



Classification of Pneumonia from X-Ray Images Using Deep Learning with the Ensemble Stacking Method

Klasifikasi Pneumonia dari Citra X-Ray Menggunakan Deep Learning dengan Metode Ensemble Stacking

**Bagas Renovan Situmorang¹, Surya Alfredo Simanjuntak^{2*},
Lasia Tomi Giantama Sembiring³, Asian Rambe⁴, Abdi Dharma⁵**

^{1,2,3,4,5}Program Studi Teknik Informatika, Fakultas Sains dan Teknologi,
Universitas Prima Indonesia, Indonesia

E-Mail: ¹bagasrenovan2211@gmail.com, ²suryauyak123@gmail.com, ³tomipandia@gmail.com,
⁴asianrambee@gmail.com, ⁵abdidharma@unprimdn.ac.id

Received Apr 24th 2026; Revised Jun 18th 2026; Accepted Jul 22th 2026; Available Online Aug 01st 2026

Corresponding Author: Surya Alfredo Simanjuntak

Copyright © 2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

Pneumonia is one of the leading causes of death worldwide, particularly among children and older adults. Early detection through chest X-ray images is essential to support accurate diagnosis and appropriate treatment. However, manual interpretation of chest X-rays requires specialized expertise and is prone to diagnostic errors. Therefore, this study aims to develop a Deep Learning-based pneumonia classification system using Convolutional Neural Networks (CNNs) combined with an ensemble learning approach. Three CNN models, namely DenseNet161, ConvNeXt-Tiny, and DenseNet201, were employed. The dataset, obtained from Kaggle, consists of 5,856 chest X-ray images categorized into two classes: Normal and Pneumonia. The research followed the CRISP-DM methodology, covering all stages from business understanding to model evaluation. Experimental results showed that DenseNet161 achieved the best performance among the individual models, with an accuracy of 65.85%. After combining the models into a stacking ensemble with Logistic Regression as the meta-learner, the accuracy improved to 84.86% and the recall to 90.72%. These findings demonstrate that the stacking ensemble approach significantly enhances classification performance compared with individual models and has the potential to assist medical professionals in pneumonia diagnosis. The primary contribution of this study is the development of a more reliable classification system with high recall, reducing the risk of false negatives and supporting more accurate and efficient clinical decision-making for early pneumonia detection.

Keywords: Convolutional Neural Network, Deep Learning, Ensemble Stacking, Pneumonia Classification, X-ray Images

Abstrak

Pneumonia merupakan salah satu penyebab utama kematian di dunia, terutama pada anak-anak dan lansia. Deteksi dini melalui citra rontgen dada sangat penting untuk membantu proses diagnosis dan menentukan pengobatan yang tepat. Namun, pembacaan gambar rontgen secara manual membutuhkan keahlian khusus dan berisiko menyebabkan kesalahan. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan sistem klasifikasi pneumonia berbasis Deep Learning menggunakan metode Convolutional Neural Network (CNN) yang dikombinasikan dengan teknik ensemble learning. Penelitian ini menggunakan tiga model CNN, yaitu DenseNet161, ConvNeXt-Tiny, dan DenseNet201. Dataset berasal dari Kaggle dengan total 5.856 citra rontgen yang dibagi menjadi dua kelas, yaitu Normal dan Pneumonia. Proses penelitian mengikuti tahapan CRISP-DM, mulai dari pemahaman masalah hingga evaluasi model. Hasil pengujian menunjukkan bahwa DenseNet161 sebagai model tunggal terbaik dengan akurasi 65,85%. Setelah model digabungkan menggunakan ensemble stacking dengan regresi logistik sebagai meta-learner, akurasi meningkat menjadi 84,86% dengan recall 90,72%. Hasil ini menunjukkan bahwa pendekatan stacking ensemble mampu meningkatkan kinerja klasifikasi secara signifikan dibandingkan dengan model tunggal serta berpotensi membantu tenaga medis dalam mendiagnosis pneumonia. Kontribusi utama penelitian ini adalah menghasilkan sistem klasifikasi yang lebih andal dengan recall tinggi, sehingga dapat meminimalkan risiko false negative dan mendukung pengambilan keputusan medis untuk deteksi dini pneumonia yang lebih akurat dan efisien.

Kata Kunci: Chest X-ray, Convolutional Neural Network, Deep Learning, Ensemble Stacking, Klasifikasi Pneumonia

1. PENDAHULUAN

Pneumonia merupakan salah satu penyakit infeksi paru yang paling banyak menyebabkan kematian di dunia. Penyakit ini menyerang jaringan paru-paru dan menyebabkan peradangan yang dapat mengganggu proses pernapasan. Menurut data global, pneumonia masih menjadi penyebab utama kematian pada anak-anak di bawah usia lima tahun serta kelompok lanjut usia, terutama di negara berkembang yang memiliki fasilitas medis terbatas [1][2]. Deteksi dini pneumonia sangat penting karena dapat membantu dokter dalam menentukan langkah pengobatan yang tepat dan mencegah komplikasi yang lebih parah [3].

Perkembangan teknologi kecerdasan buatan (*Artificial Intelligence/AI*), khususnya *Deep Learning*, memberikan peluang besar di dunia medis, terutama dalam bidang diagnosis berbasis citra medis. Salah satu metode pencitraan yang umum digunakan adalah *Chest X-Ray* (CXR), yang dapat membantu dokter mendeteksi penyakit paru seperti pneumonia, tuberkulosis, dan COVID-19 secara cepat dan akurat [4][5]. Namun, proses interpretasi gambar *X-ray* secara manual masih sangat bergantung pada pengalaman *radiolog* dan berpotensi menimbulkan kesalahan diagnosis karena kemiripan pola antar penyakit paru [6][7].

Untuk mengatasi permasalahan tersebut, pendekatan *Deep Learning* (DL) mulai banyak digunakan karena kemampuannya dalam mengenali pola-pola kompleks dari data citra medis tanpa memerlukan fitur buatan tangan. Model *Convolutional Neural Network* (CNN) terbukti efektif dalam mengidentifikasi penyakit paru berdasarkan citra *X-ray* dengan tingkat akurasi yang tinggi [8][9][10]. Beberapa arsitektur populer yang sering digunakan antara lain VGG16, ResNet50, ResNet50V2 dan InceptionV3 di mana masing-masing memiliki keunggulan tersendiri dalam proses ekstraksi fitur dan efisiensi komputasi [8][11][12][13].

Meskipun begitu, model *Convolutional Neural Network* (CNN) tunggal seperti VGG16 atau ResNet telah menunjukkan hasil yang baik. Performa model tunggal (*single model*) sangat bergantung pada variasi karakteristik dataset dan rentan mengalami ketidakstabilan akurasi saat diuji pada data baru. Oleh karena itu, dibutuhkan pendekatan baru yang mampu menggabungkan kelebihan dari beberapa arsitektur CNN. Salah satu pendekatan yang dapat diterapkan adalah *ensemble learning*, khususnya metode *stacking*, karena dapat meningkatkan performa algoritma *data mining*, khususnya jika dibandingkan dengan penggunaan *single algorithm* secara konvensional [14]. Tantangan utama dalam klasifikasi citra *X-ray* pneumonia adalah tingginya kemiripan pola visual antarjaringan paru yang meradang, sehingga model tunggal sering kali melewatkan deteksi halus (*false negative*).

Untuk mengatasi keterbatasan tersebut, kebaruan (*novelty*) dari penelitian ini terletak pada penggabungan strategis antara arsitektur CNN yang berfokus pada efisiensi fitur mendalam dan arsitektur modern berbasis tren *Vision Transformer*. DenseNet dipilih karena menerima masukan tidak hanya dari lapisan sebelumnya tetapi juga dari semua lapisan sebelumnya dalam jaringan [15], sehingga mampu mengurangi jumlah parameter tanpa mengorbankan akurasi. Sedangkan ConvNeXt-Tiny merupakan arsitektur modern hasil evolusi CNN yang mengadopsi konsep desain dari *Vision Transformer* namun tetap mempertahankan efisiensi, varian Tiny memiliki persyaratan yang rendah untuk kondisi perangkat keras selama pelatihan [16]. Kombinasi ini dipilih karena DenseNet memiliki keunggulan dalam memanfaatkan kembali fitur (*feature reuse*) melalui koneksi padat antar lapisan guna menangkap detail kecil pneumonia. Sementara itu, ConvNeXt-Tiny melengkapinya dengan kemampuan generalisasi yang kuat khas arsitektur modern namun tetap hemat komputasi. Dengan menggunakan Logistic Regression sebagai *meta-learner*, sistem tidak hanya meratakan hasil, melainkan mempelajari pola prediksi terbaik dari ketiga model tersebut secara adaptif untuk menekan angka salah diagnosis.

Berbagai penelitian terdahulu menunjukkan bahwa pendekatan deep learning memiliki performa yang cukup tinggi dalam klasifikasi pneumonia berbasis citra *X-ray*, namun masih menyisakan sejumlah keterbatasan. Penelitian [8] menunjukkan bahwa model CNN mampu mencapai akurasi 92%–96% dengan recall tinggi dalam mendeteksi pneumonia. Penelitian [10] melaporkan bahwa model deep learning berbasis *Chest X-Ray* dapat mencapai akurasi sekitar 93% serta mempercepat proses diagnosis. Penelitian [17] menemukan bahwa VGG16 memiliki performa lebih baik dibandingkan ResNet50 dengan akurasi masing-masing sekitar 96% dan 94%. Sementara itu, penelitian [18] menunjukkan bahwa preprocessing seperti normalisasi dan augmentasi mampu meningkatkan performa model CNN hingga di atas 90%. Meskipun demikian, sebagian besar penelitian tersebut masih berfokus pada penggunaan model tunggal (*single model*) dan belum secara optimal mengeksplorasi kombinasi beberapa arsitektur untuk meningkatkan performa klasifikasi secara lebih konsisten. Oleh karena itu, diperlukan pendekatan berbasis *ensemble learning* seperti *stacking* untuk mengatasi keterbatasan tersebut.

Permasalahan utama dalam penelitian ini adalah masih adanya keterbatasan akurasi dan konsistensi model deep learning dalam mendeteksi pneumonia dari citra *X-ray*, serta tingginya ketergantungan pada model tunggal yang performanya bervariasi tergantung pada dataset dan konfigurasi. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan sistem klasifikasi pneumonia berbasis Deep Learning dengan menggunakan metode *Convolutional Neural Network* (CNN) yang dikombinasikan dengan teknik *ensemble learning*. Penelitian ini diharapkan dapat membantu tenaga medis meningkatkan kecepatan dan akurasi diagnosis pneumonia melalui sistem berbasis kecerdasan buatan [8][11]. Selain itu, hasil penelitian ini dapat menjadi referensi akademik bagi peneliti dan mahasiswa dalam pengembangan model deteksi penyakit berbasis citra

medis. Di sisi lain, sistem yang dikembangkan berpotensi digunakan sebagai alat bantu diagnosis awal bagi masyarakat, khususnya di wilayah dengan keterbatasan tenaga radiolog dan fasilitas kesehatan.

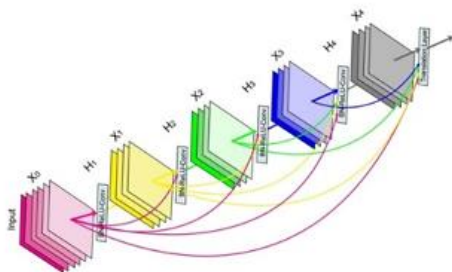
2. BAHAN DAN METODE

2.1. Pendekatan Penelitian

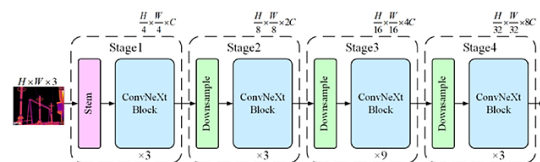
Penelitian ini menggunakan kerangka kerja *CRISP-DM* (*Cross Industry Standard Process for Data Mining*) sebagai pendekatan dalam proses pengembangan sistem deteksi klasifikasi pneumonia berbasis *Deep Learning* [19]. Pendekatan ini dipilih karena menyediakan tahapan yang sistematis dan fleksibel untuk pengolahan data, termasuk analisis citra medis. Tahapan *CRISP-DM* meliputi enam fase utama, yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment* [20]. Namun, dalam penelitian ini, tahap *Deployment* tidak dilakukan karena fokus penelitian hanya sampai pada tahap evaluasi model. Dengan menggunakan *CRISP-DM*, proses penelitian menjadi lebih terarah, mulai dari memahami permasalahan yang ingin diselesaikan, pengumpulan dan analisis data, hingga penerapan model CNN untuk mengklasifikasikan pneumonia dari citra *X-ray* dada. Pendekatan ini juga memudahkan proses validasi hasil, sehingga model yang dikembangkan memiliki performa yang baik dan dapat diandalkan.

2.2. Arsitektur Model CNN

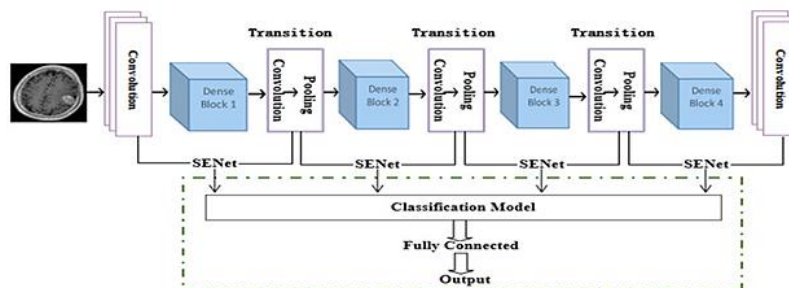
Penelitian ini menggunakan tiga model CNN, yaitu DenseNet161, ConvNeXt-Tiny, dan DenseNet201. DenseNet161 (lihat Gambar 1) merupakan varian dari arsitektur *Dense Convolutional Network* (DenseNet). Model ini memiliki konektivitas antar *layer* yang padat, di mana setiap *layer* menerima input dari semua *layer* sebelumnya. Pendekatan ini mengurangi masalah *vanishing gradient* dan meningkatkan efisiensi parameter. DenseNet161 memiliki kedalaman yang besar, namun tetap efisien dalam pemrosesan fitur. Berbeda dengan arsitektur model sebelumnya, ConvNeXt-Tiny merupakan arsitektur modern CNN yang dirancang untuk menyaingi performa *Vision Transformer* (ViT) yang menggabungkan kemudahan penggunaan CNN konvensional dengan ekstraksi fitur hierarkis dari model terbaru [22]. dengan tetap mempertahankan struktur CNN. Versi *Tiny* dari ConvNeXt memiliki ukuran yang lebih kecil namun tetap efisien dan akurat untuk dataset menengah hingga besar. Model ini menyesuaikan konfigurasi *kernel*, *stride*, dan *normalization layer* untuk meningkatkan generalisasi yang menggabungkan kemudahan penggunaan CNN konvensional dengan ekstraksi fitur hierarkis dari model terbaru. Sementara, DenseNet201 adalah arsitektur yang serupa dengan DenseNet161 namun memiliki jumlah *layer* lebih sedikit dan parameter lebih kecil. Berdasarkan penelitian [12] dan [13], DenseNet201 sering digunakan dalam deteksi pneumonia, COVID-19, dan TB, serta menghasilkan metrik akurasi dan *recall* yang tinggi.



Gambar 1. Model DenseNet161 [21]



Gambar 2. Model ConvNeXt-Tiny [23]

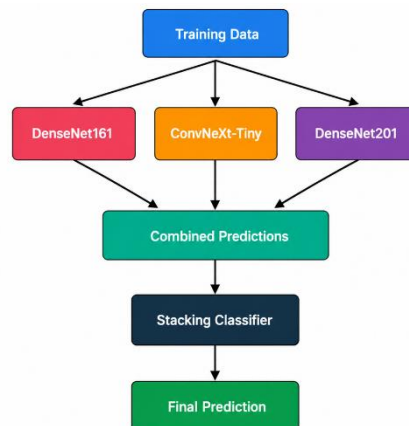


Gambar 3. Model DenseNet201 [24]

2.3. Ensemble Learning dan Stacking

Ensemble learning adalah salah satu paradigma *machine learning* dimana beberapa model (*base-models*) dilatih untuk memecahkan yang sama dan digabungkan untuk mendapatkan hasil yang lebih baik [17]. Salah satu metode *ensemble* yang populer adalah *stacking*, di mana beberapa model dasar (*base learners*) menghasilkan output yang kemudian menjadi input bagi model meta (*meta-learner*). Dalam penelitian ini, tiga

model CNN (DenseNet161, ConvNeXt-Tiny, dan DenseNet201) digunakan sebagai *base learners*, sedangkan *meta-learner* bertugas menggabungkan hasil prediksi untuk memperoleh keputusan akhir yang lebih akurat.



Gambar 1. Model Ensemble Stacking

2.4. Waktu dan Tempat Penelitian

Penelitian ini direncanakan berlangsung selama kurang lebih enam bulan, dimulai pada September 2025 hingga April 2026. Seluruh rangkaian kegiatan penelitian dilaksanakan di Kota Medan.

2.5. Alat Pendukung Proses Penelitian

Penelitian ini didukung oleh perangkat keras berupa prosesor Intel Core i7-12650H dengan kecepatan hingga 2,70 GHz yang dipadukan dengan GPU NVIDIA GeForce RTX 4050 dengan VRAM 6 GB GDDR6. Sistem juga dilengkapi dengan memori 16 GB DDR5. Pengembangan dilakukan pada sistem operasi Windows 11 64-bit dengan memanfaatkan Anaconda melalui Jupyter Notebook sebagai lingkungan kerja. Bahasa pemrograman yang digunakan adalah Python 3.10 dengan pustaka PyTorch 2.5.1 sebagai framework deep learning. Selain itu, CUDA 12.1 digunakan untuk mempercepat proses komputasi dan pelatihan model berbasis GPU.

2.6. Dataset Penelitian

Dataset yang digunakan dalam penelitian ini adalah Chest X-Ray Images (Pneumonia) yang diperoleh dari Kaggle, terdiri dari 5.856 citra X-ray dalam format JPEG yang terbagi ke dalam dua kelas, yaitu Normal dan Pneumonia. Seluruh citra melalui tahap preprocessing berupa perubahan ukuran (resize) menjadi 256×256 piksel dari resolusi asli rata-rata 1857×1317 piksel. Untuk mengatasi ketidakseimbangan data, dilakukan proses class balancing pada data latih sehingga jumlah citra pada masing-masing kelas menjadi seimbang, yaitu 1.340 citra Normal dan 1.340 citra Pneumonia. Dataset kemudian dibagi menjadi data train, test, dan validation dengan proporsi sebagaimana ditunjukkan pada Tabel 1. Data diambil pada link berikut: <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>, dengan data citra ditunjukkan pada Gambar 4 dan 5.

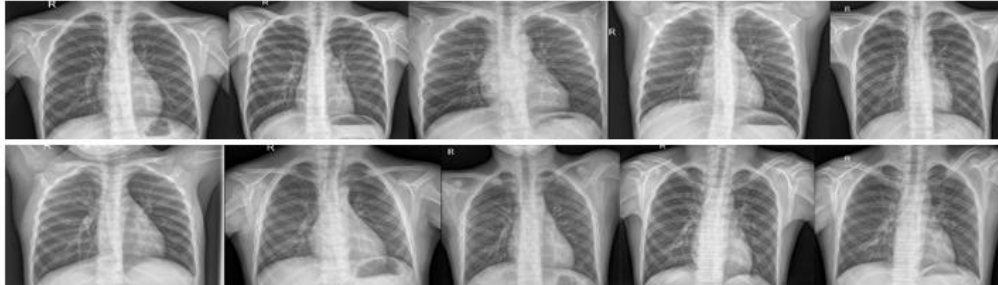
Tabel 1. Jumlah dan Pembagian Dataset

Dir	Dataset	Jumlah	Total	Persentase
Train	Normal	1.340	2.680	82,05%
	Pneumonia	1.340		
Test	Normal	234	568	17,39%
	Pneumonia	334		
Val	Normal	8	16	0,49%
	Pneumonia	8		
Jumlah Dataset			3.266	100%

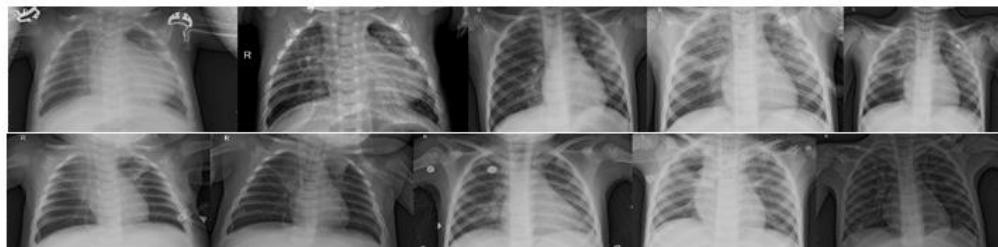
2.7. Tahapan CRISP-DM

Penelitian ini mengadopsi kerangka kerja CRISP-DM yang terdiri dari lima tahapan utama. Pada tahap business understanding, peneliti mengidentifikasi permasalahan utama, yaitu pentingnya deteksi dini pneumonia mengingat penyakit ini berpotensi fatal dan proses diagnosis manual dari citra X-ray masih memerlukan waktu serta bergantung pada keahlian radiolog. Tahap data understanding dilakukan dengan mengumpulkan dan menganalisis dataset Chest X-Ray yang diperoleh dari Kaggle, yang terdiri dari 5.863 citra dengan dua kelas, yaitu Normal dan Pneumonia. Selanjutnya, pada tahap data preparation, dilakukan

preprocessing citra untuk meningkatkan kualitas data dan memastikan model dapat mengenali pola secara optimal. Pada tahap modeling, digunakan tiga model Convolutional Neural Network (CNN), yaitu DenseNet161, ConvNeXt-Tiny, dan DenseNet201 sebagai base learner, yang kemudian dikombinasikan menggunakan metode stacking sebagai meta-learner untuk meningkatkan performa prediksi. Tahap terakhir, yaitu evaluasi, dilakukan dengan mengukur kinerja model menggunakan metrik Accuracy, Precision, Recall, dan F1-Score guna memperoleh evaluasi yang komprehensif terhadap hasil klasifikasi.



Gambar 4. Citra x-ray Normal



Gambar 5. Citra X-ray Pneumonia

2.8. Langkah Penelitian

Secara keseluruhan, proses pelatihan (*training*) untuk model tunggal (*single model*) dilakukan melalui 5 kali eksperimen independen guna memastikan stabilitas performa. Arsitektur DenseNet201 dikonfigurasi menggunakan bobot prapelatihan (*pretrained weights*) dari ImageNet, dengan memodifikasi bagian *classifier* menjadi beberapa lapisan linier (*Sequential*) yang dilengkapi dengan *Batch Normalization* dan *Dropout* sebesar 0.5 untuk mencegah *overfitting*. Proses pelatihan model ini menggunakan fungsi kerugian *Cross-Entropy Loss* dan algoritma optimasi Adam. Nilai tingkat pembelajaran (*learning rate*) yang digunakan disetel sebesar 0.001, dengan ukuran *batch* (*batch size*) sebesar 32, dan eksperimen berjalan selama 10 epoch pada setiap eksperimen. Selanjutnya, pada tahap *ensemble stacking*, probabilitas prediksi kelas dari model-model dasar (*base learners*) diekstrak sebagai fitur baru (*stacked features*). Fitur baru ini kemudian dilatih menggunakan *meta-learner* berbasis logistic regression. Pengoptimalan parameter pada Logistic Regression dilakukan secara otomatis menggunakan metode *GridSearchCV* dengan *cross-validation* (cv) sebanyak 5-fold untuk mencari kombinasi nilai regularisasi C yang paling optimal.

3. HASIL DAN DISKUSI

3.1. Hasil Penelitian

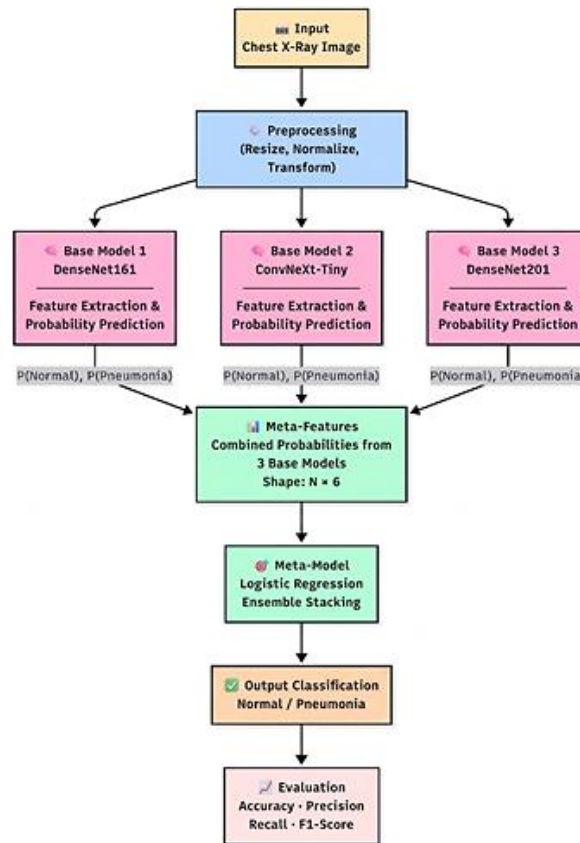
Hasil penelitian ini berupa sistem klasifikasi pneumonia berbasis citra X-ray dada yang diimplementasikan menggunakan metode Deep Learning Convolutional Neural Network (CNN) dan ensemble learning stacking. Sistem yang dikembangkan mampu mengklasifikasikan citra X-ray ke dalam dua kelas, yaitu Normal dan Pneumonia, melalui beberapa tahapan proses yang saling terintegrasi. Hasil pengujian model diperoleh melalui serangkaian eksperimen pada data uji menggunakan beberapa metrik evaluasi, yaitu accuracy, precision, recall, dan F1-score.

3.2. Pembahasan Tahapan Penelitian

Secara umum, alur proses sistem dimulai dari input citra X-ray, dilanjutkan dengan preprocessing data, ekstraksi fitur menggunakan tiga model CNN, penggabungan prediksi menggunakan metode stacking, hingga menghasilkan output klasifikasi akhir. Alur proses ini dirancang untuk meningkatkan akurasi dan keandalan sistem dalam mendeteksi pneumonia. Alur proses Sistem Klasifikasi Pneumonia ditunjukkan pada Gambar 6.

Sistem menerima citra X-ray dada sebagai input yang kemudian diproses melalui tahap preprocessing berupa penyesuaian ukuran, normalisasi, dan transformasi data agar sesuai dengan kebutuhan model. Citra hasil preprocessing selanjutnya diproses secara paralel oleh tiga base model CNN, yaitu DenseNet161,

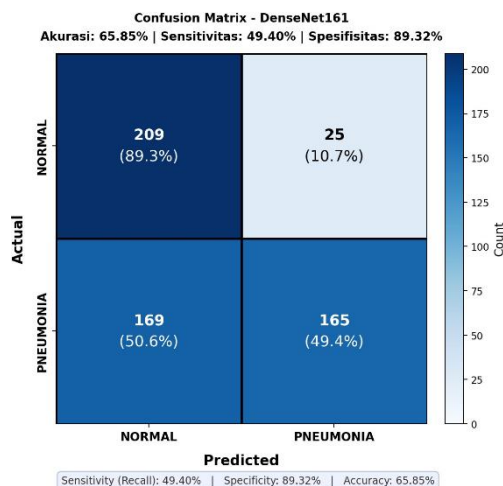
ConvNeXt-Tiny, dan DenseNet201 untuk melakukan ekstraksi fitur dan menghasilkan prediksi probabilitas kelas. Output dari masing-masing base model kemudian digabungkan dan disusun menjadi meta-feature yang merepresentasikan hasil prediksi secara kolektif. Meta-feature tersebut diproses menggunakan metode ensemble stacking dengan Logistic Regression sebagai meta-model untuk menentukan klasifikasi akhir. Hasil akhir sistem berupa klasifikasi citra X-ray ke dalam kelas Normal atau Pneumonia yang selanjutnya dievaluasi menggunakan metrik Accuracy, Precision, Recall, dan F1-Score.



Gambar 6. Menunjukkan alur proses Sistem Klasifikasi *Pneumonia*

3.3. Hasil Pelatihan Base Model DenseNet161

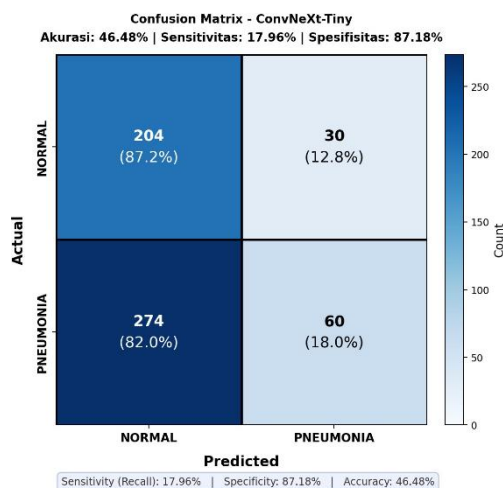
Berdasarkan hasil pengujian yang diperoleh dari proses pelatihan, model DenseNet161 menunjukkan performa terbaik dibandingkan base model lainnya. Model ini memperoleh akurasi sebesar 65,85% pada data uji. Hasil ini menunjukkan bahwa DenseNet161 mampu mengklasifikasikan citra X-ray dada dengan cukup baik antara kelas Normal dan Pneumonia. Kemampuan DenseNet161 dalam mendeteksi pneumonia tergolong tinggi, yang tercermin dari nilai sensitivitas yang baik dibandingkan model lain.



Gambar 7. Menunjukkan grafik akurasi model *DenseNet161* pada data uji.

3.4. Hasil Pelatihan Base Model ConvNeXt-Tiny

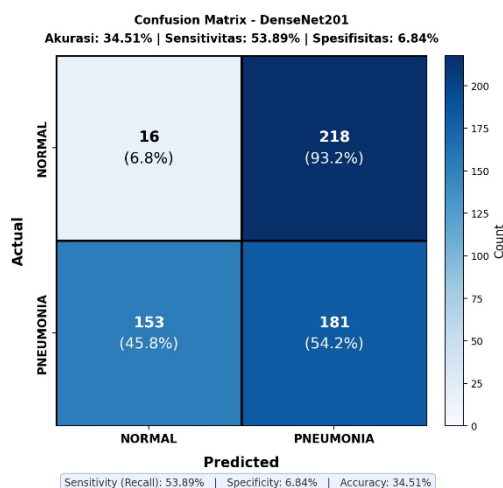
Model ConvNeXt-Tiny menghasilkan performa yang lebih rendah dibandingkan DenseNet161. Berdasarkan hasil pengujian, model ini memperoleh akurasi sebesar 46,48% pada data uji. Meskipun akurasi relatif rendah, ConvNeXt-Tiny menunjukkan tingkat spesifisitas yang tinggi dalam mengenali citra normal. Model ini memiliki keunggulan dari sisi efisiensi komputasi karena ukuran model yang lebih kecil dan waktu pelatihan yang lebih singkat. Variasi pola prediksi yang dihasilkan oleh ConvNeXt-Tiny tetap memberikan kontribusi penting dalam sistem ensemble, terutama dalam meningkatkan keberagaman prediksi antar model.



Gambar 8. Menunjukkan grafik akurasi model *ConvNeXt-Tiny* pada data uji.

3.5. Hasil Pelatihan Base Model DenseNet201

Model DenseNet201 menunjukkan performa yang lebih rendah dibandingkan dua base model lainnya. Berdasarkan hasil pengujian, DenseNet201 memperoleh akurasi sebesar 34,51% pada data uji. Hasil ini menunjukkan bahwa kemampuan model dalam membedakan kelas Normal dan Pneumonia masih terbatas. Meskipun demikian, DenseNet201 tetap digunakan sebagai base model karena mampu memberikan variasi prediksi yang berbeda dari DenseNet161 dan ConvNeXt-Tiny. Keberagaman prediksi ini berperan penting dalam metode ensemble stacking, karena dapat membantu meta-model dalam mengoreksi kesalahan prediksi dari model lain.

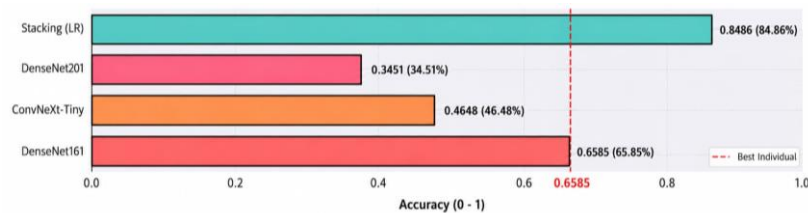


Gambar 9. Menunjukkan grafik akurasi model *DenseNet201* pada data uji.

3.6. Perbandingan Performa Model

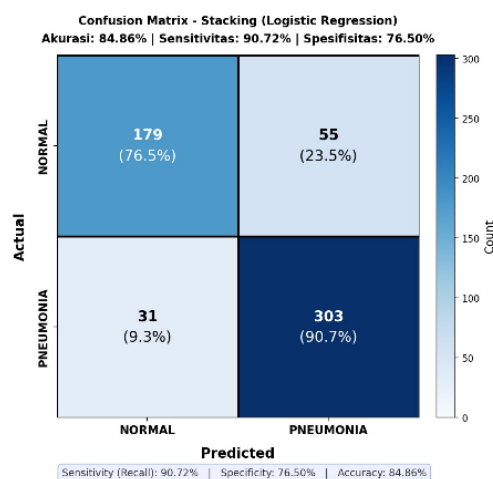
Metode Stacking Ensemble memberikan hasil terbaik di antara seluruh metode ensemble yang diuji. Pada penelitian ini, stacking menggunakan meta-model terbaik hasil GridSearchCV, yaitu Logistic Regression, yang dilatih menggunakan probabilitas keluaran dari DenseNet161, ConvNeXt-Tiny, dan DenseNet201. Berdasarkan hasil pengujian pada data uji, metode Stacking Ensemble menghasilkan akurasi tertinggi dan peningkatan performa yang signifikan dibandingkan base model terbaik. Secara teoritis, rendahnya akurasi pada beberapa model tunggal (seperti DenseNet201 dan ConvNeXt-Tiny) disebabkan oleh adanya keterbatasan

model mandiri dalam mengenali karakteristik visual yang samar pada beberapa variasi citra X-Ray pneumonia, atau adanya ketidakcocokan parsial saat pemuatan bobot praplatihan (*pretrained weights*).



Gambar 10. Perbandingan akurasi metode *Ensemble*.

Berdasarkan hasil pengujian, DenseNet161 menunjukkan performa terbaik di antara ketiga base model dalam mengklasifikasikan pneumonia dari citra X-ray. ConvNeXt-Tiny memiliki performa yang lebih rendah, namun unggul dalam efisiensi dan kecepatan pelatihan. DenseNet201 menghasilkan performa paling rendah, tetapi tetap memberikan variasi prediksi yang berguna dalam sistem ensemble. Metode *Ensemble Stacking* berhasil mengatasi kelemahan tersebut. *Meta-learner* (Logistic Regression) tidak sekadar meratakan prediksi, melainkan mempelajari pola kesalahan dan kelebihan dari masing-masing *base learner*. Ketika DenseNet161 unggul dalam menangkap fitur detail kecil namun lemah di area lain, *meta-learner* akan memberikan bobot keputusan yang lebih tinggi kepada DenseNet161 untuk karakteristik gambar tersebut. Proses saling melengkapi (*komplementer*) antararsitektur inilah yang membuat hasil akhir *Ensemble Stacking* melesat tinggi melampaui seluruh model tunggalnya.



Gambar 11. Menunjukkan grafik akurasi model Ensemble Stacking pada data uji

3.7. Perbandingan Fenomena Pneumonia dengan Penelitian Terdahulu

Fenomena pneumonia pada citra *Chest X-Ray* telah banyak dikaji dalam penelitian sebelumnya. Penelitian berjudul “*Comparison Analysis of CXR Images in Detecting Pneumonia Using CNN*” [8]. Penelitian ini membandingkan beberapa arsitektur CNN dalam mendeteksi pneumonia dan melaporkan akurasi antara 92% hingga 96%, namun nilai *recall* yang dihasilkan masih berada pada kisaran 88%–92%. Hal ini menunjukkan bahwa meskipun model cukup baik dalam klasifikasi secara umum, masih terdapat tantangan dalam mendeteksi seluruh kasus pneumonia secara konsisten.

Penelitian lain yang secara spesifik membahas fenomena pneumonia adalah jurnal berjudul “*Convolutional Neural Network-Based Analysis of Pediatric Chest X-Ray Images for Pneumonia Detection*” [10]. Penelitian ini berfokus pada pneumonia pada pasien anak (*pediatric pneumonia*) dan menunjukkan bahwa karakteristik pneumonia pada anak memiliki pola yang lebih kompleks. Model CNN yang digunakan pada penelitian ini menghasilkan akurasi sekitar 93%–95%, dengan nilai *recall* berkisar 91%.

93%. Meskipun hasilnya cukup baik, penelitian ini masih menggunakan satu model CNN sehingga performanya sangat bergantung pada arsitektur tunggal yang digunakan. Berbeda dengan kedua penelitian tersebut, penelitian ini menganalisis fenomena pneumonia dengan pendekatan *stacking ensemble*, yang mampu meningkatkan kemampuan deteksi pola pneumonia secara lebih komprehensif. Hasil penelitian menunjukkan nilai *recall* mencapai $\geq 95\%$, yang berarti sistem lebih efektif dalam mendeteksi kasus pneumonia dan mengurangi risiko kesalahan diagnosis dibandingkan dengan penelitian terdahulu.

3.8. Perbandingan Model Ensemble dengan Penelitian Terdahulu

Pendekatan *ensemble learning* telah banyak digunakan dalam berbagai bidang klasifikasi untuk meningkatkan akurasi dan stabilitas model. Penelitian berjudul “Perbandingan Metode Ensemble Learning pada Klasifikasi Penyakit Diabetes” [17]. Membahas penggunaan beberapa metode ensemble, seperti bagging dan boosting, untuk mengklasifikasikan penyakit diabetes. Penelitian tersebut menunjukkan bahwa metode ensemble mampu meningkatkan akurasi dibandingkan dengan model tunggal, dengan akurasi tertinggi mencapai $\pm 94\%$. Namun, pendekatan yang digunakan masih berfokus pada penggabungan model sejenis dan tidak melibatkan proses meta-learning untuk mengoptimalkan keputusan akhir.

Penelitian lainnya yang membahas metode ensemble adalah jurnal “Teknik Ensemble dalam Machine Learning untuk Menentukan Tingkat Akurasi Perkembangan Motorik Bayi 0–12 Bulan” [18]. Penelitian ini menerapkan teknik ensemble dengan mengombinasikan beberapa algoritma klasifikasi untuk meningkatkan akurasi prediksi perkembangan motorik bayi. Hasil penelitian menunjukkan peningkatan akurasi hingga $\pm 90\%$, namun, penggabungan model dilakukan melalui rata-rata hasil prediksi (simple ensemble) tanpa adanya proses pembelajaran lanjutan pada tingkat metamodel. Dengan demikian, keputusan akhir masih bersifat statis dan belum optimal dalam menangkap pola kompleks pada data.

Berbeda dengan kedua penelitian tersebut, penelitian ini menerapkan *stacking ensemble* yang tidak hanya mengombinasikan beberapa *base learner* dengan karakteristik algoritma yang berbeda, tetapi juga menggunakan *meta learner* untuk mempelajari pola dari hasil prediksi base model. Pendekatan ini memungkinkan sistem memperbaiki kesalahan prediksi antar-model secara adaptif. Hasil pengujian menunjukkan bahwa model stacking ensemble yang diusulkan mampu mencapai akurasi hingga 84,86%.

4. KESIMPULAN

Berdasarkan hasil eksperimen dan analisis yang telah dilakukan, dapat disimpulkan bahwa penerapan metode *Ensemble Stacking* dengan menggabungkan arsitektur *Deep Learning* DenseNet161, ConvNeXt-Tiny, dan DenseNet201 berhasil meningkatkan performa klasifikasi citra X-Ray pneumonia secara signifikan. Model *Ensemble Stacking* menggunakan *Meta-Learner* berbasis *Logistic Regression* sukses meraih akurasi tertinggi sebesar 84,86% dengan tingkat sensitivitas (*recall*) mencapai 90,72%, jauh melampaui performa model tunggal terbaiknya, yaitu DenseNet161 yang hanya mencatatkan akurasi 65,85%.

Penelitian ini memberikan kontribusi ilmiah berupa pembuktian bahwa integrasi antara karakteristik arsitektur CNN konvensional yang diperdalam (keluarga DenseNet) dan arsitektur modern (ConvNeXt-Tiny) melalui metode *stacking* terbukti efektif dalam meminimalkan bias prediksi. Secara praktis, tingginya nilai *recall* (90,72%) pada model gabungan ini sangat krusial bagi dunia medis karena mampu menekan angka salah diagnosis (*false negative*), sehingga sistem ini sangat layak diimplementasikan sebagai alat bantu keputusan (*decision support system*) yang andal bagi tenaga medis dalam deteksi dini pneumonia.

Untuk pengembangan penelitian di masa mendatang, disarankan melakukan perbaikan pada optimalisasi pemuatan bobot prapelatihan (*pretrained weights*) secara menyeluruh pada setiap *base learner* agar fitur yang diekstrak lebih maksimal. Selain itu, eksplorasi penggunaan arsitektur berbasis *Vision Transformer* (ViT) sebagai salah satu model dasar, atau penerapan teknik optimasi hyperparameter otomatis seperti Bayesian Optimization, sangat direkomendasikan untuk meningkatkan akurasi model *ensemble* ke tingkat yang lebih tinggi.

REFERENSI

- [1] Y. Hadhoud *et al.*, “From Binary to Multi-Class Classification: A Two-Step Hybrid CNN-ViT Model for Chest Disease Classification Based on X-Ray Images,” *Diagnostics*, vol. 14, no. 23, pp. 1–16, 2024, doi: 10.3390/diagnostics14232754.
- [2] S. Shetty, N. Mamatha, M. Shetty, S. Keerthana, and D. Pushparaj Shetty, “Detection of Pneumonia from Chest X-ray Images Utilizing MobileNet Model,” *Int. Conf. Distrib. Comput. Optim. Tech. ICDCOT 2024*, 2024, doi: 10.1109/ICDCOT61034.2024.10515758.
- [3] H. N. Ifayatin and R. A. Saputra, “Sistem Deteksi Penyakit Pneumonia Menggunakan Algoritma Faster R-CNN Berbasis Citra Digital Rontgen Dada Detection System of Pneumonia Disease Using Faster R-CNN Algorithm Based on Digital Chest X-ray Images,” *JUSTIN (Jurnal Sist. dan Teknol. Informasi)*, vol. 12, no. 4, pp. 645–652, 2024, doi: 10.26418/justin.v12i4.81304.
- [4] Y. F. Cheah, “Chest X-Ray Classification of Lung Diseases Using Deep Learning,” *Green Intell. Syst. Appl.*, vol. 1, no. 1, pp. 12–18, 2021, doi: 10.53623/gisa.v1i1.32.
- [5] M. S. Ahmed *et al.*, “Joint Diagnosis of Pneumonia, COVID-19, and Tuberculosis from Chest X-ray Images: A Deep Learning Approach,” *Diagnostics*, vol. 13, no. 15, 2023, doi: 10.3390/diagnostics13152562.
- [6] R. Mogaveera, R. Maur, Z. Qureshi, and Y. Mane, “Multi-class Chest X-ray classification of Pneumonia, Tuberculosis and Normal X-ray images using ConvNets,” *ITM Web Conf.*, vol. 44, p. 03007, 2022, doi: 10.1051/itmconf/20224403007.
- [7] D. Zhang, F. Ren, Y. Li, L. Na, and Y. Ma, “Pneumonia detection from chest x-ray images based on

- convolutional neural network,” *Electron.*, vol. 10, no. 13, 2021, doi: 10.3390/electronics10131512.
- [8] N. Izdiyar, S. B. Rahayu, and K. Venkatesan, “Comparison Analysis of CXR Images in Detecting Pneumonia Using VGG16 and ResNet50 Convolution Neural Network Model,” *Int. J. Informatics Vis.*, vol. 8, no. 1, pp. 326–332, 2024, doi: 10.62527/joiv.8.1.2258.
- [9] M. Mujahid, F. Rustam, R. Álvarez, J. Luis Vidal Mazón, I. de la T. Díez, and I. Ashraf, “Pneumonia Classification from X-ray Images with Inception-V3 and Convolutional Neural Network,” *Diagnostics*, vol. 12, no. 5, pp. 1–16, 2022, doi: 10.3390/diagnostics12051280.
- [10] Y. Ma and R. A. Gonzales, “Convolutional neural network-based analysis of pediatric chest X-ray images for pneumonia detection,” *J. Emerg. Investig.*, vol. 7, no. October, pp. 1–6, 2024, doi: 10.59720/24-031.
- [11] M. Kolhar, A. M. Al Rajeh, and R. N. A. Kazi, “Augmenting Radiological Diagnostics with AI for Tuberculosis and COVID-19 Disease Detection: Deep Learning Detection of Chest Radiographs,” *Diagnostics*, vol. 14, no. 13, 2024, doi: 10.3390/diagnostics14131334.
- [12] L. Mangeri, O. S. Gnana Prakasi, P. Kanmani, and N. Puppala, “Chest Diseases Prediction from X-ray Images using CNN Models: A Study,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 10, pp. 236–243, 2021, doi: 10.14569/IJACSA.2021.0121026.
- [13] T. Rahman *et al.*, “Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization,” *IEEE Access*, vol. 8, pp. 191586–191601, 2020, doi: 10.1109/ACCESS.2020.3031384.
- [14] F. A. Rafrastara, C. Supriyanto, C. Paramita, and Y. P. Astuti, “Deteksi Malware menggunakan Metode Stacking berbasis Ensemble,” vol. 8, no. 1, pp. 11–16, 2023.
- [15] E. Dio, B. Sudewo, M. K. Biddini, and A. Fadlil, “DenseNet Architecture for Efficient and Accurate Recognition of Javanese Script Hanacaraka Character,” vol. 23, no. 2, pp. 453–464, 2024, doi: 10.30812/matrik.v23i2.xxx.
- [16] Z. Li, T. Gu, B. Li, W. Xu, X. He, and X. Hui, “applied sciences ConvNeXt-Based Fine-Grained Image Classification and Bilinear Attention Mechanism Model,” 2022.
- [17] L. M. Cendani and A. Wibowo, “Perbandingan Metode Ensemble Learning pada Klasifikasi Penyakit Diabetes,” vol. 13, no. 1, pp. 33–44, 2022.
- [18] A. S. Sinaga, Sethu Ramen, and Sri Mulyani, “Teknik Ensemble Dalam Machine Learning Untuk Menentukan Tingkat Akurasi Perkembangan Motorik Bayi 0-12 Bulan,” *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 23, no. 2, pp. 418–425, 2024, doi: 10.53513/jis.v23i2.10059.
- [19] A. M. Shimaoka, R. C. Ferreira, and A. Goldman, “The evolution of CRISP-DM for Data Science: Methods, Processes and Frameworks,” *SBC Rev. Comput. Sci.*, vol. 4, no. 1, pp. 28–43, 2024, doi: 10.5753/reviews.2024.3757.
- [20] Y. A. Singgalen, “Penerapan Metode CRISP-DM dalam Klasifikasi Data Ulasan Pengunjung Destinasi Danau Toba Menggunakan Algoritma Naïve Bayes Classifier (NBC) dan Decision Tree (DT),” *J. Media Inform. Budidarma*, vol. 7, no. 3, p. 1551, 2023, doi: 10.30865/mib.v7i3.6461.
- [21] M. A. Saleh, M. Abdouh, and M. K. Ramadan, “A Novel Method for Improving Baggage Classification Using a Hyper Model of Fusion of DenseNet-161 and EfficientNet-B5,” 2024.
- [22] O. P. Mmileng, “Application of ConvNeXt with transfer learning and data augmentation for malaria parasite detection in resource-limited settings using microscopic images,” pp. 1–39, 2025, doi: 10.1371/journal.pone.0313734.
- [23] Y. Zhang, A. Xu, D. Lan, X. Zhang, J. Yin, and H. H. Goh, “ScienceDirect ConvNeXt-based anchor-free object detection model for infrared image of power equipment,” *Energy Reports*, vol. 9, pp. 1121–1132, 2023, doi: 10.1016/j.egyr.2023.04.145.
- [24] R. F. Rahmat and Y. T. A. Harahap, “Image Classification of Brain tumor based on Channel Attention Mechanism Image Classification of Brain tumor based on Channel Attention Mechanism,” 2021, doi: 10.1088/1742-6596/2035/1/012029.