$  : jumlah sampel pada kelas ke- $k$



Bobot yang diperoleh kemudian diterapkan ke parameter `class_weight` saat proses pelatihan model. Hasil perhitungan untuk masing-masing kelas dalam dataset ditampilkan pada Tabel 2.

Tabel 2. Bobot Kelas Hasil Perhitungan  
*compute\_class\_weight*

Label Kelas	Bobot Kelas
0 (bacterial pneumonia)	0.9967
1 (covid19)	0.9783
2 (normal)	0.9846
3 (tuberculosis)	1.0393
4 (viral pneumonia)	1.0033

### Pemilihan dan Pembangunan Arsitektur CNN

Penelitian ini menggunakan pendekatan transfer learning dengan memanfaatkan lima arsitektur CNN pretrained yang telah terbukti efektif dalam berbagai tugas klasifikasi citra, khususnya pada domain medis. Kelima arsitektur yang digunakan adalah VGG16, ResNet50, EfficientNetB0, DenseNet121, dan MobileNetV2. Pemilihan arsitektur ini didasarkan pada popularitas, performa, efisiensi komputasi, dan ketersediaan bobot pralatih (*pretrained weights*) dari dataset ImageNet.

Setiap model dimodifikasi dengan menghapus lapisan atas (*top layer*) dan menggantinya dengan struktur klasifikasi baru agar dapat disesuaikan dengan jumlah kelas pada dataset, yaitu lima kelas penyakit paru-paru. Proses ini dilakukan dengan menambahkan *classification head* yang terdiri dari lapisan pooling, dense, dropout, dan output softmax.

Lapisan tambahan yang diterapkan pada setiap arsitektur meliputi:

1. GlobalAveragePooling2D()
2. Dense(512, activation='relu')
3. Dropout(0.5)
4. Dense(5, activation='softmax')

Struktur ini diterapkan secara konsisten pada kelima arsitektur CNN untuk menjaga keadilan dalam proses evaluasi dan perbandingan performa.

### Pelatihan Model CNN

Model CNN yang telah dibangun pada masing-masing arsitektur kemudian dilatih menggunakan data yang telah melalui proses preprocessing dan augmentasi. Untuk memastikan evaluasi performa yang menyeluruh, pelatihan dilakukan dalam empat eksperimen berbeda berdasarkan variasi jumlah epoch, yaitu 5, 10, 15, dan 30 epoch. Tujuannya adalah untuk mengamati pengaruh durasi pelatihan terhadap performa model.

Setiap proses pelatihan model dilengkapi dengan tiga jenis *callback*, yaitu:

1. ModelCheckpoint: menyimpan model terbaik berdasarkan nilai *validation accuracy* tertinggi.
2. EarlyStopping: menghentikan pelatihan lebih awal apabila model tidak menunjukkan peningkatan performa dalam beberapa epoch berturut-turut.
3. ReduceLROnPlateau: menurunkan nilai *learning rate* secara otomatis ketika performa validasi stagnan, guna menghindari local minima.

Proses pelatihan diulang sebanyak empat kali untuk masing-masing jumlah epoch yang telah ditentukan. Hasil dari setiap pelatihan, termasuk nilai akurasi dan loss validasi, disimpan untuk dianalisis lebih lanjut pada tahap evaluasi.

### Evaluasi Model

Setelah proses pelatihan selesai, performa setiap model CNN dievaluasi menggunakan data pengujian (test set) yang telah diproses tanpa augmentasi. Evaluasi dilakukan untuk keempat eksperimen jumlah epoch (5, 10, 15, dan 30) pada masing-masing arsitektur. Penilaian dilakukan secara menyeluruh menggunakan empat metrik utama, yaitu:

1. Akurasi (Accuracy):

Akurasi digunakan sebagai metrik utama untuk membandingkan performa model pada setiap variasi jumlah epoch. Rumus akurasi didefinisikan sebagai:

$$\text{Accuracy} = \frac{\text{Jumlah prediksi benar}}{\text{Total Jumlah data uji}}$$

2. Presisi (Precision):

Mengukur proporsi prediksi yang benar pada kelas tertentu terhadap semua prediksi yang dibuat untuk kelas tersebut.

$$\text{Precision} = \frac{TP}{TP + FP}$$

3. Recall (Sensitivitas):

Mengukur kemampuan model dalam menemukan seluruh sampel dari kelas tertentu.

$$\text{Recall} = \frac{TP}{TP + FN}$$

4. F1-Score: Merupakan rata-rata harmonik antara precision dan recall, yang digunakan untuk menilai trade-off antara keduanya:

$$F1 - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

*Confusion Matrix* juga digunakan untuk memberikan gambaran visual tentang jumlah prediksi benar dan salah untuk masing-masing kelas, serta untuk mengidentifikasi pola kesalahan yang konsisten.

Untuk memantau proses pelatihan, dilakukan visualisasi terhadap *validation accuracy* dan *validation loss* dari keempat eksperimen. Kurva ini membantu

mengevaluasi stabilitas model selama proses training dan mendeteksi overfitting atau underfitting.

### Penentuan Model Terbaik

Setelah seluruh model selesai dilatih dan dievaluasi, langkah terakhir adalah menentukan model terbaik dari masing-masing arsitektur berdasarkan performa klasifikasinya. Kriteria utama yang digunakan dalam pemilihan ini adalah nilai *validation accuracy* tertinggi, yaitu akurasi model pada data validasi selama pelatihan.

Dari empat eksperimen jumlah epoch (5, 10, 15, dan 30), nilai *validation accuracy* tertinggi pada masing-masing model dicatat dan dibandingkan. Model dengan akurasi validasi terbaik dianggap sebagai representasi performa optimal dari arsitektur tersebut.

Selain akurasi, metrik tambahan seperti f1-score per kelas, confusion matrix, dan stabilitas learning curve juga diperhatikan untuk mendukung keputusan akhir dan menghindari bias terhadap satu metrik saja.

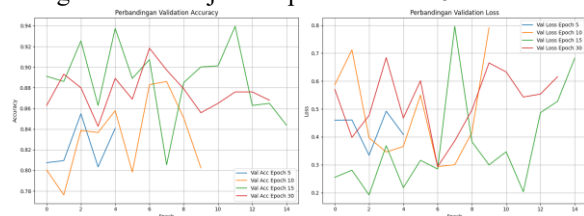
Hasil akhir dari tahap ini akan digunakan untuk membandingkan performa antar arsitektur dan menentukan arsitektur CNN yang paling optimal untuk tugas klasifikasi penyakit paru-paru pada citra radiograf toraks.

## IV. HASIL DAN PEMBAHASAN

### Hasil Evaluasi Tiap Arsitektur CNN

#### A. VGG16

Model VGG16 dievaluasi melalui empat skenario jumlah epoch (5, 10, 15, dan 30). Hasil pelatihan dan validasi menunjukkan bahwa kinerja model bervariasi pada setiap konfigurasi epoch. Untuk mengetahui tren pembelajaran model, dilakukan visualisasi learning curve berupa akurasi dan loss pada data validasi sebagaimana ditunjukkan pada Gambar 6.

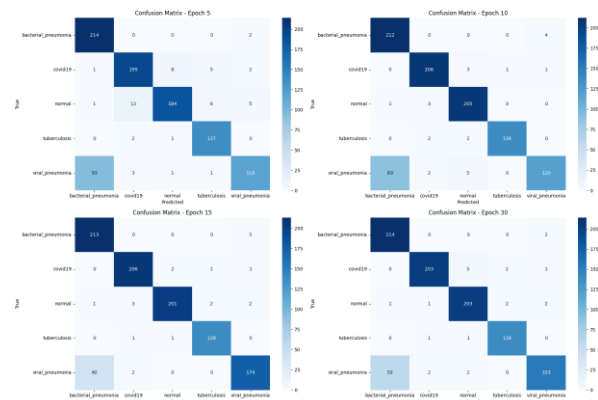


Gambar 6. Learning Curve Model VGG16 pada Data Validasi

Gambar ini menampilkan grafik perbandingan akurasi dan loss validasi pada setiap epoch. Terlihat bahwa pada epoch ke-15, model mencapai akurasi validasi tertinggi dan loss yang rendah secara stabil.

Evaluasi lebih lanjut dilakukan menggunakan confusion matrix untuk melihat distribusi prediksi

terhadap ground truth dari masing-masing kelas. Confusion matrix ini divisualisasikan pada Gambar 7.



Gambar 7. Confusion Matrix Model VGG16 pada Data Uji

Gambar ini menunjukkan performa klasifikasi model dalam memetakan setiap kelas penyakit paru-paru. Terlihat bahwa pada epoch ke-15, model berhasil meminimalkan kesalahan klasifikasi khususnya pada kelas *viral pneumonia*, yang sebelumnya banyak tertukar dengan *bacterial pneumonia*. Nilai metrik evaluasi dari tiap konfigurasi epoch untuk model VGG16 dapat dilihat pada Tabel 3.

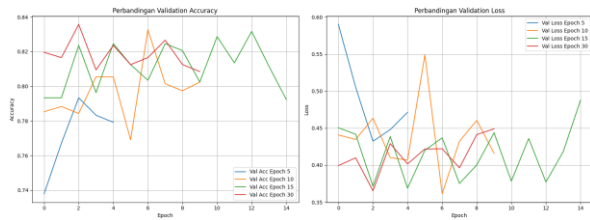
Tabel 3. Hasil Evaluasi Model VGG16 pada Data Uji

Epoch	Accuracy	Precision	Recall	F1-Score
5	0.8548	0.88	0.86	0.85
10	0.8861	0.92	0.89	0.89
15	0.9395	0.95	0.94	0.94
30	0.9183	0.93	0.92	0.92

Dari tabel tersebut terlihat bahwa model VGG16 mencapai performa terbaik pada epoch ke-15 dengan akurasi sebesar 93.95% dan F1-Score sebesar 0.94, menjadikannya sebagai konfigurasi optimal dari arsitektur ini.

#### B. EfficientNetB0

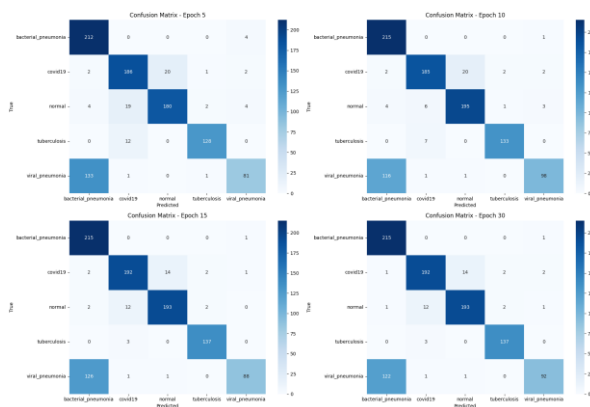
Model EfficientNetB0 dievaluasi melalui empat skenario jumlah epoch (5, 10, 15, dan 30). Hasil pelatihan dan validasi menunjukkan bahwa performa model meningkat seiring penambahan jumlah epoch, meskipun terdapat fluktuasi pada beberapa kelas tertentu. Untuk mengetahui tren pembelajaran model, dilakukan visualisasi learning curve berupa akurasi dan loss pada data validasi sebagaimana ditunjukkan pada Gambar 8.



Gambar 8. Learning Curve Model EfficientNetB0 pada Data Validasi

Gambar ini menampilkan grafik perbandingan akurasi dan loss validasi pada setiap epoch. Terlihat bahwa pada epoch ke-30, model memperoleh nilai akurasi validasi tertinggi secara konsisten dengan nilai loss yang cenderung stabil, menandakan pelatihan yang konvergen.

Evaluasi lebih lanjut dilakukan menggunakan confusion matrix untuk melihat distribusi prediksi terhadap ground truth dari masing-masing kelas. Confusion matrix ini divisualisasikan pada Gambar 9.



Gambar 9. Confusion Matrix Model EfficientNetB0 pada Data Uji

Gambar ini menunjukkan performa klasifikasi model dalam membedakan lima kelas penyakit paru-paru. Terlihat bahwa meskipun akurasi secara keseluruhan meningkat, pada beberapa epoch awal model masih mengalami kesalahan klasifikasi signifikan khususnya pada kelas viral pneumonia. Namun pada epoch ke-30, prediksi mulai membaik untuk seluruh kelas.

Nilai metrik evaluasi dari tiap konfigurasi epoch untuk model EfficientNetB0 dapat dilihat pada Tabel 4.

Tabel 4. Hasil Evaluasi Model EfficientNetB0 pada Data Uji

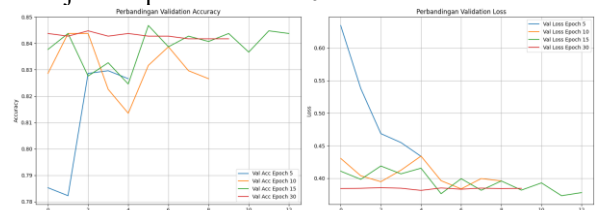
Epoch	Accuracy	Precision	Recall	F1-Score
5	0.790	0.83	0.79	0.78

10	0.832	0.87	0.83	0.82
15	0.830	0.88	0.83	0.82
30	0.836	0.88	0.84	0.82

Tabel 4 tersebut terlihat bahwa model EfficientNetB0 mencapai performa terbaik pada epoch ke-30 dengan akurasi sebesar 83.57% dan F1-Score sebesar 0.82, menjadikannya sebagai konfigurasi optimal dari arsitektur ini.

### C. DenseNet121

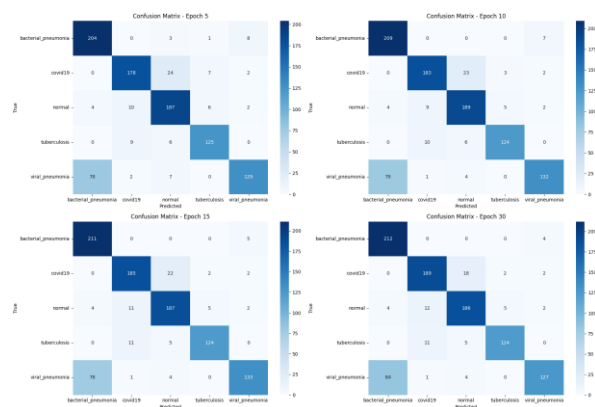
Model DenseNet121 dievaluasi dengan empat variasi epoch (5, 10, 15, dan 30) guna mengamati kestabilan performa model dalam proses pelatihan. Hasil training dan validasi kemudian divisualisasikan dalam bentuk grafik learning curve, seperti yang ditunjukkan pada Gambar 10.



Gambar 10. Learning Curve Model DenseNet121 pada Data Validasi

Gambar ini menunjukkan fluktuasi nilai akurasi dan loss pada data validasi untuk setiap konfigurasi epoch. Dapat dilihat bahwa pada epoch ke-15, model mencapai akurasi validasi tertinggi disertai penurunan nilai loss yang relatif stabil.

Evaluasi performa klasifikasi lebih lanjut dilakukan melalui confusion matrix untuk mengetahui detail prediksi terhadap masing-masing kelas. Hasil visualisasi confusion matrix ditampilkan pada Gambar 11.



Gambar 11. Confusion Matrix Model DenseNet121 pada Data Uji

Confusion matrix ini menunjukkan distribusi prediksi model terhadap lima kelas penyakit paru-paru. Kinerja terbaik dicapai pada epoch ke-15, di mana model berhasil mempertahankan akurasi tinggi serta mengurangi kesalahan klasifikasi, khususnya pada kelas viral pneumonia.

Nilai metrik evaluasi akurasi, precision, recall, dan F1-score dari keempat eksperimen epoch disajikan pada Tabel 5.

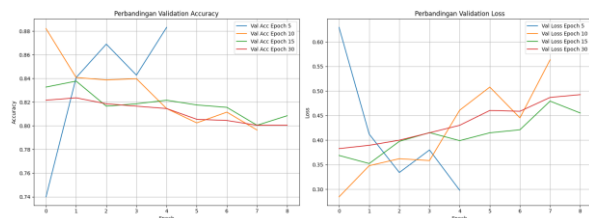
**Tabel 5. Hasil Evaluasi Model DenseNet121 pada Data Uji**

Epoch	Accuracy	Precision	Recall	F1-Score
5	0.8306	0.85	0.83	0.83
10	0.8438	0.87	0.85	0.85
15	0.8468	0.87	0.85	0.85
30	0.8448	0.87	0.85	0.84

Berdasarkan Tabel 5, dapat disimpulkan bahwa konfigurasi DenseNet121 dengan 15 epoch memberikan hasil terbaik dengan akurasi mencapai 84.68% dan nilai F1-Score sebesar 0.85, menjadikannya konfigurasi optimal pada arsitektur ini.

#### D. MobileNetV2

Model MobileNetV2 dievaluasi melalui empat skenario jumlah epoch (5, 10, 15, dan 30). Hasil pelatihan dan validasi menunjukkan bahwa performa terbaik diperoleh pada epoch ke-5. Untuk mengetahui tren pembelajaran model, dilakukan visualisasi learning curve berupa akurasi dan loss pada data validasi sebagaimana ditunjukkan pada Gambar 12.

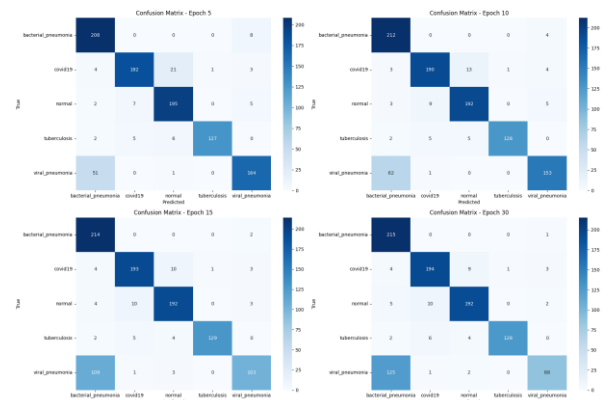


**Gambar 12. Learning Curve Model MobileNetV2 pada Data Validasi**

Gambar ini menampilkan perbandingan akurasi dan loss validasi untuk setiap konfigurasi epoch. Terlihat bahwa model dengan epoch ke-5 memiliki akurasi validasi tertinggi secara konsisten dan nilai loss yang rendah.

Evaluasi lanjutan dilakukan melalui visualisasi confusion matrix untuk setiap konfigurasi epoch. Confusion matrix ini memberikan gambaran distribusi

prediksi model terhadap ground truth dan ditampilkan pada Gambar 13.



**Gambar 13. Confusion Matrix Model MobileNetV2 pada Data Uji**

Gambar menunjukkan bahwa pada epoch ke-5 model memiliki akurasi klasifikasi terbaik, terutama dalam mengklasifikasikan kelas viral pneumonia dan bacterial pneumonia yang sebelumnya sering tertukar.

Nilai metrik evaluasi (akurasi, precision, recall, dan F1-score) untuk masing-masing epoch ditampilkan dalam Tabel 6.

**Tabel 6. Hasil Evaluasi Model MobileNetV2 pada Data Uji**

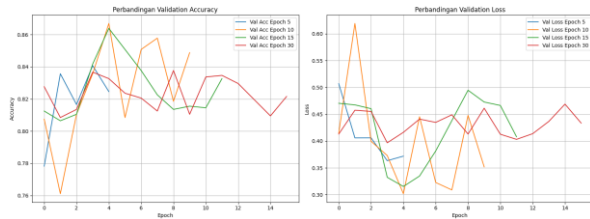
Epoch	Accuracy	Precision	Recall	F1-Score
5	0.8831	0.90	0.88	0.89
10	0.8821	0.90	0.88	0.89
15	0.8377	0.88	0.84	0.84
30	0.8236	0.88	0.82	0.82

Berdasarkan tabel tersebut, MobileNetV2 mencapai performa terbaik pada epoch ke-5 dengan akurasi sebesar 88.31% dan F1-Score 0.89, menjadikannya konfigurasi paling optimal dari arsitektur ini.

#### E. ResNet50

Model ResNet50 dievaluasi dengan empat skenario jumlah epoch yaitu 5, 10, 15, dan 30. Setiap model dilatih menggunakan strategi preprocessing, augmentasi data, dan class weighting yang seragam. Untuk melihat performa pembelajaran model pada data validasi, digunakan visualisasi learning curve berupa grafik akurasi dan loss seperti ditunjukkan pada Gambar 12.

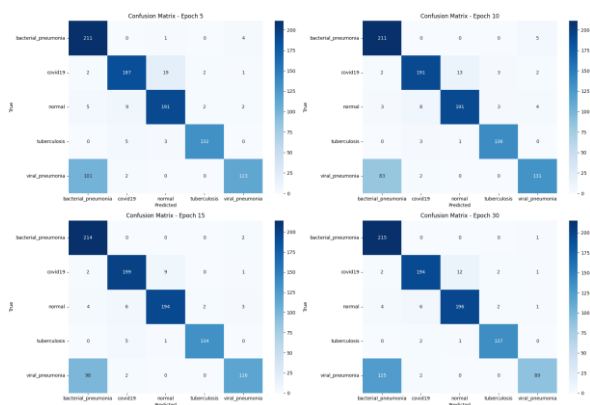




Gambar 12. *Learning Curve Model ResNet50 pada Data Validasi*

Gambar ini menunjukkan perkembangan akurasi dan loss validasi dari keempat konfigurasi epoch. Terlihat bahwa pada epoch ke-10, akurasi validasi mencapai nilai tertinggi sebesar 0.8669 dengan tren loss yang stabil dan relatif rendah dibandingkan konfigurasi lainnya.

Untuk mengetahui tingkat akurasi klasifikasi pada masing-masing kelas, digunakan confusion matrix sebagaimana ditampilkan pada Gambar 13.



Gambar 13. *Confusion Matrix Model ResNet50 pada Data Uji*

Gambar ini memperlihatkan performa klasifikasi model pada lima kelas penyakit paru-paru. Confusion matrix memperlihatkan bahwa model pada epoch ke-10 mampu mengklasifikasikan kelas normal, tuberculosis, dan covid19 dengan tingkat akurasi yang tinggi, serta mengurangi kesalahan klasifikasi pada kelas viral pneumonia.

Seluruh metrik evaluasi dari model ResNet50 untuk keempat konfigurasi epoch disajikan pada Tabel 6.

Tabel 6. Hasil Evaluasi Model ResNet50 pada Data Uji

Epoch	Accuracy	Precision	Recall	F1-Score
5	0.8407	0.88	0.85	0.84

10	0.8669	0.89	0.87	0.87
15	0.8639	0.90	0.87	0.87
30	0.8377	0.89	0.85	0.84

Dari hasil pada tabel tersebut, diketahui bahwa konfigurasi epoch ke-10 memberikan hasil terbaik dengan akurasi sebesar 86.69% dan F1-Score sebesar 0.87, sehingga model pada epoch ini ditetapkan sebagai versi terbaik untuk arsitektur ResNet50.

## KESIMPULAN DAN SARAN

### A. KESIMPULAN

Penelitian ini telah berhasil mengeksplorasi dan mengevaluasi performa lima arsitektur Convolutional Neural Network (CNN) pretrained—yaitu VGG16, ResNet50, EfficientNetB0, DenseNet121, dan MobileNetV2—dalam tugas klasifikasi multi-kelas citra radiograf toraks (CXR) penyakit paru-paru. Proses klasifikasi dilakukan terhadap lima kategori penyakit, yakni bacterial pneumonia, COVID-19, tuberculosis, viral pneumonia, dan kondisi normal.

Beberapa poin kesimpulan yang dapat diambil dari hasil penelitian ini adalah

1. Penggunaan metode preprocessing berbasis CLAHE pada kanal warna RGB secara signifikan meningkatkan kualitas kontras citra dan menonjolkan fitur jaringan paru-paru. Kombinasi ini, bersama dengan normalisasi input sesuai arsitektur CNN, berkontribusi besar dalam memperkuat proses pembelajaran model.
2. Augmentasi data terbukti mampu meningkatkan kemampuan generalisasi model terhadap variasi bentuk dan intensitas citra, sedangkan penerapan class weighting berhasil mengatasi ketidakseimbangan distribusi kelas pada dataset, sehingga menghindari bias klasifikasi terhadap kelas mayoritas.
3. Berdasarkan hasil evaluasi terhadap empat metrik utama (akurasi, precision, recall, dan F1-score), arsitektur VGG16 dengan 15 epoch menunjukkan performa klasifikasi terbaik dengan akurasi mencapai 93,95% dan F1-score sebesar 0,94. Model ini mampu secara konsisten meminimalkan kesalahan klasifikasi, khususnya pada kelas viral pneumonia yang kerap mengalami tumpang tindih visual dengan kelas lain.
4. Model MobileNetV2 menunjukkan kinerja kompetitif pada epoch ke-5 (akurasi 88,31%), menjadikannya alternatif efisien untuk sistem dengan keterbatasan komputasi. ResNet50 dan DenseNet121 juga menunjukkan hasil yang baik secara umum, meskipun tidak melampaui VGG16. Sementara itu, EfficientNetB0 menunjukkan

stabilitas pelatihan, namun dengan performa klasifikasi yang relatif lebih rendah (akurat hingga 83,57%).

5. Seluruh arsitektur yang digunakan memanfaatkan strategi transfer learning dari bobot pralatih ImageNet, yang terbukti mampu mempercepat proses pelatihan dan meningkatkan akurasi model meski jumlah data tidak terlalu besar. Hal ini menunjukkan bahwa pemanfaatan model pretrained sangat efektif dalam domain medis, khususnya pada citra radiografi toraks.

## B. SARAN

Berdasarkan hasil penelitian dan analisis yang telah dilakukan, beberapa saran dapat diajukan untuk pengembangan penelitian lanjutan dan penerapan sistem klasifikasi CXR secara lebih luas:

1. Untuk meningkatkan kepercayaan klinis, disarankan agar penelitian selanjutnya mengintegrasikan teknik Explainable AI seperti Grad-CAM atau LIME untuk memvisualisasikan area citra yang menjadi dasar keputusan model. Ini penting dalam konteks medis agar hasil klasifikasi dapat diverifikasi secara visual oleh tenaga kesehatan.
2. Penelitian ini terbatas pada satu dataset dari Kaggle. Untuk memperoleh generalisasi yang lebih kuat, disarankan untuk memperluas dataset dengan citra dari berbagai sumber rumah sakit atau institusi medis lainnya. Validasi eksternal menggunakan data dunia nyata akan memperkuat kredibilitas sistem yang dikembangkan.
3. Meskipun VGG16 menunjukkan hasil terbaik, eksplorasi terhadap arsitektur mutakhir lainnya seperti EfficientNetV2, Vision Transformer (ViT), atau strategi ensemble (kombinasi model) berpotensi meningkatkan performa klasifikasi lebih lanjut.
4. Untuk meningkatkan fokus model pada area yang relevan, penelitian mendatang dapat mempertimbangkan penerapan segmentasi paru sebelum klasifikasi, guna menghilangkan informasi latar belakang yang tidak diperlukan.
5. Hasil penelitian ini memiliki potensi untuk diimplementasikan dalam sistem decision support untuk klinik atau rumah sakit, baik dalam bentuk web-based diagnostic tools maupun integrasi ke dalam sistem PACS. Oleh karena itu, disarankan pengembangan sistem antarmuka yang ramah pengguna (user-friendly interface) bagi tenaga medis.

## REFERENSI

- [1] M. H. Al-Sheikh, O. Al Dandan, A. S. Al-Shamayleh, H. A. Jalab, and R. W. Ibrahim, "Multi-class deep learning architecture for classifying lung diseases from chest X-Ray and CT images," *Sci. Rep.*, vol. 13, no. 1, pp. 1–14, 2023.
- [2] G. M. M. Alshmrani, Q. Ni, R. Jiang, H. Pervaiz, and N. M. Elshennawy, "A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images," *Alexandria Eng. J.*, vol. 64, pp. 923–935, 2023.
- [3] T. El Leil, M. Ahsan, and J. Haider, "Detecting COVID-19 from Chest X-rays Using Convolutional Neural Network Ensembles," *Computers*, vol. 12, no. 5, pp. 1–22, 2023.
- [4] H. Çiğ, M. T. Güllüoğlu, M. B. Er, U. Kuran, and E. C. Kuran, "Enhanced Disease Detection Using Contrast Limited Adaptive Histogram Equalization and Multi-Objective Cuckoo Search in Deep Learning," *Trait. du Signal*, vol. 40, no. 3, pp. 915–925, 2023.
- [5] H. Bhatt and M. Shah, "A Convolutional Neural Network ensemble model for Pneumonia Detection using chest X-ray images," *Healthc. Anal.*, vol. 3, no. March, p. 100176, 2023.
- [6] A. Jain, A. Bhardwaj, K. Murali, and I. Surani, "A Comparative Study of CNN, ResNet, and Vision Transformers for Multi-Classification of Chest Diseases," pp. 1–8, 2024.
- [7] M. Usman, I. A. Nasir, R. Saeed, H. Nazir, and M. Asad, "A Deep Learning Approach for Multi-Label Chest X-ray Diagnosis Using DenseNet-121," *IET Conf. Proc.*, vol. 2024, no. 10, pp. 210–217, 2024.
- [8] S. Ashwini, J. R. Arunkumar, R. T. Prabhu, N. H. Singh, and N. P. Singh, "Diagnosis and multi-classification of lung diseases in CXR images using optimized deep convolutional neural network," *Soft Comput.*, vol. 28, no. 7–8, pp. 6219–6233, 2024.
- [9] M. Jain, A. Shah, P. Sharma, and M. Campisi, "CAD : Computer-Aided Detection of Pneumonia Using Convolutional Neural Networks ( CNN )," vol. 14, no. 5, pp. 87–112, 2025.
- [10] A. Kabiraj, T. Meena, P. B. Reddy, and S. Roy, *Detection and Classification of Lung Disease Using Deep Learning Architecture from X-ray Images*, vol. 13598 LNCS, no. December. Springer International Publishing, 2022.
- [11] K. Mridha, A. C. Barman, S. Biswas, S. Sarkar, S. Biswas, and M. A. Priyok, "Accuracy and Interpretability: Developing a Computer-Aided Diagnosis System for Pneumonia Detection in Chest X-Ray Images," *2nd IEEE Int. Conf. Distrib. Comput. Electr. Circuits Electron. ICDCECE 2023*, no. April, 2023.
- [12] D. E. Cahyani, A. D. Hariadi, F. F. Setyawan, L. Gumila, and S. Setumin, "COVID-19 classification using CNN-BiLSTM based on chest X-ray images," *Bull. Electr. Eng. Informatics*, vol. 12, no. 3, pp. 1773–1782, 2023.
- [13] K. Kansal, T. B. Chandra, and A. Singh, "ResNet-50 vs. EfficientNet-B0: Multi-Centric Classification of Various Lung Abnormalities Using Deep Learning 'session id: ICMLDsE.004,'" *Procedia Comput. Sci.*, vol. 235, pp. 70–80, 2024.
- [14] M. Ali *et al.*, "Pneumonia Detection Using Chest Radiographs with Novel EfficientNetV2L Model," *IEEE Access*, vol. 12, no. February, pp. 34691–34707, 2024.
- [15] A. V. Ikechukwu, S. Murali, and B. Honnaraju, "COPDNet: An Explainable ResNet50 Model for the Diagnosis of COPD from CXR Images," *2023 IEEE 4th Annu. Flagsh. India Counc. Int. Subsections Conf. Comput. Intell. Learn. Syst. INDISCON 2023*, no. October, pp. 1–7, 2023.
- [16] R. Pramanik, S. Dey, S. Malakar, S. Mirjalili, and R. Sarkar, "TOPSIS aided ensemble of CNN models for screening COVID-19 in chest X-ray images," *Sci. Rep.*, vol. 12, no. 1, pp. 1–19, 2022.

- 
- 
- [17] S. M. Anwar *et al.*, “SPCXR: Self-supervised Pretraining using Chest X-rays Towards a Domain Specific Foundation Model,” pp. 1–19, 2022.
  - [18] A. Mirzaee and P. Ghorbanzadeh, “Performance Evaluation and Comparison of Transfer Learning Models in Chest X-Ray Image Classification Using Deep Neural Networks,” vol. 7, no. 4, pp. 33–40, 2025.
  - [19] S. Hamal, B. K. Mishra, R. Baldock, W. Sayers, T. N. Adhikari, and R. M. Gibson, “A comparative analysis of machine learning algorithms for detecting COVID-19 using lung X-ray images,” *Decis. Anal. J.*, vol. 11, no. February, p. 100460, 2024.
  - [20] A. Mirugwe, L. Tamale, and J. Nyirenda, “Improving Tuberculosis Detection in Chest X-ray Images through Transfer Learning and Deep Learning: A Comparative Study of CNN Architectures,” vol. 6, pp. 1–15, 2024.
  - [21] K. Subramaniam *et al.*, “A comprehensive review of analyzing the chest X-ray images to detect COVID-19 infections using deep learning techniques,” *Soft Comput.*, vol. 27, no. 19, pp. 14219–14240, 2023.
  - [22] M. Nahiduzzaman *et al.*, “Parallel CNN-ELM: A multiclass classification of chest X-ray images to identify seventeen lung diseases including COVID-19,” *Expert Syst. Appl.*, vol. 229, no. PA, p. 120528, 2023.
  - [23] R. I. Majumder, “Efficient Classification of Pulmonary Pneumonia and Tuberculosis Alongside Normal and Non-X-ray Images with Minimal Resources Section 1 : Abstract & Introduction,” 2025.
  - [24] U. Ejaz, A. Luz, and R. Wasiu, “Comparison of CNN Architectures for COVID-19 Detection in Chest X-rays,” no. October 2023, 2025.
  - [25] M. K. Nallakaruppan, S. Ramalingam, S. R. K. Somayaji, and S. B. Prathiba, “Comparative Analysis of Deep Learning Models Used in Impact Analysis of Coronavirus Chest X-ray Imaging,” *Biomedicines*, vol. 10, no. 11, pp. 1–20, 2022.