



**Deteksi Abnormalitas Toraks yang Efisien Menggunakan
Model Deep Learning Mobile: Studi Komparatif MobileNet,
ShuffleNet, dan EfficientNet dengan Grad-CAM**

Achmad Bauravindah

23917016

Tesis diajukan sebagai syarat untuk meraih gelar Magister Komputer

Konsentrasi Sains Data

Program Studi Informatika Program Magister

Fakultas Teknologi Industri

Universitas Islam Indonesia

2025

Lembar Pengesahan Pembimbing

Deteksi Abnormalitas Toraks yang Efisien Menggunakan Model Deep Learning Mobile: Studi Komparatif MobileNet, ShuffleNet, dan EfficientNet dengan Grad- CAM




Ir. Dhomas Hatta Fudholi S.T, M.Eng, Ph.D.

Lembar Pengesahan Penguji

Deteksi Abnormalitas Toraks yang Efisien Menggunakan Model Deep Learning Mobile: Studi Komparatif MobileNet, ShuffleNet, dan EfficientNet dengan Grad- CAM

Achmad Bauravindah

23917016

Yogyakarta, Juni 2025

Tim Penguji,

Ir. Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D.

Ketua

Dr. Syarif Hidayat, S.Kom., M.IT.

Anggota I

Ir. Chandra Kusuma Dewa, S.Kom., M.Cs., Ph.D.

Anggota II

Mengetahui,

Ketua Program Studi Informatika Program Magister

Universitas Islam Indonesia



Ir. Irving Vitra Paputungan, S.T., M.Sc., Ph.D.

Abstrak

Deteksi Abnormalitas Toraks yang Efisien Menggunakan Model Deep Learning Mobile: Studi Komparatif MobileNet, ShuffleNet, dan EfficientNet dengan Grad-CAM

Tingginya beban penyakit paru di Indonesia, seperti tuberkulosis, pneumonia, dan gagal jantung, diperparah oleh keterbatasan jumlah dokter spesialis radiologi. Hal ini menuntut solusi teknologi yang efisien dan dapat dijalankan di lingkungan dengan sumber daya terbatas. Penelitian ini mengusulkan penggunaan tiga model deep learning ringan—MobileNetV2, ShuffleNetV2, dan EfficientNetB0—untuk mendeteksi lima kelainan toraks utama pada citra X-ray dada (CXR), yaitu pneumotoraks, kardiomegali, nodul/massa, konsolidasi, dan infiltrasi. Dataset ChestX-ray8 dari NIH digunakan sebagai basis data, dengan total 12.500 citra yang telah dinormalisasi dan diseimbangkan antar kelas. Pelatihan dilakukan menggunakan validasi silang 5-fold dan konfigurasi hiperparameter yang distandarisasi, termasuk batch size 512, optimizer Adam, dan fungsi loss BCEWithLogitsLoss.

Hasil validasi silang menunjukkan bahwa EfficientNetB0 secara konsisten unggul atas dua model lainnya, dengan rata-rata F1-score $0,563 \pm 0,009$, akurasi $0,773 \pm 0,005$, dan AUC-ROC $0,772 \pm 0,005$. Pada tahap pengujian akhir menggunakan 2.500 citra, model ini kembali menunjukkan performa terbaik dengan macro F1-score 0,556, recall 0,551, dan AUC-ROC 0,765. Deteksi pneumotoraks menjadi yang paling akurat (AUC-ROC 0,880, F1-score 0,678), sedangkan kardiomegali dan nodul/massa memiliki performa terendah, mengindikasikan tantangan dalam mendeteksi kelainan yang bersifat halus atau kecil.

Dari sisi efisiensi, ShuffleNetV2 mencatat waktu inferensi tercepat ($\sim 17,6$ ms per citra) dan jumlah parameter paling sedikit ($\sim 1,26$ juta), menjadikannya kandidat ideal untuk deployment di perangkat dengan keterbatasan sumber daya. Sebaliknya, meskipun EfficientNetB0 memiliki kompleksitas lebih tinggi (~ 4 juta parameter, ~ 33 ms per citra), model ini memberikan akurasi dan sensitivitas yang lebih baik, menjadikannya pilihan utama untuk sistem pendukung diagnosis berbasis AI. Visualisasi Grad-CAM menunjukkan bahwa model ini mampu mengarahkan fokus ke area patologis, terutama infiltrasi, namun tetap menunjukkan kelemahan dalam mendeteksi kardiomegali akibat kurangnya aktivasi pada area jantung.

Untuk memvalidasi asumsi penerapan di perangkat terbatas, dilakukan deployment dan pengujian langsung pada emulator Android dengan konfigurasi minimal (1 core CPU, 2 GB RAM, Android 14). Model berhasil dijalankan secara fungsional dan inference time tercatat paling efisien pada ShuffleNetV2 (rata-rata 103 ms), disusul oleh EfficientNetB0 dan MobileNetV2. Hasil ini menunjukkan bahwa seluruh model dapat dioperasikan pada perangkat edge, serta memperkuat relevansi pendekatan ini dalam mendukung sistem deteksi penyakit paru di lokasi dengan keterbatasan infrastruktur.

Kata kunci

Deteksi Abnormalitas Toraks; *Deep Learning*; Model *Lightweight*; *Chest X-ray (CXR)*; Grad-CAM

Abstract

Efficient Thoracic Abnormalities Detection Using Mobile Deep Learning Models: Comparative Study of MobileNet, ShuffleNet, dan EfficientNet with Grad-CAM

The high burden of pulmonary diseases in Indonesia—such as tuberculosis, pneumonia, and heart failure—is exacerbated by the shortage of radiology specialists, necessitating an efficient, AI-based technological solution that can operate in resource-constrained environments. This study proposes the use of three lightweight deep learning models—MobileNetV2, ShuffleNetV2, and EfficientNetB0—for detecting five major thoracic abnormalities in chest X-ray (CXR) images, namely pneumothorax, cardiomegaly, nodules/masses, consolidation, and infiltration. The NIH ChestX-ray8 dataset was used, comprising 12,500 normalized and class-balanced images. The models were trained using 5-fold cross-validation with standardized hyperparameters, including a batch size of 512, Adam optimizer, and BCEWithLogitsLoss as the loss function.

Cross-validation results demonstrated that EfficientNetB0 consistently outperformed the other two models, with an average F1-score of 0.563 ± 0.009 , accuracy of 0.773 ± 0.005 , and AUC-ROC of 0.772 ± 0.005 . In the final test phase using 2,500 unseen images, EfficientNetB0 again achieved the highest performance with a macro F1-score of 0.556, recall of 0.551, and AUC-ROC of 0.765. Pneumothorax detection was the most accurate (AUC-ROC 0.880, F1-score 0.678), while cardiomegaly and nodules/masses showed the lowest performance, indicating challenges in detecting visually subtle or small-scale abnormalities.

In terms of efficiency, ShuffleNetV2 achieved the fastest inference time (~ 17.6 ms per image) and the smallest parameter count (~ 1.26 million), making it an ideal candidate for deployment in low-resource devices. Conversely, although EfficientNetB0 has higher complexity (~ 4 million parameters, ~ 33 ms per image), it offers superior accuracy and sensitivity, making it the most reliable model for AI-assisted diagnostic systems. Grad-CAM visualizations indicate that EfficientNetB0 can accurately localize pathological regions, particularly for infiltration, though it still tends to underperform in cardiomegaly cases due to insufficient activation around the cardiac region.

To validate the assumption of real-world deployment in resource-constrained settings, the models were deployed and tested on an Android emulator with minimal configuration (1 CPU core, 2 GB RAM, Android 14). All models successfully executed, with inference times

confirming ShuffleNetV2 as the most efficient (average 103 ms), followed by EfficientNetB0 and MobileNetV2. These results demonstrate the feasibility of operating the proposed models on edge devices and support their application as AI-assisted diagnostic tools in low-infrastructure healthcare environments.

Keywords

Thoracic Abnormality Detection; Deep Learning; Lightweight Model; Chest X-Ray (CXR); Grad-CAM

Pernyataan Keaslian Tulisan

Dengan ini saya menyatakan bahwa tesis ini merupakan tulisan asli dari penulis, dan tidak berisi material yang telah diterbitkan sebelumnya atau tulisan dari penulis lain terkecuali referensi atas material tersebut telah disebutkan dalam tesis. Apabila ada kontribusi dari penulis lain dalam tesis ini, maka penulis lain tersebut secara eksplisit telah disebutkan dalam tesis ini.

Dengan ini saya juga menyatakan bahwa segala kontribusi dari pihak lain terhadap tesis ini, termasuk bantuan analisis statistik, desain survei, analisis data, prosedur teknis yang bersifat signifikan, dan segala bentuk aktivitas penelitian yang dipergunakan atau dilaporkan dalam tesis ini telah secara eksplisit disebutkan dalam tesis ini.

Segala bentuk hak cipta yang terdapat dalam material dokumen tesis ini berada dalam kepemilikan pemilik hak cipta masing-masing. Apabila dibutuhkan, penulis juga telah mendapatkan izin dari pemilik hak cipta untuk menggunakan ulang materialnya dalam tesis ini.

Yogyakarta, Juni 2025



Achmad Bauravindah

Daftar Publikasi

Publikasi berikut menjadi bagian dari penelitian ini.

Bauravindah, Achmad, Dhomas Hatta Fudholi, and Rima Tri Wahyuningrum. 2025. "Efficient Thoracic Abnormalities Detection Using Mobile Deep Learning Models." Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control. doi:10.22219/kinetik.v10i3.2268.

Kontributor	Jenis Kontribusi
Achmad Bauravindah	Menulis draft dan revisi artikel (30%) Mendesain eksperimen (25%)
Ir. Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D.	Meninjau naskah (20%) Memberi arahan metodologi (10%)
Prof. Dr. Rima Tri Wahyuningrum	Menginisiasi proyek (10%) Memberi masukan pada struktur dan interpretasi (5%)

Halaman Kontribusi

Penelitian ini memperoleh kontribusi signifikan dari Ir. Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D., selaku pembimbing utama, yang memberikan arahan mendalam dalam desain riset, pemilihan pendekatan teknis, serta evaluasi menyeluruh terhadap proses dan hasil penelitian ini. Kontribusi beliau sangat menentukan arah dan kedalaman analisis dalam tesis ini.

Kontribusi penting juga diberikan oleh Prof. Dr. Rima Tri Wahyuningrum, S.T., M.T. dari Universitas Trunojoyo Madura, yang menginisiasi ide awal proyek penelitian ini serta membantu secara aktif dalam proses review dan penyempurnaan artikel ilmiah yang terkait dengan topik tesis.

Selain itu, Dr. Syarif Hidayat, S.Kom., M.IT. dan Chandra Kusuma Dewa, S.Kom., M.Kom., Ph.D. turut memberikan masukan substansial terhadap struktur tesis dan analisis eksperimen yang meningkatkan kualitas keseluruhan karya.

Penulis juga memperoleh perspektif tambahan dari diskusi riset selama program akademik internasional di Kumamoto University dan University of South-Eastern Norway, yang memberikan wawasan teknis dan kontekstual dalam pengolahan data dan pengembangan model.

Kontribusi-kontribusi tersebut secara kolektif berperan penting dalam menghasilkan penelitian ini.

Halaman Persembahan

Tesis ini penulis persembahkan kepada orang tua tercinta yang telah menjadi sumber doa, kasih sayang, dan semangat tak terbatas sepanjang perjalanan akademik ini; kepada dosen pembimbing dan penguji yang telah membimbing dengan sabar, memberikan ilmu, arahan, dan evaluasi yang membentuk kualitas karya ini; kepada Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia atas kepercayaan dan dukungan finansial melalui Hibah Tesis Magister yang sangat membantu keberlangsungan penelitian ini; kepada institusi mitra internasional seperti Kumamoto University dan University of South-Eastern Norway yang telah membuka ruang bagi penulis untuk memperluas wawasan dan pengalaman riset di lingkungan akademik global; serta kepada seluruh rekan-rekan seperjuangan di Program Studi Magister Informatika Universitas Islam Indonesia yang telah menjadi bagian dari proses bertumbuh, berbagi, dan belajar bersama. Semoga karya ini menjadi amal jariyah dalam bentuk ilmu yang bermanfaat bagi pengembangan keilmuan dan kontribusi nyata bagi masyarakat.

Kata Pengantar

Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas rahmat dan kemudahannya sehingga penulis dapat menyelesaikan tesis ini yang berjudul “Efficient Thoracic Abnormalities Detection Using Mobile Deep Learning Models”. Tesis ini disusun sebagai salah satu syarat untuk memperoleh gelar Magister Informatika di Program Magister Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia. Penelitian ini berfokus pada penerapan model deep learning ringan untuk mendeteksi kelainan toraks dari citra X-ray, sebagai upaya untuk menyediakan solusi diagnosis berbasis kecerdasan buatan di lingkungan yang terbatas secara komputasi maupun sumber daya medis.

Penulis menyampaikan penghargaan yang sebesar-besarnya kepada Ir. Irving Vitra Paputungan, S.T., M.Sc., Ph.D., selaku Ketua Program Studi Magister Informatika, atas dukungan dan arahnya selama masa studi. Ucapan terima kasih secara khusus penulis sampaikan kepada Ir. Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D., selaku dosen pembimbing utama, yang telah membimbing dengan sabar dan penuh ketelitian dalam setiap tahapan penelitian dan penulisan tesis ini. Penulis juga mengucapkan terima kasih yang mendalam kepada Prof. Dr. Rima Tri Wahyuningrum, S.T., M.T. dari Universitas Trunojoyo Madura, yang telah menginisiasi topik penelitian ini serta memberikan kontribusi substansial dalam proses penyusunan dan peninjauan karya ilmiah yang terkait. Penghargaan yang sama penulis sampaikan kepada Dr. Syarif Hidayat, S.Kom., M.IT. dan Chandra Kusuma Dewa, S.Kom., M.Kom., Ph.D., selaku dosen penguji, atas masukan dan evaluasi yang sangat bermanfaat dalam penyempurnaan tesis ini.

Penulis juga mengucapkan terima kasih kepada Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia atas dukungan finansial melalui Hibah Tesis Magister, serta kepada institusi mitra internasional seperti Kumamoto University (Jepang) dan University of South-Eastern Norway (Norwegia) atas pengalaman akademik yang memperluas perspektif dan memperkaya pemahaman penulis dalam bidang penelitian ini. Tidak lupa, penulis menyampaikan rasa terima kasih yang tulus kepada seluruh dosen dan staf Departemen Informatika UII serta kepada keluarga dan sahabat atas doa, semangat, dan dukungan yang tidak pernah surut.

Penulis menyadari bahwa tesis ini masih memiliki kekurangan. Oleh karena itu, segala bentuk kritik dan saran yang membangun sangat diharapkan. Semoga karya ini dapat memberikan manfaat yang nyata, khususnya dalam pengembangan teknologi kecerdasan buatan di bidang kesehatan dan pelayanan medis di Indonesia.

Daftar Isi

BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	5
1.3 Batasan Masalah	6
1.4 Tujuan Penelitian	6
1.5 Manfaat Penelitian	6
1.6 Sistematika Penulisan	7
BAB 2 TINJAUAN PUSTAKA	8
2.1 Landasan Teori	8
2.1.1 Prinsip dan Peran Pencitraan Medis (CXR) dalam Diagnostik	8
2.1.2 <i>Deep Learning</i>	8
2.1.3 Convolutional Neural Networks (CNNs)	10
2.1.4 MobileNet.....	15
2.1.5 ShuffleNet.....	17
2.1.6 EfficientNet	20
2.1.7 Grad-CAM.....	21
2.1.8 K-Fold Cross Validation.....	23
2.1.9 Metrik Evaluasi.....	24
2.2 Kajian Pustaka	26
BAB 3 METODOLOGI PENELITIAN.....	34
3.1 Dataset	34
3.2 Skenario Penelitian	38
3.3 Spesifikasi Lingkungan Eksperimen	41
3.4 Konfigurasi <i>Hyperparameter</i>	41

BAB 4 HASIL DAN PEMBAHASAN.....	44
4.1 Analisis Model pada Saat Pelatihan	44
4.2 Evaluasi Kinerja Model pada Dataset Validasi	46
4.3 Evaluasi Kinerja Model pada Dataset Pengujian	48
4.3.1 Analisis Hasil Pengujian MobileNetV2	49
4.3.2 Analisis Hasil Pengujian ShuffleNetV2	51
4.3.3 Analisis Hasil Pengujian EfficientNetB0	52
4.3.4 Analisis Macro-Average pada Semua Metrik	55
4.4 Analisis Performa Model secara Keseluruhan.....	56
4.4.1 Performa Efisiensi Pelatihan	56
4.4.2 Performa Pengujian secara Menyeluruh.....	57
4.5 Analisis Visualisasi grad-CAM.....	64
BAB 5 KESIMPULAN DAN SARAN.....	66
5.1 Kesimpulan.....	66
5.2 Saran	66

Daftar Tabel

Tabel 3. 2 Dataset yang Digunakan	36
Tabel 3. 3 Pembagian Dataset untuk Pelatihan Model	36
Tabel 3. 4 Spesifikasi Perangkat Keras dan Perangkat Lunak Eksperimen	41
Tabel 3. 5 Konfigurasi <i>Hyperparameter</i>	41
Tabel 4. 1 Kinerja Model pada Data Validasi	47
Tabel 4. 2 F1-Score Setiap Fold pada Tiap-tiap Model	48
Tabel 4. 3 Pemilihan Model Terbaik Berdasarkan F1-Score Tertinggi	48
Tabel 4. 4 Hasil Pengujian MobileNetV2	49
Tabel 4. 5 Hasil Pengujian ShuffleNetV2	51
Tabel 4. 6 Hasil Pengujian EfficientNetB0	53
Tabel 4. 7 Perbandingan AUC-ROC Model.....	54
Tabel 4. 8 Macro-Average Kinerja Model	55
Tabel 4. 9 Akurasi, Waktu Pengujian, dan Kompleksitas Model.....	57
Tabel 4. 9 Hasil Prediksi dari 1 Citra pada Analisis Visualisasi Grad-CAM.....	64

Daftar Gambar

Gambar 4. 1 Learning Curve dari MobileNetV2.....	44
Gambar 4. 2 Learning Curve dari ShuffleNetV2	45
Gambar 4. 3 Learning Curve dari EfficientNetB0	46
Gambar 4. 4 ROC-Curve dari MobileNetV2.....	50
Gambar 4. 5 ROC-Curve dari ShuffleNetV2	52
Gambar 4. 6 ROC-Curve dari EfficientNetB0	54
Gambar 4. 7 Waktu Pelatihan (dalam menit) untuk Setiap Model.....	56
Gambar 4. 8 Kinerja Model secara Komprehensif dalam <i>Bubble Chart</i>	58
Gambar 4. 7 Citra Paru dengan Visualisasi Grad-CAM.	64

Glosarium

- a) ***Convolutional Neural Network (CNN)***: Sebuah arsitektur jaringan saraf yang dirancang khusus untuk pemrosesan data grid, seperti gambar, dengan menggunakan operasi konvolusi untuk mendeteksi pola-pola dalam gambar.
- b) **MobileNet**: Arsitektur deep learning yang dirancang untuk perangkat mobile dengan daya komputasi rendah, menggunakan teknik depthwise separable convolutions untuk mengurangi komputasi tanpa mengorbankan akurasi.
- c) **ShuffleNet**: Model deep learning yang mengurangi beban komputasi dengan menggunakan konvolusi grup pointwise dan pengacakan channel, dirancang untuk efisiensi pada perangkat dengan sumber daya terbatas.
- d) **EfficientNet**: Sebuah model deep learning yang menggunakan skala majemuk untuk menyeimbangkan kedalaman, lebar, dan resolusi jaringan, menghasilkan kinerja yang efisien dengan parameter lebih sedikit.
- e) **Depthwise Separable Convolution**: Teknik yang digunakan dalam arsitektur MobileNet untuk memecah operasi konvolusi menjadi dua tahap—depthwise convolution dan pointwise convolution—sehingga mengurangi jumlah parameter atau kompleksitas model dan komputasi.
- f) **Gradient-weighted Class Activation Mapping (Grad-CAM)**: Sebuah teknik interpretabilitas dalam deep learning yang memvisualisasikan area gambar yang paling berkontribusi terhadap prediksi model, membantu dalam memahami hasil diagnostik.
- g) **K-Fold Cross Validation**: Teknik evaluasi model di mana data dibagi menjadi beberapa subset (lipatan) dan model diuji pada lipatan yang berbeda untuk meminimalkan bias dan varians, meningkatkan generalisasi model.
- h) **Accuracy**: Metrik evaluasi dalam deep learning yang mengukur proporsi prediksi yang benar terhadap total jumlah prediksi, digunakan untuk menilai performa keseluruhan model.
- i) **AUC-ROC (Area Under the Receiver Operating Characteristic Curve)**: Metrik yang mengukur kemampuan model deep learning dalam membedakan antara kelas positif dan negatif, dengan mempertimbangkan trade-off antara true positive rate dan false positive rate.

BAB 1

Pendahuluan

1.1 Latar Belakang

Indonesia menghadapi tantangan kesehatan masyarakat yang serius dengan prevalensi tinggi penyakit menular dan tidak menular, terutama tuberkulosis (TB) dan pneumonia. Berdasarkan laporan Global TB Report 2023, Indonesia menduduki peringkat kedua dunia dalam kasus TB dengan estimasi sebanyak 1.060.000 kasus per tahun dan sekitar 134.000 kematian per tahun, yang setara dengan sekitar 15 kematian setiap jamnya (Wulan 2024). Pneumonia merupakan penyebab utama kematian akibat infeksi pada anak di bawah usia lima tahun di Indonesia, yang mempengaruhi sekitar setengah juta anak setiap tahun dan menyebabkan sekitar 10.000 kematian (Dayne 2021). Penyakit-penyakit ini, bersamaan dengan penyakit paru obstruktif kronik (PPOK), kanker paru-paru, gagal jantung, dan penyakit paru interstisial, secara signifikan turut menyumbang angka kesakitan (morbiditas) dan kematian (mortalitas) yang tinggi (Andarini et al. 2023; Muharram et al. 2024; Shah Gupta et al. 2023; Statista 2024). Diagnosis yang cepat dan akurat sangat penting dalam pengendalian dan pengobatan penyakit-penyakit ini, namun hal tersebut masih sulit dilakukan di Indonesia akibat keterbatasan sumber daya dan kurangnya jumlah dokter spesialis radiologi. Saat ini Indonesia hanya memiliki sekitar 2.161 radiolog terdaftar, yang berarti hanya sekitar 1,2 radiolog per 100.000 penduduk. Selain itu, Indonesia kekurangan sekitar 31.481 dokter spesialis, termasuk radiolog, yang semakin memperparah kondisi pelayanan kesehatan, terutama di daerah terpencil (nela 2023; Yunus 2024). Situasi ini menyebabkan lambatnya proses diagnosis dan penanganan pasien sehingga beban kesehatan semakin berat (nela 2023).

X-ray dada (CXR) merupakan alat pencitraan yang penting untuk mendeteksi berbagai gangguan paru, termasuk TB, pneumonia, dan penyakit serius lainnya (Bradley et al. 2019; Inui et al. 2021; Jewell 2023). Sebagai metode non-invasif yang murah dan tersedia luas, CXR berperan besar dalam diagnosis penyakit pernapasan. Namun, menginterpretasi gambar CXR membutuhkan keahlian tinggi karena kelainan pada gambar bisa samar atau saling bertumpang tindih, sehingga menyulitkan diagnosis yang tepat, khususnya di fasilitas kesehatan dengan sumber daya terbatas (Eisen et al. 2006; Rafferty 2005).

Pada pencitraan dada, TB biasanya ditandai dengan *nodule*, *mass*, dan *consolidation*; *nodule* atau *mass* tersebut dapat mengindikasikan infeksi aktif, granuloma, atau kanker paru,

sedangkan tuberkuloma ditemukan pada sekitar 5% kasus TB post-primer, kadang membentuk kavitas atau rongga (Bomanji et al. 2015; Jeong and Lee 2008). Pneumonia, terutama pada anak-anak, biasanya ditunjukkan oleh *consolidation* karena adanya cairan dalam alveoli dan *infiltration* berupa opasitas difus atau bercak-bercak, sering kali terkait dengan infeksi virus seperti influenza (Gaillard 2024; Lee et al. 2013; Pochepnia et al. 2024). PPOK umumnya menunjukkan gambaran *pneumothorax* (udara dalam rongga pleura) yang menyebabkan kolaps paru dan *infiltration* akibat peradangan yang merusak saluran napas dan jaringan paru (Cheng et al. 2016; clevelandclinic 2023; King 2015; McKnight and Burns 2025; Rangelov et al. 2020). Kanker paru tampak dalam bentuk *nodule* atau *mass*, dengan keganasan dicurigai apabila lesi lebih dari 3 cm atau menunjukkan pertumbuhan yang tidak beraturan (cancer.org 2024; utswmed 2025). Sementara gagal jantung menunjukkan gambaran kongesti paru dan *cardiomegaly* (pembesaran jantung), terlihat pada rasio kardioraks lebih dari 50% pada CXR yang mengindikasikan disfungsi ventrikel kiri atau efusi perikardium (Amin and Siddiqui 2025; Brakohiapa et al. 2017; clevelandclinic 2022; Cundrle, Olson, and Johnson 2019; Metkus 2024; Ridley 2019). Kelainan-kelainan tersebut—seperti *pneumothorax*, *cardiomegaly*, *nodule/mass*, *consolidation*, dan *infiltration*—menjadi tanda penting penyakit paru serius, namun interpretasi akurat membutuhkan keahlian radiologi yang memadai, terutama di fasilitas kesehatan terbatas.

Deep learning (pembelajaran mendalam) menawarkan solusi menjanjikan dalam menghadapi tantangan pencitraan medis, khususnya mendeteksi kelainan-kelainan tersebut. Alih-alih langsung mendiagnosis penyakit, model *deep learning* membantu analisis lebih rinci dengan mengenali tanda-tanda yang mengindikasikan penyakit tertentu. Sebagai contoh, *nodule* pada CXR bisa menunjukkan berbagai penyakit, seperti TB atau kanker paru, tergantung kondisi klinis lain. Dengan mengidentifikasi kelainan ini, model *deep learning* membantu radiolog menghasilkan diagnosis lebih cepat dan akurat, yang sangat penting di fasilitas dengan sumber daya terbatas.

Penggunaan alat kecerdasan buatan (*Artificial Intelligence/AI*) untuk proses pra-skrining dan identifikasi awal kelainan berpotensi besar mempercepat waktu yang dibutuhkan oleh radiolog dalam menyusun laporan diagnostik (Anderson et al. 2024; Guo et al. 2024; Kvak et al. 2023; Nam et al. 2021; Wang et al. 2017). Hal ini memungkinkan radiolog lebih fokus pada area yang memang membutuhkan perhatian khusus. Keuntungan ini sangat dirasakan terutama di lingkungan kerja yang padat, seperti rumah sakit besar atau selama terjadi wabah penyakit, di mana laporan diagnosis yang cepat sangat dibutuhkan.

Namun demikian, terdapat beberapa tantangan dalam implementasinya, terutama terkait keterbatasan perangkat keras di daerah yang memiliki sumber daya terbatas atau fasilitas kesehatan di wilayah pedesaan (Blantz 2010). Banyak rumah sakit dan klinik di Indonesia belum memiliki akses ke perangkat komputasi canggih seperti server atau komputer dengan kemampuan GPU tinggi (PricewaterhouseCoopers 2023), disebabkan tingginya biaya investasi teknologi tersebut (Sarkar 2021). Oleh karena itu, dibutuhkan pengembangan model-model deep learning yang tidak hanya memiliki tingkat akurasi tinggi, tetapi juga dapat beroperasi secara efektif pada perangkat dengan kapasitas komputasi yang rendah.

Model-model berbasis *mobile* seperti MobileNet, ShuffleNet, dan EfficientNet menawarkan solusi yang efisien dalam bidang pencitraan medis, karena mampu mengoptimalkan kecepatan komputasi tanpa mengurangi akurasi hasil diagnosis (Howard et al. 2017; Tan and Le 2020; Zhang et al. 2017a). Model ini memiliki efisiensi tinggi sehingga mudah diaplikasikan pada perangkat seluler atau sistem dengan konsumsi daya rendah, menjadikannya pilihan ideal bagi fasilitas kesehatan dengan keterbatasan teknologi. Kemudahan akses tersebut sangat bermanfaat bagi wilayah dengan infrastruktur kurang memadai, sebab penggunaan smartphone atau tablet memungkinkan diagnosis yang cepat dan akurat, terutama di daerah pedesaan atau wilayah berkembang (Blantz 2010).

Di samping itu, kecepatan pemrosesan yang tinggi memungkinkan radiolog untuk segera menyelesaikan laporan diagnostik, suatu keunggulan penting dalam kondisi darurat kesehatan atau volume pasien yang tinggi (van Leeuwen et al. 2022). Secara ekonomi, penggunaan model-model *mobile* ini mengurangi biaya operasional dengan menurunkan kebutuhan akan perangkat keras yang mahal, sehingga rumah sakit dapat menerapkan teknologi AI tanpa investasi infrastruktur yang besar (Howard et al. 2017). Selain itu, skalabilitas dari model-model ini memungkinkan integrasi luas dalam berbagai layanan kesehatan, mulai dari klinik kecil hingga platform telemedisin, mendukung implementasi yang lebih luas serta pemanfaatan secara jarak jauh (Rybczak and Kozakiewicz 2024).

MobileNet menggunakan teknik *depthwise separable convolution* untuk mengurangi beban komputasi sehingga sangat cocok untuk penggunaan perangkat seluler (Howard et al. 2017). ShuffleNet memanfaatkan metode *pointwise group convolution* dan *channel shuffling* yang menjaga akurasi sekaligus menekan biaya komputasi (Zhang et al. 2017a). Sedangkan EfficientNet menerapkan teknik *multi-scale compound scaling* untuk menyeimbangkan dimensi jaringan, mencapai performa optimal dengan parameter yang lebih sedikit (Tan and Le 2020). Secara keseluruhan, model-model ini menawarkan solusi diagnosis yang cepat,

terjangkau, dan mudah dikembangkan, sangat tepat digunakan dalam lingkungan layanan kesehatan dengan sumber daya terbatas.

Selain efisiensi arsitektural, model-model ini juga menunjukkan kinerja unggul dalam klasifikasi citra X-ray dada. MobileNetV2 terbukti sangat akurat dan cepat, mencapai akurasi hingga 98,65% dan recall 98,15% dalam deteksi pneumonia dan COVID-19 (Kolonne et al. 2021), serta menempati peringkat tertinggi dalam kecepatan dan akurasi di antara 11 arsitektur CNN lainnya (Akter et al. 2021). ShuffleNetV2 juga menunjukkan akurasi tinggi (91,2%) dengan parameter yang ringan, sangat cocok untuk sistem embedded (Gu and Lee 2024). Sementara itu, EfficientNetB0 mencatatkan akurasi mencapai 99,78% dalam klasifikasi abnormalitas paru multi-sumber, dan konsisten melampaui model yang lebih dalam seperti ResNet-50 (Kansal, Chandra, and Singh 2024).

Untuk meningkatkan interpretabilitas atau kemudahan interpretasi dari model-model AI dalam aplikasi medis, teknik seperti *Gradient-weighted Class Activation Mapping* (Grad-CAM) telah banyak digunakan (Nam et al. 2021; Selvaraju et al. 2020; Suara et al. 2024). Grad-CAM menyediakan visualisasi yang menunjukkan bagian-bagian gambar mana yang paling berkontribusi terhadap prediksi yang dihasilkan oleh model, sehingga membantu tenaga medis memahami sekaligus mempercayai hasil diagnosis berbasis AI. Dalam konteks deteksi kelainan toraks, misalnya, Grad-CAM mampu menunjukkan area paru-paru yang mengalami gangguan atau penyakit secara jelas, sehingga memudahkan dokter dalam mengidentifikasi lokasi patologi pada citra rontgen dada (CXR). Kemampuan visualisasi ini sangat bernilai, terutama di lingkungan klinis yang menuntut transparansi serta keterjangkauan teknologi AI sebagai prasyarat untuk memperoleh kepercayaan dan penerimaan luas dari para profesional kesehatan.

Evaluasi terhadap model *deep learning* dalam pencitraan medis umumnya dilakukan melalui serangkaian prosedur pengujian yang ketat untuk memastikan ketahanan dan kemampuan generalisasi model. Salah satu teknik yang umum digunakan adalah K-Fold Cross Validation, yaitu metode pembagian data menjadi beberapa subset atau fold, di mana model divalidasi secara bergiliran pada setiap fold untuk mengurangi potensi bias dan varians (Le et al. 2024; Refaeilzadeh, Tang, and Liu 2009).

Selain itu, berbagai metrik performa seperti *accuracy* (akurasi), *precision* (presisi), *recall* (sensitivitas), dan *F1 score* digunakan untuk mengevaluasi kinerja model secara menyeluruh, mencakup aspek ketepatan dan konsistensi prediksi (Naidu, Zuva, and Sibanda 2023). Metrik lainnya yang juga sangat penting adalah *Area Under the Receiver Operating*

Characteristic Curve (AUC-ROC), yang digunakan untuk mengukur tingkat akurasi diagnostik dari model.

Kurva AUC-ROC menunjukkan keseimbangan antara *true positive rate* dan *false positive rate*, memberikan gambaran mendalam tentang kemampuan model dalam membedakan antara dua kelas—dalam konteks ini, keberadaan atau ketiadaan kelainan toraks (Melo 2013). Evaluasi yang komprehensif semacam ini sangat krusial untuk memastikan bahwa model AI yang digunakan dalam diagnosis medis benar-benar andal dan dapat diandalkan di berbagai kondisi klinis.

Integrasi model-model *deep learning* yang ringan dengan teknik evaluasi dan visualisasi canggih bertujuan untuk meningkatkan efisiensi serta akurasi dalam deteksi kelainan, khususnya di lingkungan kesehatan yang memiliki keterbatasan sumber daya radiologi konvensional. Melalui pemanfaatan model seperti MobileNetV2, ShuffleNetV2, dan EfficientNetB0 penelitian ini bertujuan mengembangkan alat diagnosis yang efektif sekaligus ramah terhadap perangkat mobile, sehingga dapat berfungsi optimal pada peralatan dengan kemampuan komputasi terbatas. Pendekatan tersebut tidak hanya menjawab kebutuhan mendesak akan solusi diagnosis yang mudah diimplementasikan secara luas, tetapi juga turut berkontribusi pada tujuan jangka panjang untuk menjadikan teknologi pencitraan medis mutakhir lebih mudah diakses, ekonomis, dan praktis untuk digunakan secara luas dalam praktik klinis, khususnya di daerah-daerah yang masih kekurangan fasilitas kesehatan.

1.2 Rumusan Masalah

Indonesia menghadapi tantangan kesehatan masyarakat yang signifikan akibat tingginya angka prevalensi penyakit infeksi dan tidak infeksi, khususnya tuberkulosis (TBC) dan pneumonia. Tingginya angka morbiditas dan mortalitas akibat penyakit ini diperburuk oleh keterbatasan sumber daya, termasuk kekurangan ahli radiologi yang dapat menghambat proses diagnosis dini dan akurat. Hal ini menimbulkan pertanyaan tentang bagaimana teknologi *deep learning* dapat membantu dalam mendeteksi abnormalitas-abnormalitas toraks yang sering dikaitkan dengan TBC, pneumonia, penyakit paru obstruktif kronik (PPOK), kanker paru-paru, dan gagal jantung, serta bagaimana teknologi ini dapat diharapkan bekerja secara efektif di Indonesia yang memiliki keterbatasan sumber daya.

1.3 Batasan Masalah

1. Penelitian ini dibatasi pada penggunaan citra rontgen dada (CXR) untuk mendeteksi abnormalitas toraks. Citra medis lain seperti CT scan atau MRI tidak termasuk dalam penelitian ini.
2. Penelitian ini hanya fokus pada abnormalitas toraks tertentu, yaitu *pneumothorax*, *cardiomegaly*, *nodule/mass*, *consolidation*, dan *infiltration*. Abnormalitas lain tidak akan dianalisis.
3. Penelitian ini akan menggunakan model deep learning tertentu, yaitu MobileNetV2, ShuffleNetV2, dan EfficientNetB0. Model lain di luar ketiga model ini tidak akan dibahas atau dievaluasi.
4. Penelitian ini mengasumsikan bahwa implementasi model deep learning dilakukan pada perangkat dengan sumber daya terbatas, seperti perangkat mobile atau perangkat dengan daya komputasi rendah.
5. Evaluasi model akan dilakukan berdasarkan metrik kinerja tertentu, yaitu *accuracy*, *precision*, *recall*, dan *F1 score*, dan AUC-ROC. Metrik lain tidak akan digunakan dalam analisis.
6. Teknik interpretabilitas yang digunakan dalam penelitian ini akan dibatasi pada penggunaan Grad-CAM. Teknik interpretabilitas lain tidak akan dieksplorasi.
7. Penelitian ini hanya mencakup pengujian performa model, sehingga tidak dilakukan implementasi klinis secara langsung.
8. Penelitian ini akan menggunakan data CXR yang tersedia secara publik atau yang telah dikumpulkan sebelumnya. Pengumpulan data baru tidak akan dilakukan sebagai bagian dari penelitian ini.

1.4 Tujuan Penelitian

Menganalisis performa model deep learning MobileNetV2, ShuffleNetV2, dan EfficientNetB0 dalam mendeteksi abnormalitas toraks pada CXR.

1.5 Manfaat Penelitian

1. Manfaat Akademis: Memberikan kontribusi terhadap pengembangan ilmu pengetahuan di bidang kecerdasan buatan, khususnya dalam aplikasi *deep learning* untuk pencitraan medis.
2. Manfaat Praktis: Menyediakan solusi diagnostik yang efisien dan terjangkau yang dapat diterapkan di rumah sakit atau klinik di Indonesia, terutama di daerah-daerah dengan keterbatasan sumber daya.

3. Manfaat Sosial: Meningkatkan akses terhadap diagnosis medis yang cepat dan akurat, yang dapat membantu dalam menurunkan angka morbiditas dan mortalitas akibat penyakit toraks di Indonesia.
4. Manfaat Kebijakan: Memberikan rekomendasi kepada pembuat kebijakan kesehatan tentang potensi penggunaan teknologi *deep learning* dalam mendukung sistem kesehatan nasional, terutama dalam penanganan penyakit infeksi dan tidak infeksi yang mempengaruhi paru-paru.

1.6 Sistematika Penulisan

Tesis ini disusun dalam lima bab utama yang saling terhubung secara sistematis untuk membentuk alur pemikiran yang utuh dari awal hingga akhir penelitian. Bab 1 merupakan pendahuluan yang berisi latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat, serta sistematika penulisan sebagai pengantar keseluruhan isi tesis. Bab 2 menyajikan landasan teori dan kajian pustaka yang mendasari penelitian, meliputi konsep pencitraan medis menggunakan CXR, arsitektur model deep learning seperti CNN, MobileNetV2, ShuffleNetV2, dan EfficientNetB0, serta teknik interpretabilitas Grad-CAM, metode validasi K-Fold, dan metrik evaluasi yang digunakan. Bab 3 menjelaskan metodologi penelitian yang terdiri dari deskripsi dataset, skenario eksperimen, spesifikasi lingkungan uji coba, serta konfigurasi hyperparameter yang digunakan dalam proses pelatihan model. Bab 4 menyajikan hasil eksperimen dan analisis performa model berdasarkan proses pelatihan dan pengujian, termasuk perbandingan antar model, efisiensi waktu, serta visualisasi prediksi menggunakan Grad-CAM. Terakhir, Bab 5 memuat kesimpulan dari hasil penelitian serta saran untuk pengembangan dan studi lanjutan yang relevan.

BAB 2

Tinjauan Pustaka

2.1 Landasan Teori

2.1.1 Prinsip dan Peran Pencitraan Medis (CXR) dalam Diagnostik

Dalam konteks diagnostik, pencitraan foto toraks (CXR) menjadi alat kunci dalam mendeteksi berbagai abnormalitas toraks seperti infeksi, tumor, dan perubahan struktural. Prinsip dasar fisika pencitraan sinar-X serta pengaturan paparan dan posisi pasien sangat penting untuk memastikan kualitas gambar yang dihasilkan. Ahli radiologi menggunakan pendekatan sistematis untuk menganalisis gambar CXR, mencari tanda-tanda abnormalitas seperti *pneumothorax*, *cardiomegaly*, *nodule/mass*, *consolidation*, dan *infiltration* yang memiliki implikasi klinis signifikan dalam menentukan diagnosis dan perawatan (Smithuis and Delden 2022). Dalam praktiknya, interpretasi CXR mengikuti pedoman standar yang membantu memastikan konsistensi dalam diagnosis, sementara teknologi kecerdasan buatan (AI) semakin banyak digunakan untuk mendukung akurasi dan kecepatan interpretasi, terutama di daerah dengan keterbatasan akses ke ahli radiologi (Mahmoud et al. 2023). Integrasi AI dalam proses diagnostik ini penting untuk meningkatkan hasil kesehatan di lingkungan dengan sumber daya terbatas, memungkinkan deteksi dini dan penanganan yang lebih tepat waktu.

2.1.2 Deep Learning

a) Pengantar Deep Learning

Deep learning adalah cabang dari machine learning yang berfokus pada penggunaan jaringan saraf dengan banyak lapisan (sering disebut sebagai jaringan “dalam”) untuk memodelkan pola kompleks dalam data (Talaie Khoei, Ould Slimane, and Kaabouch 2023). Teknik ini telah menjadi pendekatan yang dominan di berbagai bidang, termasuk pengenalan gambar (Archana and Jeevaraj 2024), pengolahan suara (Purwins et al. 2019), pemrosesan bahasa alami (Li 2018), dan diagnosis medis (Patrício, Neves, and Teixeira 2022).

Secara historis, *deep learning* berakar dari penelitian awal tentang jaringan saraf buatan, yang terinspirasi oleh struktur dan fungsi otak manusia (Wang and Raj 2017). Pengembangan *backpropagation* merupakan tonggak penting yang memungkinkan jaringan saraf belajar dari data dengan lebih efektif (Du and Swamy 2019). Namun, baru pada saat tersedia dataset yang besar, GPU yang kuat, dan arsitektur baru, *deep learning* mulai mengungguli metode *machine learning* tradisional, terutama dalam tugas yang

membutuhkan representasi data berdimensi tinggi, seperti klasifikasi gambar (Alzubaidi et al. 2021; Shiri et al. 2024).

Perbedaan utama antara *deep learning* dan *machine learning* tradisional terletak pada cara fitur diekstraksi dari data. Algoritma *machine learning* tradisional sering kali mengandalkan fitur yang dibuat secara manual, yang memerlukan keahlian khusus (Alzubaidi et al. 2021). Sebaliknya, model *deep learning* secara otomatis mempelajari representasi hierarkis dari data mentah, memungkinkan mereka menangkap pola rumit dengan sedikit intervensi manusia. Kemampuan ini telah menyebabkan adopsi luas *deep learning* di berbagai domain, termasuk pencitraan medis, di mana teknik ini terbukti sangat efektif dalam mengidentifikasi pola kompleks yang mungkin sulit dideteksi oleh ahli manusia (Zhang and Qie 2023).

b) Arsitektur Jaringan Syaraf

Di inti *deep learning* terdapat jaringan saraf, sebuah model komputasi yang terdiri dari lapisan-lapisan yang saling terhubung. Jaringan saraf umumnya terdiri dari tiga jenis lapisan: lapisan input, lapisan tersembunyi, dan lapisan output (Talaie Khoei et al. 2023). Lapisan input menerima data mentah yang kemudian diproses oleh satu atau lebih lapisan tersembunyi. Lapisan tersembunyi ini, yang terdiri dari banyak neuron, adalah tempat jaringan belajar untuk mengekstraksi dan mentransformasikan fitur. Akhirnya, lapisan output menghasilkan prediksi berdasarkan data yang telah diubah.

Neuron dalam lapisan-lapisan ini adalah unit komputasi dasar. Setiap neuron menerima input dari lapisan sebelumnya, menerapkan bobot pada setiap input, menjumlahkan input berbobot, dan kemudian melewati hasilnya melalui fungsi aktivasi (Alzubaidi et al. 2021). Fungsi aktivasi, seperti Rectified Linear Unit (ReLU), Sigmoid, dan Tanh, memperkenalkan non-linearitas ke dalam jaringan, memungkinkan jaringan untuk memodelkan hubungan kompleks dalam data (Dubey, Singh, and Chaudhuri 2022). ReLU sangat populer dalam *deep learning* karena kesederhanaannya dan efektivitasnya dalam mengatasi masalah gradien yang menghilang (*vanishing gradient*), di mana gradien menjadi terlalu kecil untuk jaringan belajar secara efektif (Agarap 2018).

c) Proses Pelatihan

Pelatihan model *deep learning* melibatkan pengoptimalan parameternya agar dapat secara akurat memetakan input ke output yang diinginkan. Proses ini dimulai dengan propagasi maju (*forward propagation*), di mana data input dilewatkan melalui jaringan, lapis demi lapis, untuk menghasilkan output. Perbedaan antara output jaringan dan target yang sebenarnya diukur menggunakan fungsi kerugian (*loss function*). Fungsi kerugian yang

umum digunakan termasuk *Mean Squared Error* (MSE) untuk tugas regresi dan *Cross-Entropy Loss* untuk tugas klasifikasi (Talaie Khoei et al. 2023).

Jaringan belajar dengan meminimalkan kerugian ini melalui proses yang disebut *backpropagation*. Selama *backpropagation*, jaringan menghitung gradien kerugian terhadap setiap parameter menggunakan aturan rantai (*chain rule*) dari kalkulus. Gradien ini menunjukkan seberapa besar setiap parameter harus disesuaikan untuk mengurangi kerugian (Rosenbaum 2022). Algoritma optimisasi, seperti *Stochastic Gradient Descent* (SGD) atau *Adaptive Moment Estimation* (Adam), kemudian digunakan untuk memperbarui parameter jaringan secara iteratif. SGD memperbarui parameter berdasarkan subset data acak (mini-batch), yang membantu dalam konvergensi lebih cepat dan menghindari minimum lokal (Tian, Zhang, and Zhang 2023). Adam, sebagai perluasan dari SGD, menggabungkan laju pembelajaran adaptif untuk setiap parameter, membuatnya lebih efisien dalam banyak kasus (Kingma and Ba 2017).

d) Ekstraksi Fitur dan Pembelajaran Representasi

Ekstraksi fitur adalah aspek penting dalam *machine learning*, di mana fitur relevan diambil dari data mentah untuk meningkatkan kinerja model. Dalam *deep learning*, proses ini diotomatisasi melalui pembelajaran representasi. Pembelajaran representasi memungkinkan model untuk menemukan fitur-fitur rumit pada berbagai tingkat abstraksi, yang penting untuk prediksi yang akurat (Elharrouss et al. 2022).

Fitur dangkal, yang umumnya ditemukan dalam *machine learning* tradisional, sering kali dibuat secara manual dan mewakili atribut dasar dari data. Sebaliknya, model *deep learning* mempelajari fitur mendalam yang menangkap pola-pola tingkat tinggi dan kompleks. Sebagai contoh, dalam pengenalan gambar, lapisan awal dari jaringan dalam mungkin belajar mendeteksi tepi dan tekstur, sementara lapisan yang lebih dalam mungkin mengenali struktur yang lebih kompleks seperti bentuk atau objek. Ekstraksi fitur hierarkis ini adalah salah satu alasan mengapa model *deep learning* unggul dalam tugas seperti pengenalan gambar dan suara (Zhan et al. 2024).

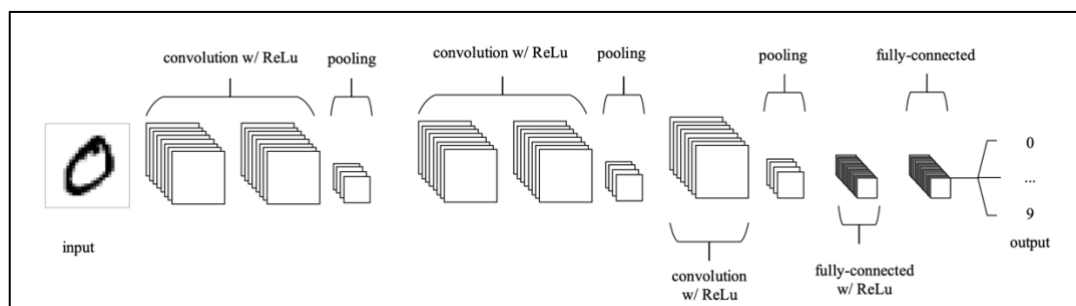
2.1.3 Convolutional Neural Networks (CNNs)

a) Pengantar CNNs

Jaringan Saraf Konvolusional (CNNs) adalah jenis jaringan saraf yang dirancang khusus untuk memproses struktur data grid, seperti gambar. Tidak seperti jaringan saraf penuh (*fully connected networks*), di mana setiap neuron dalam satu lapisan terhubung dengan setiap neuron di lapisan berikutnya, CNNs memanfaatkan hierarki spasial melalui penggunaan lapisan konvolusional. Struktur ini memungkinkan CNNs mendeteksi pola lokal

dalam data, seperti tepi atau tekstur dalam gambar, dan membangun fitur yang lebih kompleks seiring bertambahnya kedalaman jaringan (O'Shea and Nash 2015).

Perbedaan utama antara CNNs dan jaringan saraf penuh terletak pada arsitekturnya. CNNs menggunakan serangkaian lapisan konvolusional dan pooling untuk secara bertahap mengekstraksi fitur tingkat tinggi dari data input. Pendekatan ini mengurangi jumlah parameter atau kompleksitas komputasi, membuat CNNs sangat efisien untuk tugas pemrosesan gambar (O'Shea and Nash 2015). Gambar 2. 1. merupakan contoh arsitektur CNNs, yang terdiri dari input citra, lapisan konvolusional, lapisan pooling, lapisan jaringan saraf penuh dan output.



Gambar 2. 1 Arsitektur CNN (O'Shea and Nash 2015)

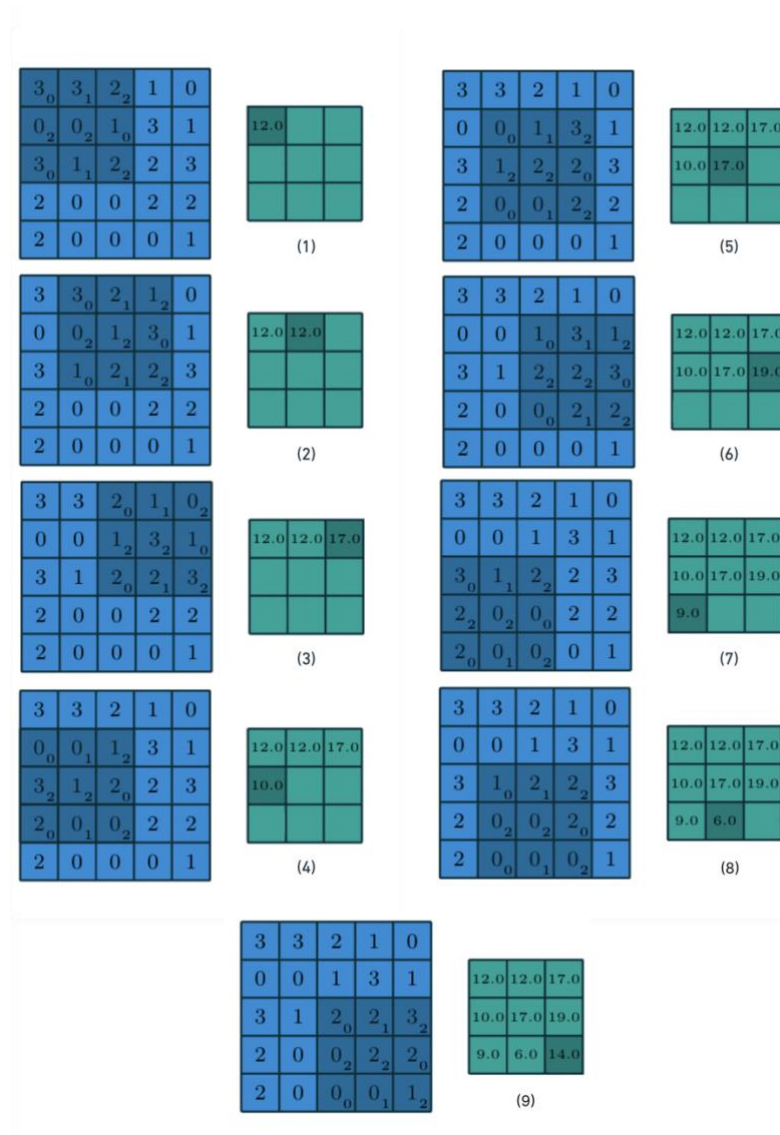
b) Lapisan Konvolusional

Lapisan konvolusional adalah blok bangunan utama dari CNN. Lapisan ini melakukan operasi matematika yang disebut konvolusi, di mana sebuah matriks kecil yang disebut kernel atau filter meluncur di atas data input, melakukan perkalian elemen demi elemen dan penjumlahan untuk menghasilkan peta fitur. Ukuran kernel, bersama dengan *stride* (ukuran langkah di mana filter bergerak), dan *padding* (penambahan piksel ekstra di sekitar input), menentukan dimensi output dari peta fitur (O'Shea and Nash 2015).

Lapisan konvolusional sangat efektif dalam mendeteksi pola lokal, seperti tepi, sudut, atau tekstur, yang penting untuk tugas pengenalan gambar. Beberapa lapisan konvolusional yang disusun bersama dapat menangkap fitur yang semakin kompleks, memungkinkan jaringan untuk mempelajari representasi hierarkis dari data (O'Shea and Nash 2015).

0	1	2
2	2	0
0	1	2

Gambar 2. 2 Contoh kernel berukuran 3x3 (Dumoulin and Visin 2018)

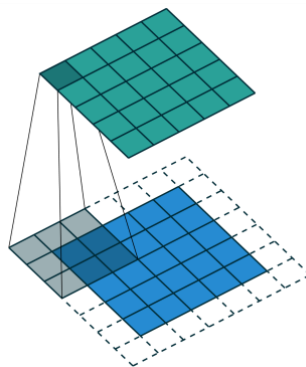


Gambar 2. 3 Proses Konvolusi dengan Ukuran Kernel 3x3 dan Jumlah *Strides* 1 (Dumoulin and Visin 2018)

Gambar 2. 3 merupakan proses konvolusi penuh yang dilakukan dengan kernel (berwarna keabu-abuan seperti Gambar 2. 2) terhadap citra input (berwarna biru) berukuran 5x5 sehingga menghasilkan matriks hasil konvolusi yang disebut dengan *feature map* (berwarna hijau) berukuran 3x3. Perhitungan dilakukan dengan perkalian dot matriks antara matriks kernel dengan matriks citra input pada saat pergeseran kernel. Kernel bergeser dari arah kiri ke kanan menuju bawah terhadap citra input, hal ini dilakukan terus menerus hingga semua bagian pada citra terkonvolusi. Pergeseran yang dilakukan oleh kernel disebut dengan *strides*, pada proses konvolusi pada Gambar 2. 3 memiliki *strides* berjumlah 1 karena

pergeseran dilakukan pada tiap 1 piksel citra. Oleh sebab itu semakin besar jumlah *strides* maka ukuran *feature map* yang dihasilkan semakin kecil.

Pada proses konvolusi ukuran citra akan semakin mengecil sehingga jika dilakukan konvolusi secara berulang maka ukuran citra akan sama dengan ukuran kernel, hal ini dikhawatirkan terdapat banyak informasi piksel yang hilang sehingga pembelajaran model tidak maksimal. Oleh sebab itu, untuk menangani masalah ini adalah melakukan *padding* pada proses konvolusi yang ditunjukkan pada Gambar 2.4. *Padding* melakukan penambahan nilai 0 pada *input* citra (*zero-padding*) sehingga jika dilakukan proses konvolusi, jumlah piksel atau matriks input citra tetap sama setelah proses konvolusi.



Gambar 2. 4 Proses Konvolusi dengan *Zero-Padding* dengan Ukuran Kernel 3x3 dan Jumlah *Strides* 1 (Dumoulin and Visin 2018)

c) Lapisan *Pooling*

Lapisan *pooling* digunakan untuk mengurangi dimensi spasial dari peta fitur, sehingga mengurangi jumlah parameter dan beban komputasi. Operasi *pooling* yang paling umum adalah *max pooling*, yang memilih nilai maksimum dalam jendela yang ditentukan, dan *average pooling*, yang menghitung nilai rata-rata (O'Shea and Nash 2015). *Global pooling* adalah varian yang mengurangi seluruh peta fitur menjadi satu nilai per saluran fitur (Lin, Chen, and Yan 2014). Gambar 2. 5 merupakan contoh proses *max-pooling* dengan menggunakan kernel 3x3 dan mempunyai *strides* 1, dimana ukuran citra akan mengecil 2 kali lipat.

Dengan mengurangi dimensi spasial, lapisan *pooling* membantu jaringan menjadi invarian terhadap translasi dan distorsi kecil dalam data input, yang sangat berguna dalam tugas pengenalan gambar di mana posisi fitur yang tepat kurang penting dibandingkan dengan kehadirannya (O'Shea and Nash 2015).

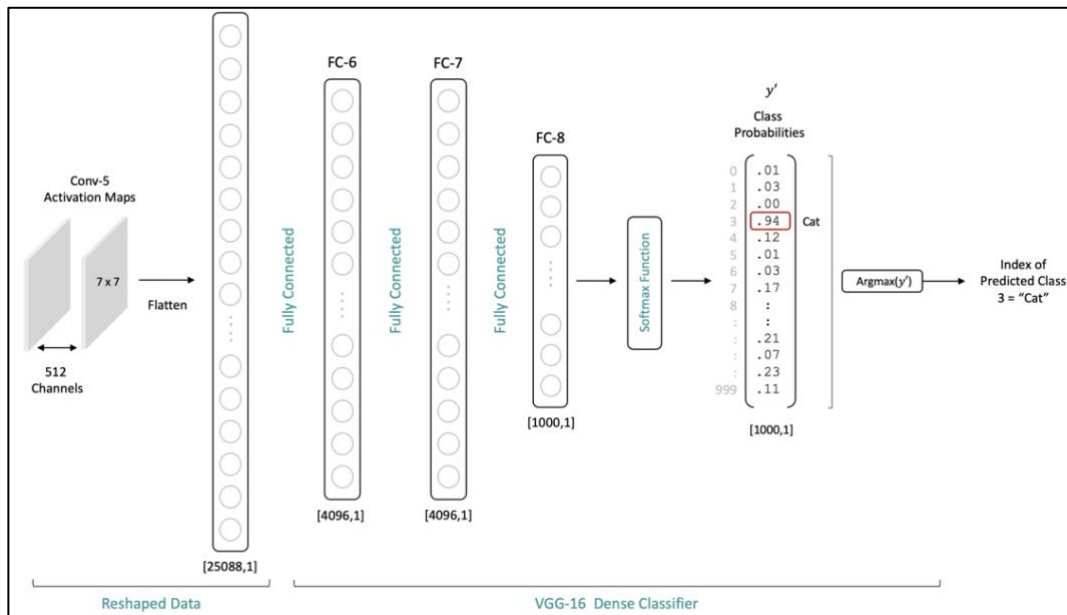


Gambar 2. 5 Proses *Max-Pooling* dengan Ukuran Kernel 3x3 dan Jumlah *Strides* 2

d) Lapisan *Fully-connected* dan Output

Setelah lapisan konvolusional dan *pooling* mengekstraksi fitur, data biasanya dilewatkan melalui satu atau lebih lapisan *fully connected* untuk melakukan klasifikasi atau regresi. Lapisan-lapisan ini bertindak sebagai bagian “pengambil keputusan” dari jaringan, di mana fitur-fitur yang diekstraksi digabungkan untuk menghasilkan output akhir (Zhao et al. 2024). Dalam tugas klasifikasi, lapisan output sering menggunakan fungsi aktivasi *softmax* (Zhao et al. 2024), yang mengubah skor mentah menjadi probabilitas di antara berbagai kelas dan *sigmoid*, yang mengubah nilai menjadi skala 0-1 yang digunakan pada *binary classification* (Zhao et al. 2024). Lapisan ini diilustrasikan pada Gambar 2.6.

Transisi dari lapisan konvolusional ke lapisan *fully connected* menandai pergeseran dari ekstraksi fitur spasial ke pengambilan keputusan tingkat tinggi. Kombinasi ini memungkinkan CNNs menjadi alat yang kuat dalam klasifikasi gambar, deteksi objek, dan tugas lainnya yang memerlukan pemahaman data visual.



Gambar 2. 6 Ilustrasi *Fully Connected Layer* pada CNN (Understanding Convolutional Neural Network: A Complete Guide 2023)

e) CNNs dalam Analisis Citra Medis

CNNs telah merevolusi analisis citra medis dengan menyediakan alat yang dapat secara otomatis mendeteksi dan mengklasifikasikan abnormalitas pada citra medis, seperti X-ray dada (CXR), MRI, dan CT scan. Sifat hierarkis dari CNNs menjadikannya sangat cocok untuk tugas-tugas pencitraan medis, di mana perbedaan halus dalam pola citra dapat menunjukkan adanya penyakit (Kshatri and Singh 2023).

Dalam deteksi abnormalitas toraks, misalnya, CNNs dapat dilatih untuk mengenali fitur-fitur yang terkait dengan berbagai abnormalitas, seperti *nodule*, *pleural effusion*, dan *cardiomegaly*. Studi kasus menunjukkan bahwa CNNs dapat mencapai tingkat akurasi yang sebanding atau bahkan melampaui ahli radiologi dalam mendeteksi penyakit ini (Gakhar and Aggarwal 2022; Rajpurkar et al. 2017). Selain itu, CNNs dapat digunakan untuk memproses sejumlah besar citra dengan cepat, yang penting untuk penerapan dalam situasi klinis di mana keputusan cepat diperlukan.

2.1.4 MobileNet

MobileNet adalah arsitektur ringan yang sangat populer, khususnya di aplikasi mobile dan tertanam. Arsitektur ini dirancang oleh Google untuk memberikan keseimbangan optimal antara akurasi dan efisiensi komputasi (Howard et al. 2017).

Salah satu inovasi utama dalam MobileNet adalah penggunaan konvolusi yang dapat dipisahkan secara mendalam (*depthwise separable convolutions*) yang ditunjukkan pada

Gambar 2.7. (Howard et al. 2017). Konvolusi ini membagi operasi konvolusi standar menjadi dua tahap :

- *Depthwise Convolution*: Di tahap pertama, setiap filter diterapkan secara individual ke setiap saluran input. Hal ini memungkinkan jaringan untuk mengekstraksi fitur-fitur dari setiap saluran secara independen, sehingga mengurangi kompleksitas komputasi (Howard et al. 2017).
- *Pointwise Convolution*: Tahap kedua adalah konvolusi 1x1 yang menggabungkan keluaran dari *depthwise convolution*. Konvolusi 1x1 ini bertindak untuk menyusun kembali informasi yang telah dipisahkan pada tahap sebelumnya, menggabungkan fitur dari berbagai saluran untuk menghasilkan representasi yang lebih kaya (Howard et al. 2017).

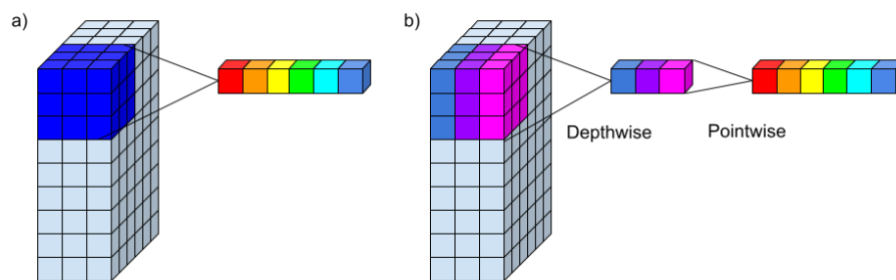
Pendekatan ini secara drastis mengurangi jumlah parameter dan operasi yang diperlukan, sehingga memungkinkan penerapan pada perangkat dengan kemampuan komputasi yang terbatas. Meskipun lebih efisien, *depthwise separable convolutions* tetap mempertahankan akurasi yang baik untuk banyak tugas.

Dalam studi ini, penulis menggunakan arsitektur MobileNetV2 seperti yang dikembangkan oleh Sandler et al. (2018). MobileNetV2 merupakan versi yang ditingkatkan dari MobileNet sebelumnya, dengan berbagai penyempurnaan arsitektural yang secara signifikan meningkatkan akurasi dan efisiensi model (Sandler et al. 2018). Beberapa peningkatan utama dari MobileNetV2 meliputi:

- *Inverted Residuals*: Alih-alih menggunakan koneksi residual standar, MobileNetV2 menerapkan struktur koneksi shortcut pada ruang berdimensi rendah. Pendekatan ini memungkinkan penggunaan ulang fitur yang lebih efektif dengan biaya komputasi yang rendah (Sandler et al. 2018).
- *Linear Bottlenecks*: Penggunaan aktivasi linear pada lapisan bottleneck membantu menjaga informasi penting yang seringkali hilang akibat transformasi non-linear, sehingga meningkatkan performa keseluruhan model (Sandler et al. 2018).
- *Efisiensi yang Ditingkatkan*: Dibandingkan dengan versi sebelumnya, MobileNetV2 menawarkan kompleksitas komputasi yang lebih rendah namun tetap mempertahankan akurasi tinggi. Hal ini menjadikannya sangat cocok untuk aplikasi waktu nyata dan perangkat dengan sumber daya terbatas (Sandler et al. 2018).

Sejumlah penelitian telah membuktikan efektivitas MobileNetV2 dalam bidang pencitraan medis, yang semakin memperkuat alasan pemilihan model ini. Misalnya, Velu (2023) mengembangkan model berbasis MobileNetV2 untuk mendeteksi COVID-19 dari citra rontgen dada (CXR), dan berhasil mencapai akurasi validasi sebesar 92% dan 93% dengan presisi 88%, menunjukkan keandalannya dalam mendiagnosis gangguan paru-paru (Velu 2023). Kolonne et al. (2021) juga mengadaptasi MobileNetV2 untuk klasifikasi CXR menjadi kategori normal, pneumonia, dan pneumonia akibat COVID-19, dan melaporkan akurasi sebesar 98,65% serta recall 98,15% melalui validasi silang 5-fold (Kolonne et al. 2021). Sementara itu, Akter et al. (2021) menggunakan MobileNetV2 dalam deteksi COVID-19 dan mencapai akurasi klasifikasi yang sangat tinggi, yakni 98% (Akter et al. 2021).

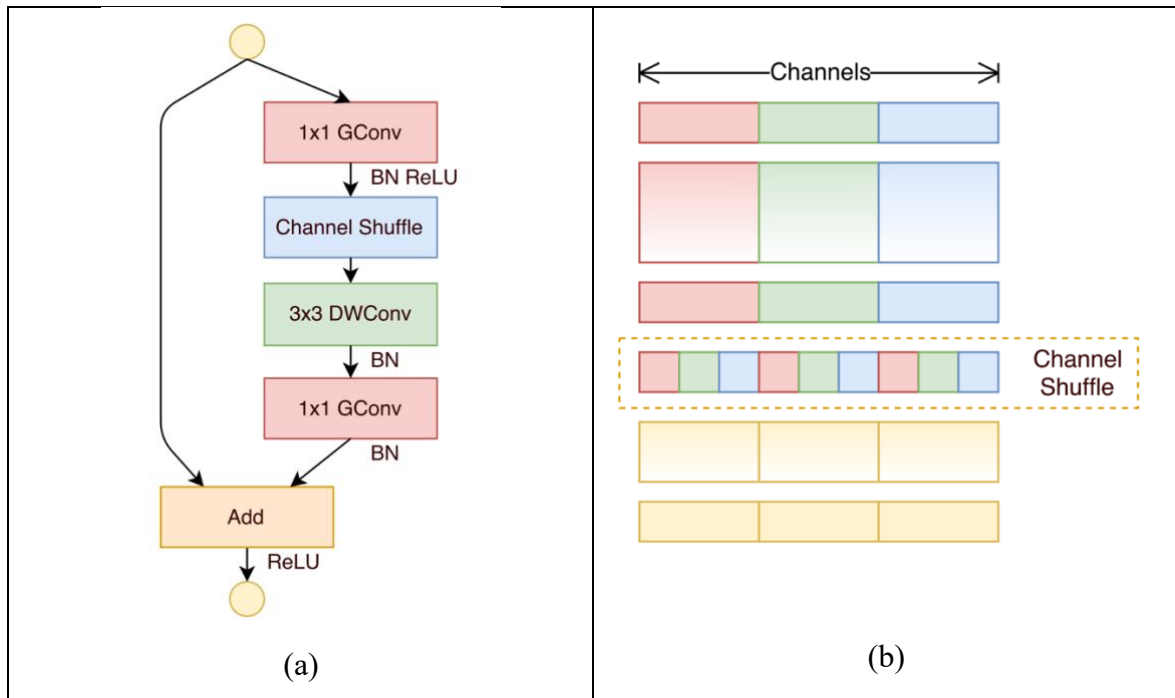
Berdasarkan efisiensi komputasi, tingkat akurasi yang tinggi, serta keberhasilannya dalam berbagai aplikasi analisis citra medis, MobileNetV2 dipandang sebagai pilihan yang paling tepat untuk penelitian ini, karena mampu melakukan ekstraksi fitur yang kuat tanpa membebani sistem secara berlebihan.



Gambar 2. 7 a) Teknik Konvolusi Konvensional b) Teknik Depthwise Separable Convolution (menjadi dasar arsitektur MobileNet) (Päpper 2021)

2.1.5 ShuffleNet

ShuffleNet adalah arsitektur lain yang dirancang untuk mencapai efisiensi komputasi maksimum sambil tetap mempertahankan kinerja yang kuat. ShuffleNet menggunakan pendekatan unik dalam desain arsitekturnya yang mencakup konvolusi grup pointwise dan teknik shuffling channel (Zhang et al. 2017b). Gambar 2. 8 merupakan penerapan konvolusi grup dan detail mengenai *channel shuffle*.



Gambar 2. 8 Arsitektur ShuffleNet a) Unit ShuffleNet dengan Konvolusi Grup Pointwise (GConv) dan *Channel Shuffle* b) Detail dari *Channel Shuffle* (Zhang et al. 2017b)

- Konvolusi Grup *Pointwise*

Konvolusi grup pointwise adalah inovasi utama yang digunakan dalam ShuffleNet. Teknik ini membagi saluran input menjadi beberapa grup dan menerapkan konvolusi 1x1 secara terpisah pada setiap grup (Zhang et al. 2017b). Dengan cara ini, jumlah parameter dan biaya komputasi dapat dikurangi secara signifikan. Namun, pendekatan ini dapat menyebabkan keterbatasan dalam berbagi informasi antar saluran dalam grup yang berbeda.

- Teknik *Shuffling Channel*

Untuk mengatasi keterbatasan konvolusi grup, ShuffleNet memperkenalkan teknik shuffling channel (Zhang et al. 2017b). Setelah konvolusi grup dilakukan, saluran output dari setiap grup dicampur (*shuffled*) sebelum diteruskan ke lapisan berikutnya. Teknik ini memastikan bahwa informasi dapat mengalir di antara grup-grup yang berbeda, memungkinkan model untuk mempelajari representasi yang lebih kaya dan kompleks. Kombinasi konvolusi grup pointwise dan shuffling channel ini membuat ShuffleNet mampu mencapai efisiensi yang tinggi tanpa mengorbankan akurasi secara signifikan.

- Aplikasi dan Keunggulan ShuffleNet

ShuffleNet dirancang untuk menjadi sangat efisien dalam hal komputasi, menjadikannya sangat cocok untuk aplikasi *real-time* di perangkat *mobile* atau lingkungan dengan sumber daya terbatas. Dalam studinya (Zhang et al. 2017b), ShuffleNet telah terbukti mampu mencapai akurasi yang sebanding dengan arsitektur ringan lainnya seperti

MobileNet, tetapi dengan kompleksitas komputasi yang lebih rendah. Ini membuat ShuffleNet menjadi pilihan yang kuat untuk aplikasi seperti pengenalan gambar, analisis citra medis, dan pemrosesan citra *real-time*, di mana waktu respons cepat dan efisiensi daya sangat penting.

Dalam penelitian ini, ShuffleNetV2 akan digunakan sebagai pengganti dari versi awal ShuffleNetV1. Meskipun ShuffleNetV1 telah memperkenalkan operasi efisien seperti *pointwise group convolutions* dan *channel shuffling*, ShuffleNetV2 dikembangkan untuk mengoptimalkan kinerja komputasi secara lebih lanjut serta mengatasi berbagai keterbatasan pada pendahulunya (Ma et al. 2018).

ShuffleNetV2 menghadirkan empat peningkatan utama dibandingkan ShuffleNetV1, yaitu:

- Desain Lebar Kanal yang Seimbang (*Equal Channel Width Design*): Pendekatan ini memastikan biaya komputasi yang merata di setiap lapisan jaringan, sehingga meningkatkan efisiensi keseluruhan (Ma et al. 2018).
- Penghapusan *Group Convolution* pada *Pointwise Layer*: Dengan menghilangkan *group convolutions*, ShuffleNetV2 mampu mengurangi *memory access cost* (MAC), yang seringkali menjadi sumber inefisiensi komputasi pada beberapa arsitektur perangkat keras (Ma et al. 2018).
- Penyederhanaan Jaringan dengan Lebih Sedikit Operasi: Hal ini membuat implementasi model menjadi lebih sederhana dan ramah terhadap perangkat keras (Ma et al. 2018).
- Aliran Informasi yang Lebih Baik: Dengan menghindari fragmentasi berlebihan pada *feature map*, ShuffleNetV2 meningkatkan kemampuan model dalam mempelajari representasi fitur yang lebih kaya dan bermakna (Ma et al. 2018).

Pemilihan ShuffleNetV2 dalam studi ini didasarkan pada efisiensinya yang tinggi, kemampuannya beradaptasi dengan sistem tertanam (*embedded systems*), serta performanya yang kuat dalam pencitraan medis. Meskipun tetap menggunakan prinsip *pointwise convolution* dan *channel shuffling*, ShuffleNetV2 mampu meminimalkan beban komputasi tanpa mengorbankan akurasi (An et al. 2022). Desainnya yang ringan menjadikannya sangat cocok untuk aplikasi waktu nyata, terutama pada perangkat berdaya rendah seperti NVIDIA Jetson Xavier. Penelitian oleh Gu dan Lee (2024) menunjukkan bahwa penggunaan *transfer learning* dapat meningkatkan akurasi deteksi pneumonia dari 78,2% menjadi 91,2%, bahkan melampaui kinerja MobileNetV2 dan ResNet18 (Gu and Lee 2024).

Meskipun ShuffleNetV2 memiliki konvergensi yang sedikit lebih lambat dan nilai recall yang sedikit lebih rendah, keseimbangan antara kecepatan, akurasi, dan efisiensi menjadikannya pilihan optimal untuk sistem diagnosis medis berbasis perangkat mobile dan embedded.

2.1.6 EfficientNet

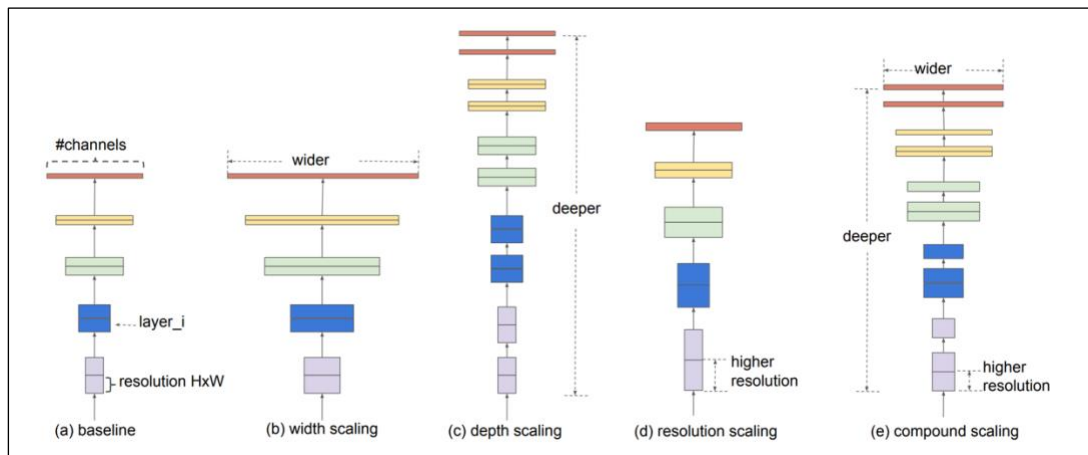
EfficientNet adalah arsitektur ringan yang telah dikenal luas karena kemampuannya untuk mencapai keseimbangan optimal antara kinerja dan efisiensi. EfficientNet memperkenalkan pendekatan baru dalam penskalaan model yang disebut compound scaling (Tan and Le 2020).

EfficientNet menggunakan metode compound scaling (dilihat pada Gambar 2. 9) yang sistematis untuk memperbesar model secara proporsional di tiga dimensi: kedalaman jaringan (*depth*), lebar jaringan (*width*), dan resolusi gambar input (*resolution*) (Tan and Le 2020). Berbeda dengan pendekatan scaling tradisional yang hanya memperbesar satu dimensi pada satu waktu, compound scaling mengatur semua dimensi ini secara serentak menggunakan koefisien scaling tetap. Hal ini memungkinkan model untuk meningkatkan kinerja secara efisien tanpa menyebabkan overfitting atau underfitting.

EfficientNet mengurangi kompleksitas proses pengaturan model yang biasanya dilakukan secara manual, dengan mengoptimalkan skala yang tepat untuk berbagai tugas (Tan and Le 2020). Dengan cara ini, EfficientNet menghasilkan arsitektur yang lebih seimbang dan efisien dibandingkan dengan model yang hanya mengatur satu dimensi pada satu waktu.

EfficientNet dirancang untuk mencapai kinerja tinggi dengan menggunakan lebih sedikit parameter dan daya komputasi (Tan and Le 2020). Pendekatan ini menghasilkan model yang tidak hanya lebih kecil, tetapi juga lebih cepat saat inferensi. Model EfficientNet, seperti varian B0 hingga B7, menawarkan berbagai opsi bagi pengguna tergantung pada kebutuhan sumber daya dan tingkat akurasi yang diinginkan. Varian yang lebih kecil, seperti B0, cocok untuk perangkat dengan keterbatasan daya, sementara varian yang lebih besar, seperti B7, menawarkan akurasi yang lebih tinggi untuk aplikasi yang lebih menuntut.

EfficientNet juga menunjukkan efisiensi energi yang tinggi, yang membuatnya ideal untuk aplikasi di perangkat *mobile*, di mana penggunaan baterai merupakan faktor penting. Dengan demikian, EfficientNet sangat cocok untuk digunakan dalam analisis citra medis pada perangkat yang mungkin memiliki daya komputasi terbatas tetapi membutuhkan hasil yang akurat.



Gambar 2. 9 Skala Model. (a) adalah contoh jaringan dasar; (b)-(d) adalah metode skala konvensional yang hanya meningkatkan satu dimensi dari lebar jaringan, kedalaman, atau resolusi. (e) adalah metode skala majemuk dari EfficientNet, yang secara seragam memperbesar ketiga dimensi tersebut dengan rasio tetap. (Tan and Le 2020)

2.1.7 Grad-CAM

Grad-CAM (*Gradient-weighted Class Activation Mapping*) adalah teknik penting yang meningkatkan interpretabilitas model deep learning, terutama dalam bidang yang kompleks dan berisiko tinggi seperti pencitraan medis (Suara et al. 2024). Dalam aplikasi klinis, di mana keputusan yang diambil oleh model AI dapat berdampak signifikan pada hasil pasien, kemampuan untuk memahami dan mempercayai keputusan tersebut menjadi sangat penting. Grad-CAM membantu mengatasi sifat *black-box* dari model *deep learning* dengan memberikan penjelasan visual atas keputusan model, sehingga membuat alat berbasis AI lebih transparan dan meningkatkan kepercayaan di kalangan tenaga medis seperti radiolog dan dokter.

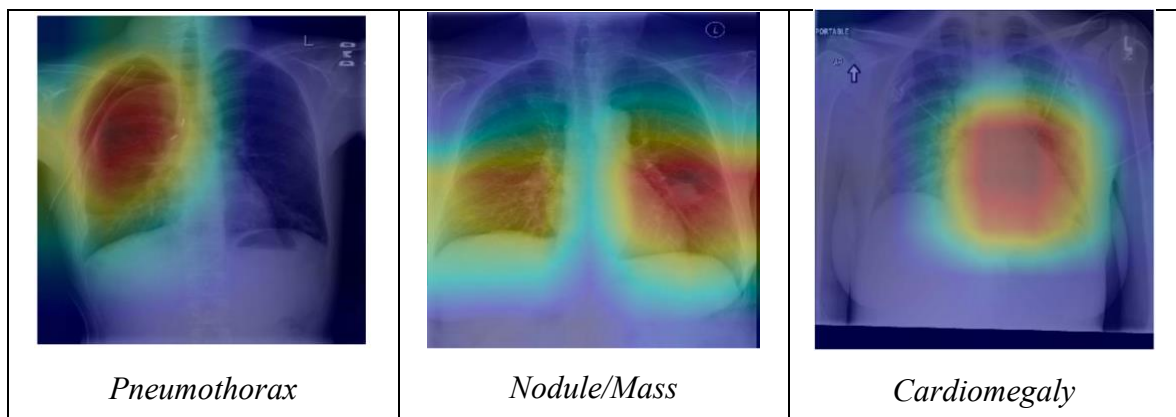
Pada dasarnya, Grad-CAM bekerja dengan menghasilkan *heatmap* yang menyoroti area pada gambar input yang paling relevan dengan prediksi model. Berbeda dengan *Class Activation Mapping* (CAM) tradisional yang membutuhkan modifikasi pada arsitektur model, Grad-CAM memanfaatkan gradien yang mengalir ke lapisan konvolusi terakhir dari jaringan saraf. Gradien ini digunakan untuk menghasilkan peta lokal yang menunjukkan bagian gambar yang penting dalam memprediksi kelas tertentu (Selvaraju et al. 2020). Ini sangat berguna dalam CNN yang digunakan untuk tugas pencitraan medis, karena memberikan umpan balik spasial yang bermakna mengenai alasan model memilih fitur tertentu untuk mengklasifikasikan abnormalitas.

Grad-CAM beroperasi dengan memberikan bobot pada peta aktivasi dari lapisan konvolusi terakhir berdasarkan gradien kelas target. Visualisasi yang dihasilkan

menunjukkan area gambar yang secara signifikan berkontribusi pada prediksi (Selvaraju et al. 2020). Dengan cara ini, *heatmap* yang dihasilkan dapat digabungkan dengan gambar asli, sehingga tenaga medis dapat melihat dengan jelas area spesifik pada citra seperti rontgen dada yang menyebabkan model mendeteksi kelainan seperti *cardiomegaly*, *pneumothorax*, atau *pleural effusion*. Visualisasi yang terlokalisasi ini sangat penting untuk membangun kepercayaan pada sistem otomatis, karena memberikan penjelasan yang dapat dipahami untuk keputusan model, bukan hanya sekadar output tanpa alasan yang jelas.

Dalam konteks pencitraan medis, interpretabilitas model menjadi faktor yang sangat signifikan. Para profesional medis sering kali ragu menggunakan alat AI karena sifatnya yang tidak transparan, di mana proses pengambilan keputusan model tidak jelas (Gerdes 2024). Ketidakjelasan ini dapat menyebabkan keraguan dalam mengadopsi alat AI, terutama di lingkungan yang berisiko tinggi seperti radiologi, di mana kesalahan diagnosis dapat berakibat serius. Grad-CAM membantu menjembatani kesenjangan ini dengan membuat keputusan AI lebih dapat dipahami dan diinterpretasikan, sehingga memberikan kepercayaan lebih kepada tenaga medis untuk mengintegrasikan AI dalam alur kerja mereka.

Sebagai contoh, dalam mendiagnosis kelainan toraks seperti yang terlihat pada rontgen dada, tidak cukup jika model AI hanya menunjukkan adanya kelainan. Tenaga medis perlu memahami area mana dari gambar yang menjadi dasar diagnosis tersebut dan bagaimana AI sampai pada kesimpulannya. Grad-CAM memfasilitasi pemahaman ini dengan memberikan umpan balik visual (yang dapat dilihat pada Gambar 2. 10), menunjukkan area spesifik paru-paru, jantung, atau mediastinum yang menjadi fokus model saat mendiagnosis kondisi seperti kardiomegali, atelektasis, atau pneumotoraks. Transparansi ini tidak hanya membantu memvalidasi keputusan model, tetapi juga memungkinkan tenaga medis untuk memeriksa ulang area yang difokuskan oleh model berdasarkan keahlian mereka, memastikan bahwa AI mengevaluasi fitur anatomi yang benar.



Gambar 2. 10 Contoh aplikasi Grad-CAM dalam mendeteksi 3 abnormalitas berbeda
(Rajpurkar et al. 2017)

Dalam penelitian ini yang berfokus pada pengembangan model deep learning ringan seperti MobileNet, EfficientNet, dan ShuffleNet untuk mendeteksi kelainan toraks pada rontgen dada, Grad-CAM menjadi alat penting untuk meningkatkan interpretabilitas model. Karena penelitian ini bertujuan untuk menerapkan model-model ini di lingkungan dengan sumber daya terbatas, di mana ketersediaan ahli radiolog mungkin terbatas, Grad-CAM dapat berfungsi sebagai alat bantu bagi tenaga kesehatan yang tidak memiliki pelatihan khusus dalam radiologi.

Grad-CAM tidak hanya membantu dalam validasi model, tetapi juga memiliki implikasi praktis bagi pengguna akhir—tenaga medis di lingkungan dengan sumber daya terbatas. Misalnya, di daerah pedesaan di Indonesia, di mana fokus penelitian saya, radiolog tidak selalu hadir untuk meninjau setiap rontgen. Alat AI yang dilengkapi dengan Grad-CAM dapat membantu dokter umum atau tenaga kesehatan dengan tidak hanya mengidentifikasi potensi kelainan, tetapi juga menunjukkan dasar visual dari temuan tersebut. Interpretabilitas ini dapat membantu pengguna yang tidak spesialis membuat keputusan yang lebih tepat dan meningkatkan kepercayaan mereka pada alat AI. Selain itu, kemampuan Grad-CAM untuk menyoroti area relevan dalam gambar juga penting untuk keperluan pelatihan, membantu tenaga kesehatan mengenali fitur radiologis yang menunjukkan kelainan toraks seperti efusi pleura atau nodul/massa.

2.1.8 K-Fold Cross Validation

K-Fold Cross Validation adalah metode statistik yang digunakan untuk mengevaluasi dan membandingkan algoritma pembelajaran mesin. Metode ini melibatkan pembagian data menjadi k segmen, di mana satu segmen digunakan untuk validasi dan sisanya digunakan untuk pelatihan dalam setiap iterasi (Refaeilzadeh et al. 2009).

K-Fold Cross Validation dapat memberikan estimasi kinerja yang akurat dan dapat diandalkan, memanfaatkan semua data yang tersedia, serta mencegah overfitting atau pengestimasi varian data (Refaeilzadeh et al. 2009).

K-Fold Cross Validation dapat diterapkan untuk estimasi kinerja, pemilihan model, dan penyesuaian parameter model pembelajaran. Metode ini secara luas digunakan dalam data mining dan pembelajaran mesin sebagai prosedur standar (Refaeilzadeh et al. 2009).

Penerapan K-Fold Cross Validation memiliki dampak signifikan dalam memastikan keakuratan dan kehandalan penilaian kinerja algoritma pembelajaran mesin. Oleh karena itu, penelitian ini memberikan landasan yang kuat untuk memastikan bahwa hasil eksperimen

yang dihasilkan dapat diandalkan dan dapat diterapkan secara luas dalam pengembangan model pembelajaran mesin (Refaeilzadeh et al. 2009).

Gambar 2. 11 merupakan contoh dari pembagian dataset menjadi 5 bagian, hal ini dikarenakan k dalam cross validation adalah 5 sehingga pembagian setiap split atau bagian akan mempunyai perbandingan data latih dan validasi masing-masing adalah 80% dan 20%.

Split Ke- 1	Fold #1	Fold #2	Fold #3	Fold #4	Fold #5
Split Ke- 2	Fold #1	Fold #2	Fold #3	Fold #4	Fold #5
Split Ke- 3	Fold #1	Fold #2	Fold #3	Fold #4	Fold #5
Split Ke- 4	Fold #1	Fold #2	Fold #3	Fold #4	Fold #5
Split Ke- 5	Fold #1	Fold #2	Fold #3	Fold #4	Fold #5

Gambar 2. 11 *K-Fold Cross Validation* dengan nilai $k=5$

2.1.9 Metrik Evaluasi

Evaluasi model deep learning dalam pencitraan medis sangat penting untuk memastikan keandalannya dalam diagnosis. Beberapa metrik evaluasi yang umum digunakan meliputi akurasi, presisi, recall (sensitivitas), F1-Score, dan kurva ROC, yang masing-masing memiliki peran penting dalam menilai kinerja model.

- Akurasi

Akurasi mengukur persentase prediksi yang benar dari total prediksi, namun bisa menyesatkan pada dataset yang tidak seimbang (Juba and Le 2019). Misalnya, pada kasus kelainan toraks yang langka, model yang hanya memprediksi hasil negatif bisa mencapai akurasi tinggi meskipun gagal mendeteksi kelainan yang sebenarnya ada. Oleh karena itu, akurasi harus dipertimbangkan bersama metrik lain untuk menghindari kesalahan dalam diagnosis.

- Presisi

Presisi mengukur seberapa banyak prediksi positif yang benar-benar benar (Terven et al. 2024). Ini sangat penting ketika kesalahan positif (*false positive*) dapat menimbulkan biaya atau risiko tambahan, seperti tes lanjutan atau tindakan medis yang tidak perlu. Dalam lingkungan dengan sumber daya terbatas, seperti sistem kesehatan Indonesia, presisi yang tinggi membantu mengurangi diagnosis yang salah dan pengobatan yang tidak perlu.

- Recall (Sensitivitas)

Recall menunjukkan kemampuan model untuk mendeteksi semua kasus positif yang sebenarnya (Terven et al. 2024). Di bidang medis, recall sangat penting karena mengurangi kesalahan negatif (false negative) yang berisiko menunda diagnosis dan pengobatan. Model dengan recall tinggi memastikan tidak ada penyakit yang terlewatkan, yang krusial dalam kondisi darurat seperti pneumotoraks atau kardiomegali.

- **F1-Score**

F1-score adalah ukuran yang digunakan untuk menilai seberapa baik model bisa mendeteksi sesuatu dengan benar, terutama saat data tidak seimbang. F1-score menggabungkan dua hal penting: presisi (seberapa banyak hasil positif yang benar) dan recall (seberapa banyak kasus positif yang berhasil ditemukan) (Christen, Hand, and Kirielle 2023). Jika hanya salah satu yang tinggi, hasilnya belum tentu baik. F1-score membantu mencari keseimbangan antara keduanya. Dalam dunia medis, ini penting karena kita tidak ingin model melewatkan penyakit yang ada (kesalahan negatif), tapi juga tidak ingin salah memberi tahu orang bahwa mereka sakit padahal tidak (kesalahan positif).

- **Kurva AUC-ROC**

Kurva ROC menggambarkan trade-off antara true positive rate (sensitivitas) dan false positive rate pada berbagai threshold. AUC-ROC memberikan ringkasan kinerja model, di mana nilai mendekati 1 menunjukkan kemampuan diagnostik yang baik. AUC-ROC yang tinggi sangat penting dalam memastikan model dapat menyeimbangkan sensitivitas dan kesalahan positif, terutama dalam tugas diagnostik yang kompleks (Naidu et al. 2023).

Untuk melengkapi pemahaman mengenai metrik evaluasi yang telah dijelaskan sebelumnya, berikut ini disajikan rumus-rumus matematis yang digunakan untuk menghitung masing-masing metrik. Metrik akurasi, precision, recall, F1 score, dan TPR dan FPR pada ROC masing-masing ditunjukkan pada persamaan 1, 2, 3, 4, dan 5.

Untuk semua metrik di bawah ini digunakan notasi:

- **True Positive (TP):** Model memprediksi ada kelainan, dan memang benar pasien memiliki kelainan. Contoh: Model mendeteksi *cardiomegaly*, dan hasil radiolog juga menyatakan ada *cardiomegaly*.
- **True Negative (TN):** Model memprediksi tidak ada kelainan, dan benar pasien sehat. Contoh: Model mengatakan tidak ada kelainan toraks, dan hasil pemeriksaan mendukung itu.

- False Positive (FP): Model memprediksi ada kelainan, padahal sebenarnya tidak ada. Contoh: Model menyatakan ada *pneumotorax*, tapi pasien sebenarnya sehat. Ini bisa menyebabkan tes lanjutan yang tidak perlu.
- False Negative (FN): Model memprediksi tidak ada kelainan, padahal pasien sebenarnya sakit. Contoh: Model melewati mendeteksi *nodule/mass* yang nyata, sehingga diagnosis tertunda.

Dengan notasi tersebut, rumus metrik evaluasi sebagai berikut:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

$$\text{TPR} = \frac{TP}{TP+FN} \quad \& \quad \text{FPR} = \frac{FP}{FP+TN} \quad (5)$$

2.2 Kajian Pustaka

Dalam beberapa tahun terakhir, *deep learning* telah merevolusi bidang pencitraan medis, terutama dalam diagnostik dan pengambilan keputusan klinis. *Convolutional Neural Networks* (CNNs) telah memainkan peran penting, khususnya dalam analisis gambar medis seperti rontgen dada (CXR) (Sarvamangala and Kulkarni 2022). (Krizhevsky, Sutskever, and Hinton 2012) memperkenalkan CNN untuk tugas *computer vision*, yang menjadi dasar penerapan teknologi ini di bidang kesehatan. Sejak itu, CNN telah digunakan secara luas dalam tugas klasifikasi dan segmentasi gambar, memungkinkan ekstraksi fitur otomatis dan akurat dari gambar medis. Misalnya, dalam deteksi abnormalitas toraks, CNN telah menunjukkan kinerja luar biasa dalam mendeteksi kondisi seperti pneumonia, tuberkulosis (TBC), dan *pneumotorax*, memberikan diagnostik yang lebih cepat dan lebih andal dibandingkan interpretasi manual oleh radiolog (Rajaraman, Kim, and Antani 2020; Rajpurkar et al. 2017).

Keberhasilan *deep learning* dalam bidang kesehatan disebabkan oleh kemampuannya untuk melampaui metode tradisional seperti ekstraksi fitur manual dan sistem berbasis aturan, yang rentan terhadap kesalahan manusia. Seperti yang dicatat oleh (Litjens et al. 2017), CNN telah mengubah diagnostik radiologi, terutama dalam deteksi penyakit toraks, dengan secara otomatis mempelajari fitur hierarkis dari data piksel mentah. Otomatisasi ini sangat penting di wilayah seperti Indonesia, di mana terdapat kekurangan

ahli radiologi dan prevalensi penyakit seperti TBC dan pneumonia yang tinggi (Dayne 2021; Wulan 2024). Riset dari Google dalam deteksi abnormalitas rontgen dada menggunakan AI juga menyoroti bagaimana *deep learning* dapat digunakan dalam penyaringan skala besar, yang mampu membantu radiolog secara *real-time* dan mengurangi kesalahan diagnostik (Nabulsi 2021).

Namun, model deep learning tradisional, seperti ResNet dan DenseNet, menghadapi tantangan di lingkungan dengan sumber daya terbatas karena kebutuhan komputasi dan memori yang tinggi (Zhang et al. 2020). Meskipun akurat, model ini tidak praktis untuk diterapkan di wilayah dengan infrastruktur komputasi yang terbatas. (Liu et al. 2024) menunjukkan bahwa karakteristik intensif sumber daya dari model-model ini membatasi penggunaannya di daerah berkembang. Di tempat-tempat seperti Indonesia, di mana banyak rumah sakit tidak memiliki perangkat keras canggih (Sarkar 2021), diperlukan model yang lebih efisien. (Bursa, Durmaz İncel, and Işıklar Alptekin 2023) juga menyoroti pentingnya menggunakan model yang lebih ringan dan teknik kuantisasi untuk mengatasi keterbatasan perangkat keras.

Model-model mobile seperti MobileNet, ShuffleNet, dan EfficientNet menawarkan solusi efisien untuk tantangan dalam pencitraan medis dengan mengoptimalkan kecepatan komputasi tanpa mengorbankan akurasi (Howard et al. 2017; Tan and Le 2020; Zhang et al. 2017b). Model-model ini sangat efisien sehingga memungkinkan implementasi pada perangkat bergerak dan sistem dengan daya rendah, menjadikannya ideal untuk fasilitas kesehatan dengan keterbatasan teknologi. Kemampuan aksesibilitas ini memberikan manfaat besar bagi wilayah dengan infrastruktur terbatas, karena perangkat seperti ponsel pintar atau tablet dapat digunakan untuk melakukan diagnosis secara cepat dan akurat di daerah pedesaan atau negara berkembang (Blantz 2010).

Selain itu, kecepatan pemrosesan model-model ini membantu tenaga medis, seperti ahli radiologi, dalam menghasilkan laporan diagnostik dengan cepat, yang sangat penting saat menghadapi lonjakan pasien atau kondisi darurat kesehatan (van Leeuwen et al. 2022). Dari segi ekonomi, model-model mobile ini mengurangi kebutuhan akan perangkat keras mahal, sehingga rumah sakit dapat mengadopsi teknologi kecerdasan buatan tanpa harus berinvestasi dalam infrastruktur yang kompleks (Howard et al. 2017). Skalabilitasnya juga memungkinkan integrasi ke berbagai sistem pelayanan kesehatan, mulai dari klinik hingga platform telemedis, sehingga penggunaannya dapat meluas dan diakses secara jarak jauh (Bhatta 2023).

Secara teknis, MobileNet menggunakan teknik *depthwise separable convolution* untuk mengurangi beban komputasi, menjadikannya efisien untuk perangkat mobile (Howard et al. 2017). ShuffleNet memanfaatkan *pointwise group convolution* dan *channel shuffling* untuk menjaga akurasi sambil menurunkan biaya komputasi (Zhang et al. 2017b). Sementara itu, EfficientNet mengadopsi pendekatan *compound scaling* multiskala untuk menyeimbangkan dimensi jaringan, sehingga dapat mencapai performa tinggi dengan jumlah parameter yang lebih sedikit (Tan and Le 2020). Secara keseluruhan, model-model ini menyediakan solusi diagnostik yang cepat, dapat diskalakan, dan hemat biaya, sangat cocok untuk diterapkan di lingkungan layanan kesehatan dengan sumber daya terbatas.

Selain efisiensi arsitekturalnya, model-model ini juga menunjukkan kinerja empiris yang kuat dalam tugas klasifikasi X-ray dada. Khususnya, MobileNetV2 telah terbukti sangat akurat dan cepat, sehingga sangat cocok untuk analisis citra medis. Sebagai contoh, MobileNetV2 mencapai akurasi sebesar 98,65% dan recall 98,15% dalam mendeteksi pneumonia dan COVID-19 (Kolonie et al. 2021), serta menempati peringkat tertinggi di antara 11 model Convolutional Neural Networks (CNN) dalam hal akurasi dan kecepatan (Akter et al. 2021). Velu (2023) juga melakukan penyempurnaan (*fine-tuning*) terhadap MobileNetV2 untuk mendeteksi COVID-19 secara akurat dan cepat melalui citra X-ray dada, dengan hasil akurasi pelatihan sebesar 92,5% dan akurasi validasi sebesar 93,75% (Velu 2023). Hasil ini melampaui kinerja CNN yang dilatih dari awal (81,4%) maupun ResNet50 yang disesuaikan (80,6%). Selain itu, tinjauan sistematis oleh (Iqbal et al. 2024) melaporkan bahwa MobileNetV2 mencapai tingkat akurasi sebesar 94% pada dataset CXR-14, menunjukkan konsistensi kinerjanya di berbagai studi. (Gu and Lee 2024) juga membuktikan efektivitas model ini dengan pendekatan transfer learning, yang menghasilkan akurasi 90,9% dalam mendeteksi pneumonia.

ShuffleNetV2, sebagai model ringan dan efisien lainnya, juga menunjukkan hasil yang menjanjikan dalam bidang klasifikasi penyakit. (Gu and Lee 2024) melaporkan bahwa ShuffleNetV2 berhasil mencapai akurasi sebesar 91,2% dalam deteksi pneumonia dengan pendekatan transfer learning, yang menegaskan potensinya untuk klasifikasi penyakit secara cepat dan akurat. Selain itu, (An et al. 2022) menyoroti keunggulan ShuffleNetV2 dari sisi jumlah parameter yang rendah dan bobot model yang ringan, menjadikannya sangat cocok untuk aplikasi tertanam (*embedded systems*). Karakteristik ini sangat penting dalam konteks layanan kesehatan dengan sumber daya terbatas, karena memungkinkan implementasi teknologi diagnostik berbasis AI tanpa memerlukan perangkat keras yang kompleks atau mahal (An et al. 2022).

EfficientNetB0 juga telah menunjukkan performa yang luar biasa dengan mampu menyeimbangkan akurasi tinggi dan efisiensi komputasi. Model ini mencapai akurasi klasifikasi sebesar 99%, melampaui model yang lebih dalam seperti ResNet-50 dan VGG-19 dalam hal efisiensi maupun akurasi (Mzoughi et al. 2023). Lebih lanjut, (Kansal et al. 2024) melaporkan bahwa EfficientNetB0 mengungguli ResNet-50 dengan akurasi pengujian sebesar 99,62% pada dataset Kaggle dan 99,78% pada dataset Mendeley untuk klasifikasi abnormalitas paru *multi-centric* (sumber yang bervariasi). Temuan ini diperkuat oleh (Iqbal et al. 2024), yang mencatat bahwa EfficientNetB0 mencapai akurasi 98% dalam deteksi COVID-19.

Selain itu, (An, Chen, and Shao 2024) menunjukkan bahwa penggabungan EfficientNetB0 dengan DenseNet121 yang diperkaya dengan *attention mechanism* menghasilkan akurasi diagnostik tinggi sebesar 95,19%, dengan nilai presisi sebesar 98,38% dan skor F1 sebesar 96,06%. Hal ini menegaskan bahwa EfficientNetB0 bukan hanya efisien dan ringan, tetapi juga sangat andal untuk aplikasi diagnostik berbasis citra medis dalam berbagai kondisi klinis.

Temuan-temuan dari penelitian sebelumnya mendukung pemilihan MobileNetV2, ShuffleNetV2, dan EfficientNetB0 dalam studi ini serta memperkuat relevansi dan kesesuaiannya untuk aplikasi klinis di dunia nyata, khususnya di wilayah dengan keterbatasan layanan kesehatan. Ketiga model ini menunjukkan keseimbangan yang optimal antara akurasi, efisiensi komputasi, dan kemampuan adaptasi melalui teknik fine-tuning dan transfer learning, menjadikannya kandidat unggulan untuk tugas-tugas diagnostik yang menuntut kecepatan sekaligus ketepatan.

Seluruh literatur penting yang telah dibahas dirangkum dalam Tabel 2. 1, yang menyajikan ringkasan kinerja model-model serta temuan utama dari berbagai studi secara ringkas dan terstruktur.

Penggunaan model *pre-trained* dari domain umum seperti ImageNet terbukti sangat efektif dalam konteks pencitraan medis. (Velu 2023) menunjukkan bahwa arsitektur MobileNetV2 yang dilatih awal menggunakan transfer learning mampu mengenali penyakit paru-paru dengan akurasi tinggi meski menggunakan dataset yang terbatas dan beragam sumber data. Hal serupa ditegaskan oleh (Gu and Lee 2024), yang dalam studi kasus pneumonia menunjukkan bahwa fitur yang dipelajari oleh model dari citra umum (ImageNet) dapat berhasil ditransfer ke domain citra medis melalui teknik real-world feature transfer learning, menghasilkan kinerja yang melampaui model yang dilatih dari awal. Oleh karena itu, dalam studi ini, seluruh model (MobileNetV2, ShuffleNetV2, dan

EfficientNetB0) diimplementasikan dengan pendekatan transfer learning berbasis model yang telah pre-trained, agar dapat memanfaatkan representasi fitur yang telah matang serta mempercepat proses pelatihan tanpa memerlukan dataset besar yang sulit diperoleh dalam domain kesehatan lokal.

Meskipun arsitektur deep learning terus mengalami pembaruan, peningkatan versi model tidak selalu sejalan dengan peningkatan performa. Studi oleh (Breve 2022) menunjukkan bahwa MobileNet generasi awal justru menunjukkan performa lebih baik dibandingkan MobileNetV2 dalam klasifikasi COVID-19 pada dataset COVIDx8B, dengan akurasi masing-masing 97,1% dan 94,85%. Hal serupa juga dilaporkan oleh (Almutairi, Maher, and Bchir 2022), di mana model-model ringan dari EfficientNet belum mampu menyamai performa varian yang telah di-pretrained, dengan akurasi maksimal 96% pada versi lightweight dan 99% pada EfficientNet-lite-B0. Pada kasus yang lain, (Susanto et al. 2022) menemukan bahwa MobileNetV2 memberikan hasil terbaik dalam klasifikasi COVID-19 berbasis CT-scan, dengan akurasi 81%, F1-score 0.81, mengungguli MobileNetV3 Large (maksimum akurasi 80%) dan Small (maksimum akurasi 76%). Dukungan terhadap kecenderungan ini juga muncul dalam studi (Taha et al. 2022), yang membandingkan performa MobileNetV1, V2, dan V3 dalam klasifikasi tumor otak, dan melaporkan bahwa MobileNetV2 mencapai akurasi uji tertinggi sebesar 97,86%, lebih unggul dibandingkan MobileNetV3Small (94,95%) maupun MobileNetV3Large (97,40%). Temuan-temuan ini menunjukkan bahwa dalam sejumlah kasus, MobileNetV2 tetap mampu bersaing dan bahkan menunjukkan keunggulan relatif dibandingkan versi yang lebih baru seperti MobileNetV3, tergantung pada konteks dan karakteristik data yang digunakan.

Evaluasi model deep learning dalam pencitraan medis umumnya dilakukan melalui prosedur pengujian yang ketat untuk memastikan ketangguhan (*robustness*) dan kemampuan generalisasi model. Salah satu teknik yang umum digunakan adalah K-Fold Cross Validation, yaitu metode yang membagi data menjadi beberapa subset atau fold, kemudian melakukan validasi silang untuk setiap bagian guna meminimalkan bias dan variansi (Le et al. 2024; Refaeilzadeh et al. 2009). Selain itu, sejumlah metrik evaluasi seperti akurasi, presisi, recall, dan skor F1 digunakan untuk menilai kinerja model secara menyeluruh. Metrik-metrik ini memberikan gambaran mengenai ketepatan serta konsistensi prediksi model dalam mengklasifikasikan data (Naidu et al. 2023). Lebih lanjut, metrik seperti Area Under the Receiver Operating Characteristic Curve (AUC-ROC) juga sangat penting dalam mengukur akurasi diagnostik model. Kurva AUC-ROC menunjukkan keseimbangan antara *true positive rate* dan *false positive rate*, sehingga memberikan pemahaman mendalam

tentang kemampuan model dalam membedakan antar kelas—dalam konteks ini, antara keberadaan dan ketidakhadiran kelainan toraks (Melo 2013).

Terakhir, interpretabilitas model tetap menjadi perhatian penting. Teknik seperti Grad-CAM meningkatkan transparansi model AI dengan memvisualisasikan area gambar yang mempengaruhi prediksi model. Dalam pencitraan medis, di mana radiolog perlu memahami alasan di balik diagnosis model, alat interpretabilitas seperti Grad-CAM sangat penting. Grad-CAM meningkatkan kepercayaan pada sistem AI dengan memungkinkan tenaga medis untuk memvalidasi diagnostik otomatis, terutama dalam deteksi abnormalitas toraks (Rajpurkar et al. 2017; Selvaraju et al. 2020).

Sebagai kesimpulan, pengembangan dan adopsi model deep learning yang ringan, dikombinasikan dengan teknik seperti Grad-CAM, memiliki implikasi yang signifikan untuk masa depan pencitraan medis, terutama di lingkungan dengan sumber daya terbatas. Model seperti MobileNetV2, ShuffleNetV2, dan EfficientNetB0 menawarkan keseimbangan antara kinerja dan efisiensi, sehingga cocok untuk diterapkan di daerah kurang terlayani. Seiring dengan adopsi AI yang semakin meluas dalam sistem kesehatan global, model-model ini akan memainkan peran penting dalam memperluas akses ke diagnostik berkualitas, terutama di wilayah yang menghadapi kekurangan ahli radiologi dan infrastruktur komputasi yang terbatas.

Tabel 2. 1 Ringkasan dari studi literatur yang signifikan

Fokus	Author	Objektivitas	Relevansi
MobileNet V2, ShuffleNet V2	(An et al. 2022)	Membangun jaringan saraf dalam ringan (E-TBNet) untuk deteksi otomatis tuberkulosis menggunakan pencitraan X-ray DR, yang dioptimalkan untuk perangkat dengan spesifikasi perangkat keras rendah.	MobileNetV2 mencapai akurasi tertinggi (90%) di antara model ringan, sementara ShuffleNetV2 unggul dalam ukuran dan efisiensi, dan E-TBNet memberikan keseimbangan antara akurasi (85%) dan efisiensi.
MobileNet V2	(Velu 2023)	Membangun model MobileNetV2 yang disesuaikan (<i>fine-tuned</i>) untuk deteksi COVID-19 secara	Model MobileNetV2 yang telah disesuaikan (<i>fine-tuned</i>) menunjukkan kinerja terbaik dengan akurasi pelatihan

		akurat dan cepat dari citra X-ray dada.	92,5% dan akurasi validasi 93,75%, melampaui CNN yang dilatih dari awal (81,4%) dan ResNet50 yang disesuaikan (80,6%).
MobileNet V2	(Aker et al. 2021)	Membangun model pembelajaran mendalam untuk mendeteksi COVID-19 dari citra X-ray dada menggunakan arsitektur CNN.	Dari model yang dievaluasi (MobileNetV2, VGG16, ResNet50), MobileNetV2 meraih akurasi tertinggi yaitu 98%.
MobileNet V2	(Kolonne et al. 2021)	Membangun model berbasis MobileNetV2 untuk klasifikasi kondisi normal, pneumonia, dan COVID-19 dari citra X-ray dada.	MobileNetV2 tanpa pembelajaran transfer mencapai akurasi tertinggi (98,65%) dibandingkan pendekatan pembelajaran transfer (97,89%).
MobileNet V2, EfficientNetB0	(Iqbal et al. 2024)	<i>Systematic review</i> metode AI untuk deteksi penyakit paru-paru menggunakan citra X-ray dada.	MobileNetV2 mencapai akurasi 94% pada dataset CXR-14, sementara EfficientNet-B0 meraih akurasi 98% dalam deteksi COVID-19.
Transfer Learning Utilization, MobileNet V2, ShuffleNet V2	(Gu and Lee 2024)	Memanfaatkan pembelajaran transfer mendalam dari dataset umum seperti ImageNet untuk mengklasifikasikan pneumonia pada citra X-ray.	Model ringan seperti ShuffleNetV2 dan MobileNetV2 menunjukkan peningkatan performa signifikan dengan pembelajaran transfer, masing-masing mencapai akurasi 91,2% dan 90,9%.
GradCAM Use	(Gakhar and	Membangun alur dua tahap untuk deteksi abnormalitas toraks dan klasifikasi penyakit	Pendekatan berbasis fusi yang diusulkan (ThoraciNet) mencapai AUC 0,99 untuk

	Aggarwal 2022)	menggunakan kombinasi jaringan saraf konvolusional dalam (fusion DCNNs).	triase CXR dan 0,79 untuk klasifikasi penyakit, mengungguli DCNN alur tunggal. Visualisasi GradCAM digunakan untuk meningkatkan interpretabilitas.
EfficientNetB0	(Kansal et al. 2024)	Membandingkan kinerja EfficientNet-B0 dan ResNet-50 dalam klasifikasi abnormalitas paru-paru secara multi-sentral.	EfficientNet-B0 mengungguli ResNet-50 dengan akurasi pengujian 99,62% pada dataset Kaggle dan 99,78% pada dataset Mendeley.
EfficientNetB0, XAI Implementation	(Sahin et al. 2024)	Menggunakan model CNN dengan Grad-CAM++ untuk mendeteksi pneumonia dari citra X-ray dada.	EfficientNetB0 mencatat akurasi tertinggi (95,03%) dan F-measure (96,12%), sementara VGG-19 dan Inception-V3 juga menunjukkan performa yang kompetitif.
EfficientNetB0	(An et al. 2024)	Membangun model CNN yang menggabungkan EfficientNetB0 dan DenseNet121, ditingkatkan dengan mekanisme atensi untuk deteksi pneumonia.	Akurasi diagnostik tinggi (95,19%) dicapai dengan presisi yang ditingkatkan (98,38%) dan skor F1 sebesar 96,06% berkat integrasi mekanisme atensi.

BAB 3

Metodologi Penelitian

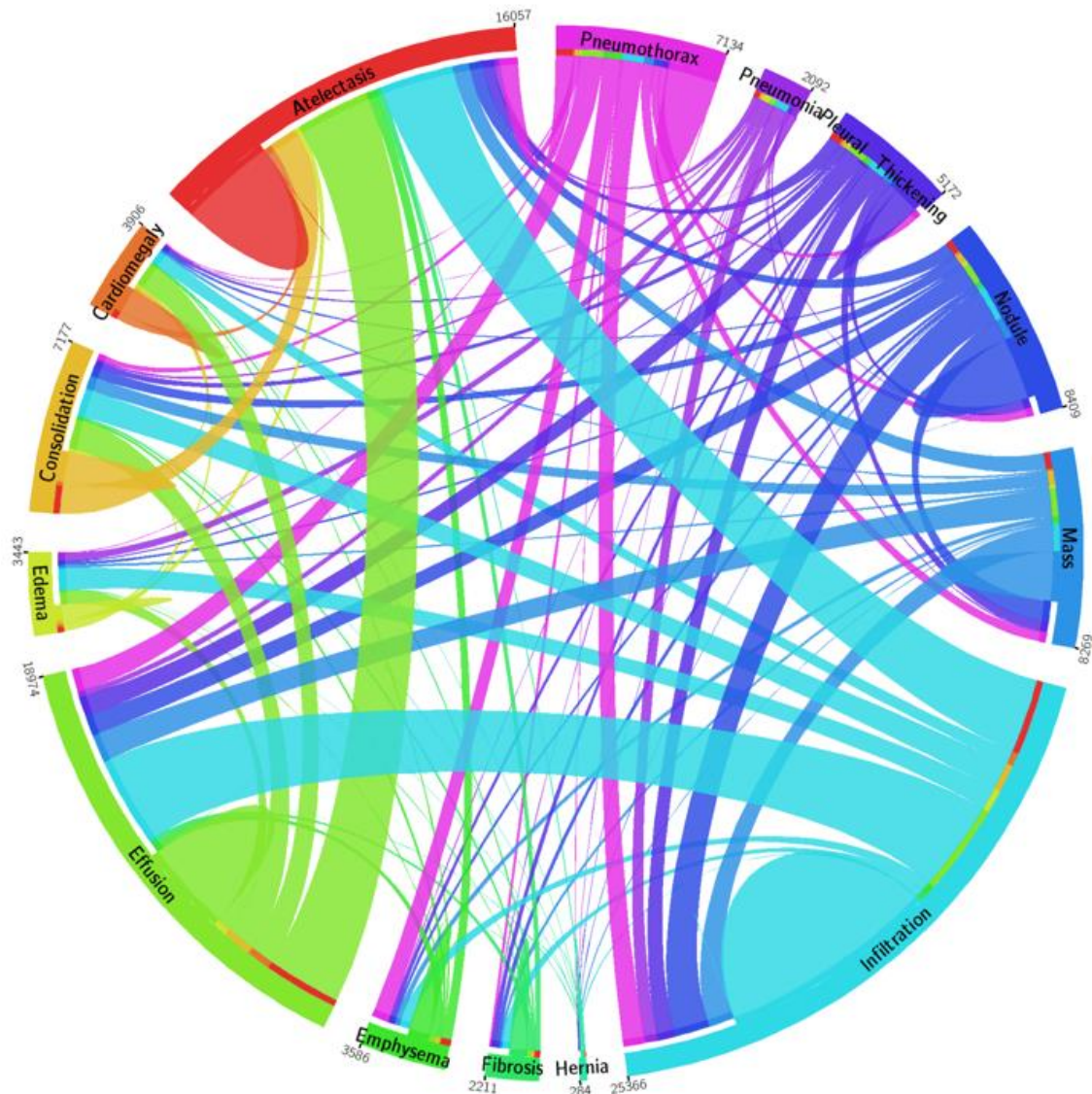
3.1 Dataset

Dataset yang digunakan dalam penelitian ini adalah ChestX-ray8, yang diperkenalkan oleh Wang dan rekan-rekannya dari National Institutes of Health (NIH) (Wang et al. 2017). Dataset ini merupakan salah satu koleksi citra rontgen dada (*chest X-ray*) terbesar yang tersedia secara publik, terdiri dari 108.948 citra rontgen dada tampak depan yang berasal dari 32.717 pasien. Setiap citra diberi label satu atau lebih dari delapan jenis penyakit toraks yang umum, yaitu: *Atelectasis*, *Cardiomegaly*, *Effusion*, *Infiltration*, *Mass*, *Nodule*, *Pneumonia*, dan *Pneumothorax*. Label-label ini diperoleh secara otomatis dari laporan radiologi menggunakan metode pemrosesan bahasa alami (*Natural Language Processing/NLP*), khususnya dengan bantuan alat seperti MetaMap dan DNorm, serta penerapan deteksi negasi dan ketidakpastian untuk menghindari pemberian label yang keliru. Dataset ini bersifat multi-label, artinya satu citra dapat mengandung lebih dari satu jenis kelainan atau penyakit. Gambar 3.1 memperlihatkan korelasi antara label yang dihasilkan. Visualisasi ini menunjukkan adanya hubungan antara berbagai jenis patologi, yang sejalan dengan pengetahuan klinis para radiolog. Sebagai contoh, *consolidation* sering kali berkaitan dengan kondisi *mass* dan *infiltration*.

Setiap citra rontgen dada dalam dataset ChestX-ray8 awalnya berukuran 1024×1024 piksel. Namun, dalam studi ini, seluruh citra diubah ukurannya menjadi 224×224 piksel agar sesuai dengan kebutuhan input model deep learning yang telah dilatih sebelumnya pada dataset ImageNet, yang digunakan sebagai arsitektur dasar. Proses pengubahan ukuran ini dilakukan untuk memastikan kompatibilitas dengan jaringan saraf konvolusional (CNN) yang telah dipra-latih (dengan *imagenet weight*), tanpa menghilangkan fitur visual penting yang diperlukan untuk klasifikasi penyakit.

Dalam penelitian ini, fokus diarahkan pada deteksi lima jenis kelainan utama, yaitu: *Pneumothorax*, *Cardiomegaly*, *Nodule/Mass*, *Consolidation*, dan *Infiltration*. Kelima kelainan ini dipilih karena memiliki relevansi klinis yang tinggi, frekuensi kemunculan yang representatif dalam dataset, serta berkaitan dengan penyakit yang umum ditemukan di Indonesia. Untuk membentuk subset data yang seimbang dan sesuai untuk pelatihan serta evaluasi model, dipilih masing-masing 2.500 citra untuk setiap kelainan, dengan total 12.500 citra. Jumlah ini ditentukan berdasarkan ketersediaan maksimal citra untuk kategori Kardiomegali (2.776 citra), yang kemudian disamakan menjadi 2.500 citra untuk menjaga

konsistensi antar kelas. Meskipun dataset ini tetap bersifat multi-label, pendekatan ini memberikan representasi yang setara untuk masing-masing dari kelima kelainan yang menjadi fokus penelitian.



Gambar 3. 1 Diagram melingkar menunjukkan proporsi dan keterkaitan label pada citra di 14 kelas patologi (Wang et al. 2017).

Setelah dilakukan proses penyeimbangan, dataset disusun agar memiliki jumlah citra yang sama untuk masing-masing kelainan, yaitu 2.500 citra per kategori. Meskipun demikian, karena sifat dataset yang multi-label, tumpang tindih antar label (*overlap*) tetap dipertahankan, sehingga kompleksitas tugas klasifikasi tidak hilang. Oleh karena itu jumlah citra pada setiap label memiliki jumlah yang berbeda-beda (seperti ditampilkan pada **Error! Not a valid bookmark self-reference.**), dicontohkan pada label infiltration, yang

mempunyai 4.678 citra yang sangat berbeda dengan jumlah citra yang di tentukan awal, yaitu 2500 citra. Pendekatan ini memastikan bahwa model tetap belajar dari kondisi yang menyerupai situasi klinis nyata, di mana satu citra dapat menunjukkan lebih dari satu kelainan.

Tabel 3. 1 Dataset yang Digunakan

Abnormality	Number of Images
Pneumothorax	2534
Cardiomegaly	2525
Nodule/Mass	3494
Consolidation	2953
Infiltration	4678

Seperti yang terlihat pada Tabel 3. 2, dataset ini dibagi menjadi dua bagian, yaitu data pelatihan dan data pengujian, dengan proporsi 80:20. Rincian pembagian ini ditampilkan pada Tabel 2, di mana 80% (10.000 citra – sebelum k-fold) digunakan untuk melatih model, sedangkan 20% (2.500 citra) digunakan untuk pengujian.

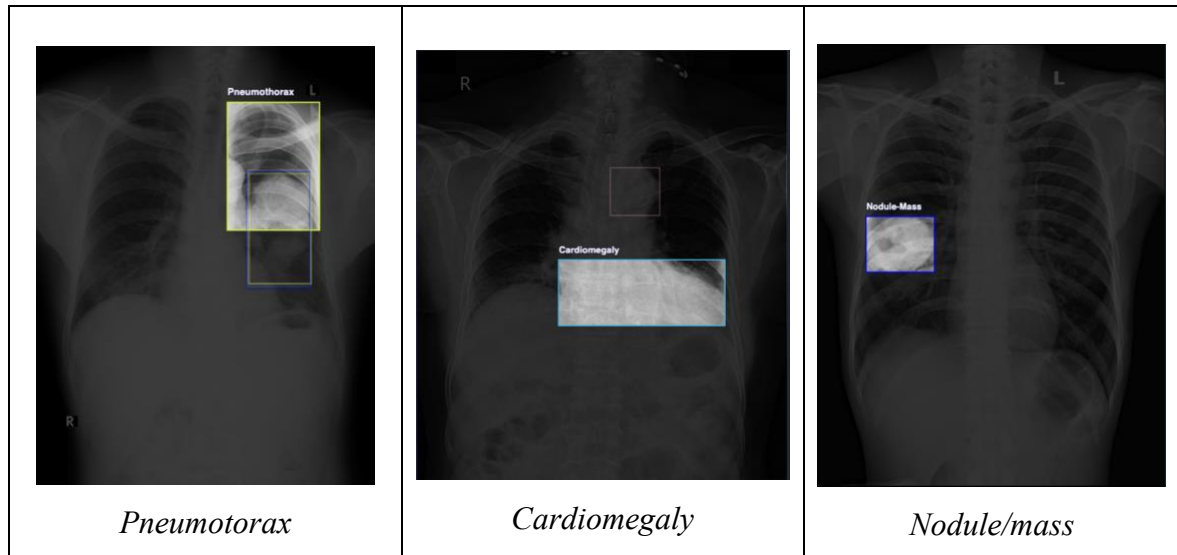
Tabel 3. 2 Pembagian Dataset untuk Pelatihan Model

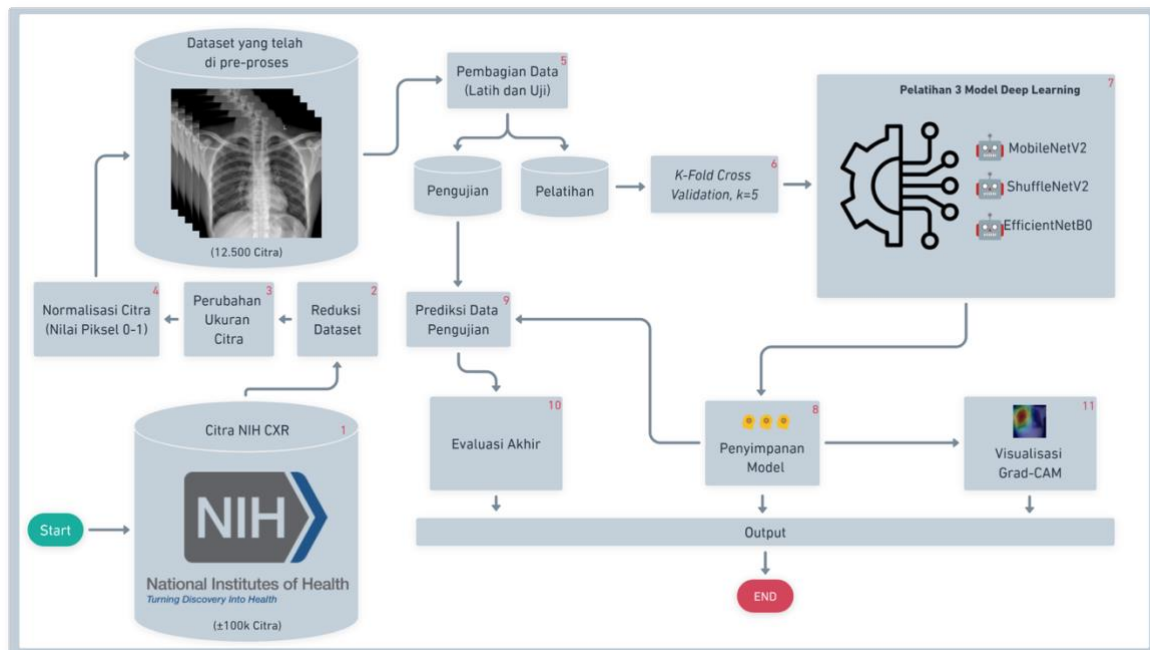
Subset	Number of Images
Training	8000
Testing	2500
Validation (k-fold)	2000
Total	12500

Selain itu, untuk meningkatkan kemampuan generalisasi model dan mengurangi risiko overfitting, kami menerapkan metode cross-validation 5-fold ($K=5$) dengan teknik pembagian acak (*shuffled splitting*). Dalam metode ini, data pelatihan diacak terlebih dahulu sebelum dibagi menjadi lima subset. Model kemudian dilatih pada empat subset dan divalidasi pada subset kelima, dan proses ini diulang hingga semua subset digunakan sebagai data validasi secara bergantian.

Konfigurasi dataset ini menghasilkan input yang semi-seimbang, multi-label, dan relevan secara klinis untuk pelatihan model *deep learning*, yang mendukung tujuan pengembangan sistem deteksi kelainan toraks yang efisien, terutama untuk diterapkan di lingkungan layanan kesehatan dengan sumber daya terbatas.

Pada interpretasi radiografi toraks yang ditunjukkan pada Gambar 3.2, identifikasi abnormalitas sangat penting untuk diagnosis penyakit paru dan kardiovaskular. Beberapa temuan umum yang dapat dikenali melalui rontgen dada meliputi pneumotoraks, kardiomegali, nodul atau massa, konsolidasi, dan infiltrasi. Pneumotoraks ditandai dengan adanya udara dalam rongga pleura yang menyebabkan kolaps paru, sering terlihat pada pasien PPOK. Kardiomegali tampak sebagai pembesaran jantung dengan rasio kardiotoraks lebih dari 50% pada citra rontgen. Nodul atau massa dapat mengindikasikan keganasan paru atau tuberkuloma, terutama pada infeksi tuberkulosis post-primer, yang kadang berhubungan dengan kavitas. Konsolidasi menunjukkan adanya cairan dalam alveoli, umum ditemukan pada kasus pneumonia, terutama pada populasi anak-anak. Sedangkan infiltrasi muncul sebagai opasitas difus atau bercak pada parenkim paru akibat peradangan atau infeksi, yang seringkali disebabkan oleh infeksi virus. Pada kasus ini, dataset yang digunakan terkadang memiliki beberapa abnormalitas dalam satu radiografi toraks, hal ini ditunjukkan pada Gambar 3.3 yang memiliki abnormalitas *nodule/mass* , *infiltration* dan *cardiomegaly* Interpretasi yang akurat terhadap abnormalitas ini sangat krusial untuk menentukan diagnosis dan terapi yang tepat.





Gambar 3. 4 Skenario Penelitian

1. Repositori Dataset

Dataset yang digunakan dalam penelitian ini diambil dari repositori NIH (*National Institutes of Health*), sebagaimana dijelaskan dalam subbagian 3.1. Repositori ini menyediakan kumpulan besar citra rontgen dada (CXR) yang digunakan sebagai data mentah sebelum dilakukan praproses.

2. Reduksi Dataset

Untuk mengakomodasi keterbatasan sumber daya pelatihan, dataset dikurangi menjadi 12.500 citra, masing-masing terdiri dari 2.500 citra untuk lima jenis kelainan utama: *pneumothorax*, *cardiomegaly*, *nodule/mass*, *consolidation*, dan *infiltration*. Pemilihan ini didasarkan pada relevansi klinis, distribusi yang representatif dalam dataset, serta prevalensi di Indonesia. Meskipun bersifat multi-label, pendekatan ini memastikan representasi yang seimbang antar kelas dan mendukung pelatihan model secara optimal.

3. Perubahan Ukuran Citra

Ukuran citra disesuaikan dengan arsitektur model pra-latih (seperti MobileNetV2, ShuffleNetV2, EfficientNetB0) guna memastikan kompatibilitas tanpa perlu modifikasi struktur.

4. Normalisasi Citra

Sebelum proses pelatihan model dimulai, seluruh citra akan dinormalisasi, yakni nilai pikselnya diubah ke dalam rentang antara 0 hingga 1. Langkah ini bertujuan untuk

meningkatkan efisiensi proses pelatihan dan mencegah terjadinya overflow pada gradien, yang dapat muncul akibat penggunaan nilai gradien yang tidak relevan.

5. Pembagian Dataset

Dataset yang telah dipraproses akan dibagi menjadi dua subset utama: data pelatihan dan data pengujian. Data pelatihan akan digunakan untuk membangun dan melatih model, sedangkan data pengujian digunakan untuk mengevaluasi performa akhir dari model yang telah dilatih.

6. K-Fold Cross Validation

Setelah dataset dibagi, subset pelatihan akan kembali dibagi menjadi data pelatihan dan data validasi menggunakan teknik validasi silang (*cross-validation*) K-Fold dengan nilai $K = 5$. Artinya, data akan dibagi menjadi lima bagian dengan proporsi 80% untuk pelatihan dan 20% untuk validasi, secara bergiliran.

7. Pelatihan Model

Dalam studi ini, tiga arsitektur deep learning akan digunakan untuk membangun model deteksi, yaitu: MobileNet, ShuffleNet, dan EfficientNet. Masing-masing model akan dilatih menggunakan data yang telah disiapkan melalui proses sebelumnya.

8. Penyimpanan Model

Setelah proses pelatihan selesai, model yang telah dilatih akan disimpan. Penyimpanan ini memungkinkan penggunaan ulang model untuk pengujian lebih lanjut maupun implementasi pada sistem nyata.

9. Prediksi Data Pengujian

Model yang telah disimpan akan digunakan untuk memprediksi data pengujian. Ini bertujuan untuk mengukur kemampuan generalisasi model terhadap data baru yang tidak digunakan selama proses pelatihan.

10. Evaluasi Akhir

Kinerja masing-masing model akan dievaluasi berdasarkan hasil prediksi pada data pengujian. Evaluasi ini akan memberikan informasi mengenai efektivitas dan akurasi model dalam mengidentifikasi berbagai jenis kelainan pada citra CXR.

11. Visualisasi Grad-CAM

Untuk mendukung interpretasi klinis, dilakukan visualisasi Grad-CAM pada sebagian citra yang diprediksi. Teknik ini menghasilkan *heatmap* yang menyoroti area-area penting dalam citra yang dianggap berkontribusi terhadap prediksi model. Hasil ini dapat membantu radiolog dalam memahami lokasi kelainan yang terdeteksi oleh model.

3.3 Spesifikasi Lingkungan Eksperimen

Seluruh proses pelatihan dan evaluasi dalam penelitian ini dilakukan menggunakan platform Google Colab Pro, dengan memanfaatkan perangkat keras berkinerja tinggi serta lingkungan pemrograman *deep learning* yang optimal. Rincian spesifikasi perangkat keras dan perangkat lunak yang digunakan ditunjukkan pada Tabel 3.3 berikut.

Tabel 3. 3 Spesifikasi Perangkat Keras dan Perangkat Lunak Eksperimen

Komponen	Spesifikasi
CPU	Intel(R) Xeon(R) CPU @ 2.20GHz
Jumlah Core	6 Core Kinerja / 12 Core Logis
GPU	NVIDIA A100-SXM4-40GB (40 GB VRAM)
Memori GPU	40960 MiB
Versi CUDA	12.5 (Rilis 12.5, V12.5.82)
RAM	90 GB
Penyimpanan Disk	253 GB
Sistem Operasi	Ubuntu 22.04.4 LTS
Versi Kernel	6.1.123+
Versi Python	3.11.12
Framework Deep Learning	PyTorch 2.6.0+cu124

Pelatihan model dilakukan menggunakan PyTorch 2.6.0 dengan akselerasi GPU berbasis CUDA, sehingga mampu menangani dataset besar seperti ChestX-ray8 dengan efisien. Penggunaan GPU NVIDIA A100 dengan kapasitas memori 40 GB memungkinkan pelatihan menggunakan *batch size* besar tanpa mengalami kendala kehabisan memori.

3.4 Konfigurasi *Hyperparameter*

Untuk menjaga konsistensi dan keadilan antar semua model yang dibandingkan (MobileNetV2, ShuffleNetV2, dan EfficientNetB0), Tabel 3.4 menunjukkan satu set hiperparameter standar digunakan selama pelatihan.

Tabel 3. 4 Konfigurasi *Hyperparameter*

Hiperparameter	Nilai	Keterangan
----------------	-------	------------

Ukuran Gambar Input	(224, 224)	Ukuran gambar setelah diubah (lebar × tinggi) sesuai kebutuhan <i>backbone pretrained model</i>
Jumlah Channel	3	Gambar berformat RGB dengan tiga <i>channel</i> warna
Jumlah Kelas	5	<i>Pneumothorax, Cardiomegaly, Nodule/Mass, Consolidation, Infiltration</i>
Ukuran Batch	512	Ukuran batch besar untuk stabilitas pelatihan dan efisiensi memori
Jumlah Epoch	50	Jumlah siklus pelatihan tetap untuk evaluasi adil
Learning Rate Awal	1.00E-04	Laju awal pembaruan bobot saat optimisasi
Optimizer	Adam	Optimizer adaptif karena keunggulannya dalam menangani dataset medis
Fungsi Loss	BCEWithLogitsLoss	Digunakan untuk klasifikasi multi-label
Threshold Label	0.5	Ambang batas prediksi probabilitas menjadi label positif
Freeze Backbone	Tidak (False)	Semua layer model dilatih kembali (<i>fine-tuning</i> penuh)

Berikut ini adalah penjelasan komprehensif konfigurasi hiperparameter yang digunakan:

- Ukuran Gambar (224×224): Disesuaikan dengan kebutuhan arsitektur *pretrained* (seperti EfficientNet dan MobileNet) agar kompatibel tanpa perubahan struktur, sekaligus menjaga keseimbangan antara detail visual dan kebutuhan memori.
- Ukuran Batch (512): Dengan memanfaatkan kapasitas besar dari GPU A100, batch besar membantu menstabilkan pembaruan bobot (gradien) dan mempercepat konvergensi pelatihan (Keskar et al. 2017).
- Jumlah Epoch (50): Jumlah epoch sebanyak 50 dipertimbangkan karena tidak memakan waktu komputasi yang terlalu lama dan juga berdasarkan percobaan sebelumnya yang menggunakan jumlah epoch 100, hasil evaluasi metrik tidak jauh berbeda dengan 50 epoch.

- Optimizer (Adam): Optimizer Adam dipilih karena keunggulannya dalam menangani dataset medis yang sering kali memiliki distribusi fitur yang jarang (sparse) (Akter et al. 2021; Kolonne et al. 2021). Adam secara otomatis menyesuaikan learning rate untuk setiap parameter.
- Learning Rate Awal ($1e-4$): Learning rate kecil membantu menghindari *overshooting* (melompati minima global), penting untuk dataset medis berukuran terbatas (Akter et al. 2021; Kolonne et al. 2021).
- Fungsi Loss (BCEWithLogitsLoss): BCEWithLogitsLoss menggabungkan aktivasi *sigmoid* dan *binary cross-entropy* dalam satu fungsi yang stabil secara numerik, ideal untuk tugas klasifikasi multi-label seperti deteksi banyak abnormalitas dalam satu citra dada (Kufel et al. 2023).
- Threshold Label (0.5): Probabilitas di atas 0.5 diklasifikasikan sebagai positif. Threshold ini umum digunakan dalam aplikasi diagnosis medis untuk mengoptimalkan keseimbangan sensitivitas dan spesifisitas.
- Pelatihan Backbone Tanpa Pembekuan (Full Fine-Tuning): Semua layer, termasuk fitur low-level seperti edge dan tekstur, dilatih ulang. Hal ini memungkinkan model beradaptasi secara lebih efektif terhadap karakteristik unik citra X-ray, yang berbeda dari citra alami (seperti ImageNet) (Davila, Colan, and Hasegawa 2024).

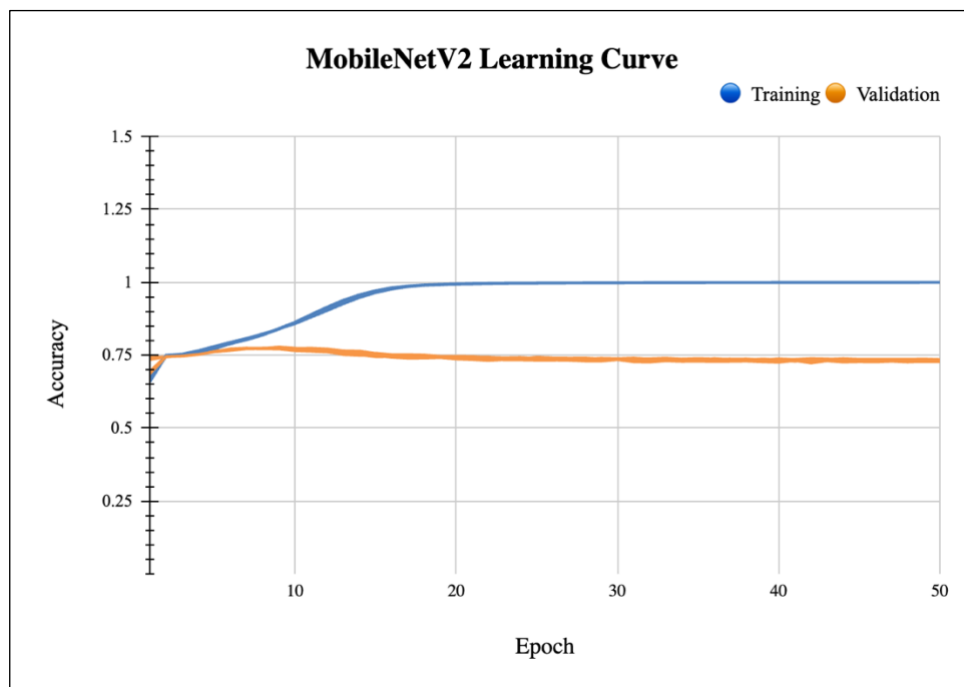
Dengan pendekatan ini, hasil yang diperoleh diharapkan mampu diterapkan secara nyata dalam praktik klinis, khususnya di daerah dengan keterbatasan fasilitas radiologi.

BAB 4

Hasil dan Pembahasan

4.1 Analisis Model pada Saat Pelatihan

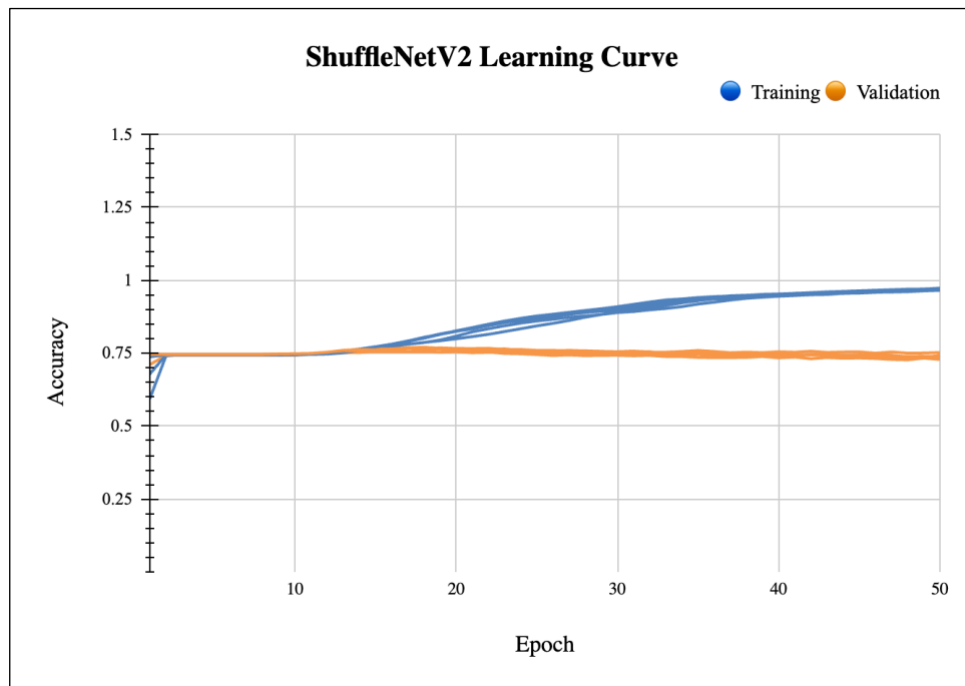
Untuk mengevaluasi kestabilan pelatihan dan kemampuan generalisasi dari masing-masing model, dilakukan analisis terhadap kurva pembelajaran (*learning curve*) yang menunjukkan akurasi pelatihan dan validasi pada setiap fold. Gambar 4. 1 – 4.3 memperlihatkan kurva akurasi dari lima fold pelatihan dan validasi untuk masing-masing model: MobileNetV2, ShuffleNetV2, dan EfficientNetB0. Pendekatan ini memungkinkan identifikasi kecenderungan overfitting atau ketidakstabilan pelatihan yang mungkin tidak terlihat jika hanya ditampilkan nilai rata-rata.



Gambar 4. 1 Learning Curve dari MobileNetV2

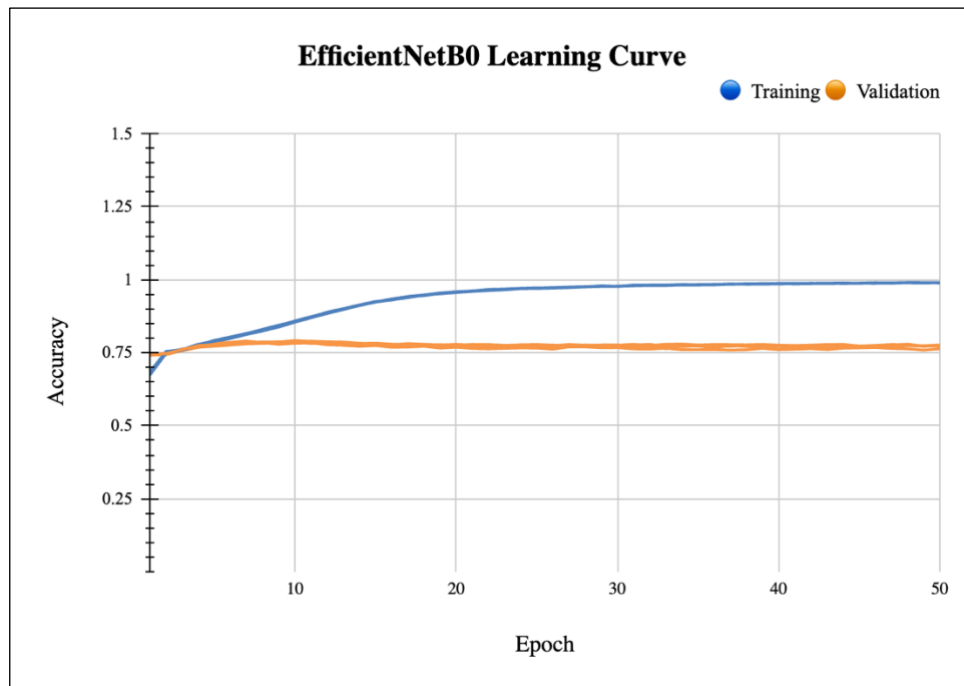
Model MobileNetV2 menunjukkan proses pelatihan yang konsisten di semua fold, dengan akurasi pelatihan mencapai lebih dari 0.99 sekitar epoch ke-20. Namun, akurasi validasi tetap jauh lebih rendah (sekitar 0.75) dan cenderung menurun seiring bertambahnya epoch, yang dimulai pada epoch ke-10 dan seterusnya. Jarak yang semakin besar antara akurasi pelatihan dan validasi ini menandakan adanya overfitting yang signifikan, di mana model sangat baik dalam menghafal data pelatihan namun kurang mampu menggeneralisasi pada data baru. Pola ini terjadi secara konsisten di semua fold, menunjukkan bahwa

permasalahan ini bukan kebetulan. Hal ini mengindikasikan bahwa kapasitas model terlalu rendah terhadap kompleksitas data.



Gambar 4. 2 Learning Curve dari ShuffleNetV2

Berbeda dengan itu, ShuffleNetV2 memperlihatkan variasi yang lebih besar antar fold dalam hal akurasi pelatihan. Akurasi pelatihan akhir bervariasi dari sekitar 0.85 hingga hampir 0.98, menandakan bahwa beberapa fold mengalami kesulitan untuk mencapai konvergensi optimal. Sementara itu, akurasi validasi cenderung datar dan stabil pada kisaran 0.74–0.76 untuk semua fold. Pola ini menunjukkan konvergensi yang lebih lambat dan kurang stabil, yang bisa disebabkan oleh desain arsitektur model yang lebih sederhana atau sensitivitas terhadap distribusi data per fold. Meskipun overfitting tidak terlalu tampak, performa keseluruhan ShuffleNetV2 relatif rendah, kemungkinan karena kapasitas model yang terbatas atau kendala dalam proses optimisasi.



Gambar 4. 3 Learning Curve dari EfficientNetB0

EfficientNetB0 menunjukkan perilaku pelatihan yang paling seimbang. Akurasi pelatihan meningkat secara stabil dan merata hingga mendekati ~ 0.98 di semua fold pada sekitar epoch ke-35 hingga ke-40. Akurasi validasi juga relatif konstan dan lebih mendekati akurasi pelatihan (~ 0.76), dengan selisih yang hampir sama dengan model-model sebelumnya.

4.2 Evaluasi Kinerja Model pada Dataset Validasi

Untuk menentukan model deep learning yang paling efektif dalam mendeteksi kelainan toraks, tiga arsitektur ringan—MobileNetV2, ShuffleNetV2, dan EfficientNetB0—telah dievaluasi berdasarkan metrik F1-score, akurasi, presisi, recall, dan AUC-ROC. Hasil evaluasi yang diperoleh merupakan rata-rata dari proses validasi silang K-fold, sebagaimana ditampilkan pada Tabel 4. 1 Model dengan performa terbaik akan dilanjutkan ke tahap pengujian akhir untuk analisis yang lebih komprehensif.

Berdasarkan Tabel 4. 1, EfficientNetB0 secara konsisten menunjukkan performa terbaik dibandingkan dua model lainnya pada seluruh metrik evaluasi, dengan pencapaian F1-score tertinggi ($0,563 \pm 0,009$), akurasi ($0,773 \pm 0,005$), recall ($0,558 \pm 0,010$), dan AUC-ROC ($0,772 \pm 0,005$). Nilai standar deviasi yang rendah pada setiap metrik menunjukkan bahwa performa model ini stabil di seluruh fold validasi. Sebaliknya, MobileNetV2 menunjukkan performa yang cukup baik dengan F1-score sebesar $0,494 \pm 0,008$, namun memiliki kelemahan pada metrik presisi ($0,486 \pm 0,009$), yang mengindikasikan tingkat *false positive* yang lebih tinggi.

Tabel 4. 1 Kinerja Model pada Data Validasi (Rata-rata dan Standar Deviasi dari Validasi Silang K-Fold)

Model	F1 Score	Accuracy	Precision	Recall	AUC-ROC
MobileNetV2	0.494 ± 0.008	0.732 $\pm \mathbf{0.003}$	0.486 ± 0.009	0.505 ± 0.009	0.72 ± 0.005
ShuffleNetV2	0.463 ± 0.014	0.74 ± 0.008	0.512 ± 0.014	0.431 ± 0.016	0.697 ± 0.01
EfficientNetB0	0.563 ± 0.009	0.773 ± 0.005	0.573 ± 0.007	0.558 ± 0.01	0.772 ± 0.005

ShuffleNetV2 menunjukkan performa terendah, dengan F1-score ($0,463 \pm 0,014$) dan recall ($0,431 \pm 0,016$) paling rendah, yang menandakan seringnya model ini belum bisa mendeteksi kasus positif. Hal ini merupakan keterbatasan serius dalam konteks diagnostik medis.

MobileNetV2 menunjukkan stabilitas yang cukup baik, terutama pada metrik akurasi ($\pm 0,003$), namun variabilitas yang lebih tinggi pada metrik presisi ($\pm 0,009$) dan recall ($\pm 0,009$) mengindikasikan adanya inkonsistensi dalam mendeteksi kelainan. Sebaliknya, ShuffleNetV2 memperlihatkan fluktuasi performa tertinggi, seperti terlihat dari standar deviasi F1-score ($\pm 0,014$) dan recall ($\pm 0,016$), menjadikannya model paling tidak andal karena tingkat false negative yang tidak konsisten. Ketidakstabilan ini menjadi risiko besar dalam pencitraan medis, mengingat kelalaian dalam mendeteksi kelainan bisa berdampak fatal.

Meskipun hasil rata-rata memberikan gambaran umum mengenai kinerja model, pemilihan model terbaik juga mempertimbangkan F1-score tertinggi yang dicapai pada setiap fold validasi. Tabel 4. 2 menyajikan nilai F1-score per fold untuk masing-masing model.

Tabel 4. 2 F1-Score Setiap Fold pada Tiap-tiap Model

Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Best
MobileNetV2	0.488	0.496	0.501	0.504	0.483	Fold 4 (0.504)
ShuffleNetV2	0.454	0.453	0.448	0.473	0.485	Fold 5 (0.485)
EfficientNetB0	0.546	0.566	0.566	0.568	0.569	Fold 5 (0.569)

Berdasarkan analisis per fold, EfficientNetB0 berhasil mencapai F1-score tertinggi sebesar 0,569 pada Fold ke-5, melampaui MobileNetV2 (0,504 pada Fold ke-4) dan ShuffleNetV2 (0,485 pada Fold ke-5). Selain itu, EfficientNetB0 secara konsisten memperoleh nilai F1-score di atas 0,546 pada seluruh fold, sedangkan kedua model lainnya menunjukkan variabilitas kinerja yang lebih besar antar fold.

Untuk menyempurnakan proses pemilihan model, Tabel 4. 3 merangkum F1-score tertinggi yang dicapai oleh masing-masing model dan menentukan apakah model tersebut memenuhi syarat sebagai model dengan performa terbaik.

Tabel 4. 3 Pemilihan Model Terbaik Berdasarkan F1-Score Tertinggi

Model	F1 Score	Model Terbaik?
MobileNetV2	Fold 4 (0.504)	Tidak
ShuffleNetV2	Fold 5 (0.485)	Tidak
EfficientNetB0	Fold 5 (0.569)	Ya

Dengan F1-score tertinggi pada satu fold (0,569 di Fold ke-5) serta performa unggul secara keseluruhan pada seluruh metrik evaluasi, EfficientNetB0 secara meyakinkan terpilih sebagai model terbaik untuk deteksi kelainan toraks. Nilai recall yang tinggi ($0,558 \pm 0,010$) menjadi keunggulan penting dalam pencitraan medis, karena mengurangi kemungkinan false negative sangat krusial untuk mencegah diagnosis yang terlewat. Selain itu, nilai AUC-ROC sebesar 0,772 semakin memperkuat kemampuan model ini dalam membedakan antara kasus normal dan abnormal pada semua ambang batas.

Dengan terpilihnya EfficientNetB0 sebagai model terbaik, langkah selanjutnya adalah mengevaluasi kinerjanya pada dataset pengujian.

4.3 Evaluasi Kinerja Model pada Dataset Pengujian

Untuk menilai penerapan model dalam situasi nyata, MobileNetV2, ShuffleNetV2, dan EfficientNetB0 dievaluasi menggunakan dataset pengujian. Evaluasi difokuskan pada

kemampuan masing-masing model dalam mengklasifikasikan lima jenis kelainan toraks: pneumotoraks, kardiomegali, nodul/massa, konsolidasi, dan infiltrasi. Kinerja model diukur menggunakan metrik akurasi, presisi, recall, F1-score, dan AUC-ROC, guna memberikan penilaian yang seimbang terhadap efektivitas klasifikasi.

Analisis ini mencakup performa masing-masing kelas secara individual, yang kemudian diikuti dengan perbandingan menggunakan nilai rata-rata makro (macro-average) untuk menentukan model yang paling efektif dalam mendeteksi kelainan toraks.

4.3.1 Analisis Hasil Pengujian MobileNetV2

Kami mengevaluasi kinerja MobileNetV2 dalam mengklasifikasikan lima jenis kelainan toraks—pneumotoraks, kardiomegali, nodul/massa, konsolidasi, dan infiltrasi—dengan menggunakan metrik akurasi, presisi, recall, F1-score, dan AUC-ROC. Ringkasan hasil evaluasi disajikan pada Tabel 4.4:

Tabel 4. 4 Hasil Pengujian MobileNetV2

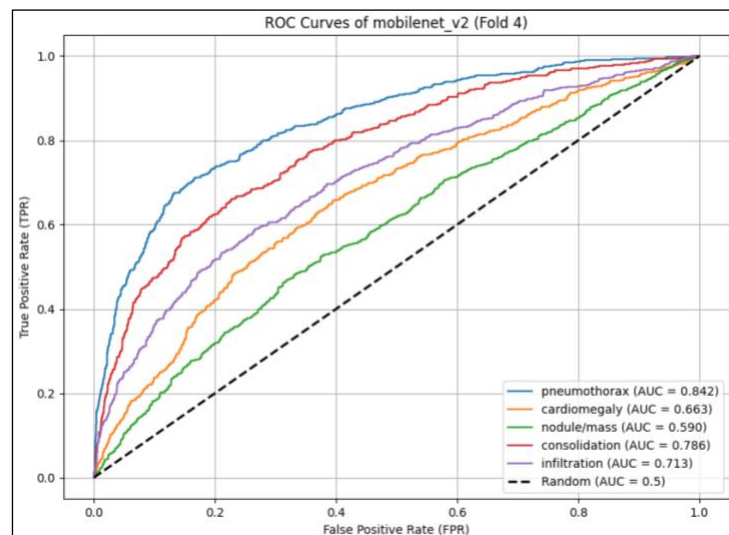
Label	Accuracy	Precision	Recall	F1-Score	ROC-AUC
pneumothorax	0.835	0.609	0.590	0.599	0.842
cardiomegaly	0.710	0.398	0.415	0.406	0.663
nodule/mass	0.592	0.457	0.490	0.473	0.590
consolidation	0.802	0.480	0.528	0.503	0.786
infiltration	0.731	0.468	0.506	0.486	0.713
MacroAVG	0.734	0.483	0.506	0.494	0.719

Model ini menunjukkan performa yang sedang, ditunjukkan oleh nilai rata-rata makro F1-score sebesar 0,494 dan AUC-ROC sebesar 0,719. Nilai recall (0,506) yang lebih tinggi dibandingkan presisi (0,483) mengindikasikan bahwa MobileNetV2 mampu mendeteksi lebih banyak kasus abnormal, meskipun dengan konsekuensi meningkatnya jumlah false positive. Deteksi pneumotoraks merupakan yang paling menonjol, dengan akurasi tertinggi (0,835), presisi (0,609), dan AUC-ROC (0,842), kemungkinan karena ciri radiologisnya yang jelas dan mudah dikenali.

Sebaliknya, kinerja terburuk terlihat pada deteksi kardiomegali, dengan nilai presisi (0,398), recall (0,415), dan F1-score (0,406) yang paling rendah, menandakan bahwa pembesaran jantung sulit dikenali akibat tampilan radiologisnya yang lebih halus. Deteksi konsolidasi menunjukkan recall yang cukup baik (0,528), meskipun nilai presisinya lebih

rendah (0,480), yang mencerminkan performa sedang. Deteksi nodul/massa (AUC-ROC = 0,590) dan infiltrasi (AUC-ROC = 0,713) menunjukkan efektivitas klasifikasi yang lebih rendah, kemungkinan disebabkan oleh karakteristik yang bervariasi dan tidak terlalu jelas secara radiologis.

Kurva AUC-ROC pada Gambar 4.4 semakin memperjelas kemampuan MobileNetV2 dalam membedakan masing-masing abnormalitas. Pneumotoraks menunjukkan performa terbaik (AUC-ROC = 0,842), menegaskan kemampuannya yang kuat dalam klasifikasi. Konsolidasi mengikuti dengan AUC-ROC sebesar 0,786, mengindikasikan sensitivitas yang cukup baik. Infiltrasi mencetak skor 0,713, sedangkan kardiomegali (0,663) dan nodul/massa (0,590) menunjukkan performa terendah, mendekati kemampuan klasifikasi acak. Keterbatasan model dalam mendeteksi kardiomegali dan nodul/massa kemungkinan disebabkan oleh variasi bentuk yang halus pada kasus kardiomegali serta ukuran kecil dan keragaman tampilan nodul paru pada kasus nodul/massa.



Gambar 4. 4 ROC-Curve dari MobileNetV2

Untuk meningkatkan performa model, berbagai upaya perbaikan dapat dilakukan, seperti augmentasi data guna meningkatkan generalisasi khususnya pada kasus kardiomegali, serta penerapan teknik ekstraksi fitur lanjutan untuk mendeteksi lesi kecil seperti nodul atau massa. Meskipun MobileNetV2 menunjukkan performa yang baik dalam mendeteksi pneumotoraks dan berpotensi dimanfaatkan dalam sistem triase otomatis di lingkungan dengan sumber daya terbatas, model ini masih memerlukan optimisasi lebih lanjut untuk meningkatkan akurasi dalam mendeteksi kardiomegali dan nodul/massa.

4.3.2 Analisis Hasil Pengujian ShuffleNetV2

Kinerja ShuffleNetV2 dievaluasi secara konsisten seperti pada bagian sebelumnya (4.1.1). Hasil evaluasi tersebut disajikan pada Tabel 4. 5.

Tabel 4. 5 Hasil Pengujian ShuffleNetV2

Label	Accuracy	Precision	Recall	F1-Score	ROC-AUC
pneumothorax	0.859	0.692	0.588	0.636	0.827
cardiomegaly	0.732	0.424	0.331	0.372	0.662
nodule/mass	0.608	0.473	0.449	0.461	0.614
consolidation	0.815	0.516	0.455	0.483	0.759
infiltration	0.743	0.487	0.426	0.455	0.704
MacroAVG	0.751	0.518	0.450	0.481	0.713

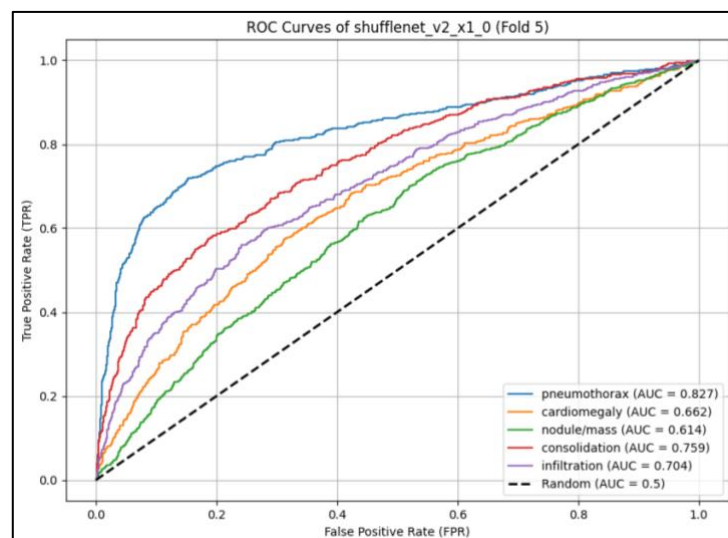
Nilai rata-rata makro F1-score sebesar 0,481 menunjukkan performa klasifikasi yang tergolong sedang, sedikit lebih rendah dibandingkan MobileNetV2 (0,494). Meskipun begitu, ShuffleNetV2 memiliki presisi yang lebih tinggi (0,518), namun disertai recall yang lebih rendah (0,450), yang berarti model ini menghasilkan lebih sedikit false positive tetapi cenderung melewati lebih banyak kasus positif. Nilai terbaik model ini terdapat pada akurasi (0,751) dan AUC-ROC (0,713), yang menunjukkan kemampuan sedang dalam membedakan antara kasus normal dan abnormal. Namun, recall yang rendah (0,450) menjadi kelemahan utama, karena mengindikasikan banyak kasus positif yang tidak terdeteksi.

ShuffleNetV2 menunjukkan kinerja terbaik pada deteksi pneumotoraks, dengan akurasi tertinggi (0,859) dan AUC-ROC (0,827). Presisi sebesar 0,692 merupakan yang tertinggi di antara semua kondisi yang diuji. Meskipun jumlah false positive rendah, nilai recall (0,588) masih belum optimal. Sebaliknya, deteksi kardiomegali merupakan kelemahan utama model ini, dengan recall (0,331), F1-score (0,372), dan AUC-ROC (0,662) terendah, yang mencerminkan kesulitan model dalam membedakan antara kardiomegali dan kondisi normal.

Performa pada deteksi nodul/massa (AUC-ROC = 0,614, F1-score = 0,461) dan konsolidasi (AUC-ROC = 0,759, F1-score = 0,483) tergolong sedang. Terdapat sedikit peningkatan dibandingkan MobileNetV2 dalam mendeteksi nodul/massa, namun hasilnya sedikit lebih buruk untuk konsolidasi. Deteksi infiltrasi (AUC-ROC = 0,704, F1-score =

0,455) juga berada di bawah performa MobileNetV2, menunjukkan bahwa ShuffleNetV2 kesulitan dalam mengenali kondisi paru-paru yang menyebar secara difus.

Kurva AUC-ROC yang ditampilkan pada Gambar 4. 5 semakin memperjelas kemampuan klasifikasi ShuffleNetV2 untuk masing-masing kondisi. Pneumotoraks (AUC = 0,827) merupakan kelas dengan performa terbaik, menunjukkan pemisahan yang kuat antara kasus normal dan abnormal. Konsolidasi (AUC = 0,759) dan infiltrasi (AUC = 0,704) menunjukkan kemampuan klasifikasi yang sedang, meskipun masih terdapat beberapa kesalahan klasifikasi. Sebaliknya, kardiomegali (AUC = 0,662) dan nodul/massa (AUC = 0,614) tetap menjadi kelas dengan performa terlemah, karena kesulitan model dalam membedakannya dari kondisi normal.



Gambar 4. 5 ROC-Curve dari ShuffleNetV2

Untuk meningkatkan performa klasifikasi ShuffleNetV2, optimisasi recall pada deteksi kardiomegali menjadi hal yang penting, salah satunya melalui penerapan fungsi loss berbobot guna meningkatkan sensitivitas. Selain itu, peningkatan ekstraksi fitur menggunakan teknik seperti filter konvolusional multi-skala dan mekanisme atensi berpotensi membantu dalam mendeteksi abnormalitas berukuran kecil seperti nodul.

4.3.3 Analisis Hasil Pengujian EfficientNetB0

Pada bagian ini, kami menganalisis kinerja EfficientNetB0 pada data pengujian, dengan fokus pada kemampuannya dalam mengklasifikasikan kelainan toraks sebagaimana ditampilkan pada Tabel 4. 6 dan divisualisasikan pada Gambar 4. 6.

Tabel 4. 6 Hasil Pengujian EfficientNetB0

Label	Accuracy	Precision	Recall	F1-Score	ROC-AUC
pneumothorax	0.868	0.696	0.660	0.678	0.880
cardiomegaly	0.756	0.490	0.430	0.458	0.699
nodule/mass	0.618	0.488	0.519	0.503	0.646
consolidation	0.844	0.587	0.598	0.592	0.835
infiltration	0.772	0.547	0.548	0.548	0.764
MacroAVG	0.772	0.562	0.551	0.556	0.765

EfficientNetB0 menunjukkan performa unggul dibandingkan MobileNetV2 dan ShuffleNetV2 dalam seluruh kondisi yang diuji, dengan pencapaian F1-score rata-rata makro tertinggi (0,556) dan AUC-ROC tertinggi (0,765), melampaui MobileNetV2 (0,494; 0,719) dan ShuffleNetV2 (0,481; 0,713). Nilai recall yang tertinggi (0,551) juga menunjukkan bahwa EfficientNetB0 lebih efektif dalam mengidentifikasi kasus abnormal dan mengurangi kemungkinan false negative.

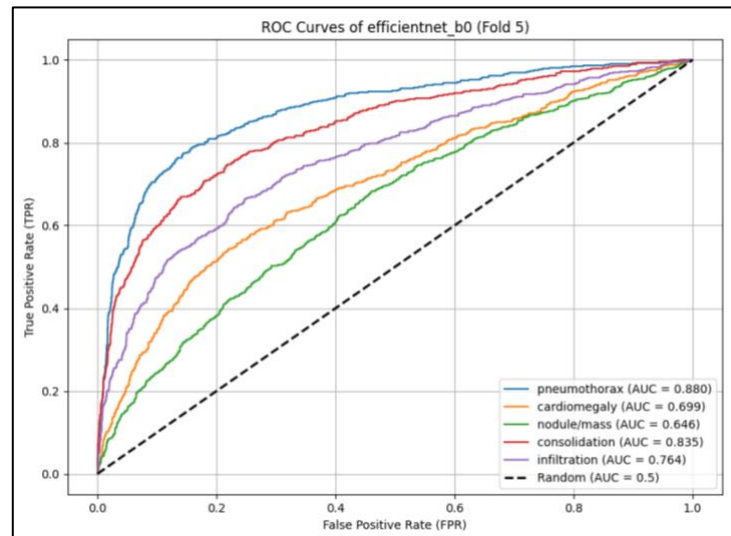
Performa terbaik terlihat pada deteksi pneumotoraks, dengan akurasi (0,868), presisi (0,696), F1-score (0,678), dan AUC-ROC (0,880) tertinggi. Hal ini menunjukkan bahwa pneumotoraks tetap menjadi kondisi yang paling mudah dikenali oleh model berbasis CNN, dengan EfficientNetB0 secara signifikan melampaui MobileNetV2 (AUC = 0,842) dan ShuffleNetV2 (AUC = 0,827).

Pada deteksi nodul/massa, EfficientNetB0 menunjukkan peningkatan paling signifikan, dengan recall tertinggi (0,519) dan AUC-ROC lebih baik (0,646) dibandingkan MobileNetV2 (0,590) dan ShuffleNetV2 (0,614). Ini menunjukkan sensitivitas yang lebih baik dalam mendeteksi abnormalitas paru berukuran kecil, meskipun performa keseluruhan masih tergolong sedang dan perlu dioptimalkan lebih lanjut.

EfficientNetB0 juga unggul dalam deteksi konsolidasi, dengan recall (0,598), presisi (0,587), dan AUC-ROC (0,835) tertinggi, menjadikannya sangat efektif untuk diagnosis pneumonia atau infeksi paru. Deteksi infiltrasi juga menunjukkan peningkatan signifikan, dengan recall yang lebih tinggi (0,548), F1-score tertinggi (0,548), dan AUC-ROC sebesar 0,764, yang lebih baik dibandingkan MobileNetV2 (0,713) dan ShuffleNetV2 (0,704).

Meskipun deteksi kardiomegali tetap menjadi tantangan, EfficientNetB0 tetap mencatatkan hasil terbaik dengan AUC-ROC (0,699) dan F1-score (0,458), melampaui

MobileNetV2 (0,663; 0,438) dan ShuffleNetV2 (0,662; 0,431). Walaupun pembesaran jantung sulit dikenali karena perbedaan bentuk yang halus, EfficientNetB0 menunjukkan peningkatan yang cukup berarti dalam klasifikasi kondisi ini.



Gambar 4. 6 ROC-Curve dari EfficientNetB0

EfficientNetB0 secara konsisten mencapai nilai AUC-ROC tertinggi untuk seluruh jenis kelainan, sebagaimana ditunjukkan pada kurva di Gambar 8 dan perbandingan antar model pada Tabel 4. 7, yang menegaskan keunggulannya dalam kemampuan klasifikasi. Peningkatan performa yang paling signifikan terlihat pada deteksi pneumotoraks, nodul/massa, dan konsolidasi. Meskipun demikian, kardiomegali tetap menjadi tantangan terbesar dalam proses klasifikasi, menunjukkan bahwa kondisi ini masih sulit dibedakan dari citra normal.

Tabel 4. 7 Perbandingan AUC-ROC Model

Label	MobileNetV2	ShuffleNetV2	EfficientNetB0
pneumothorax	0.842	0.827	0.880
cardiomegaly	0.663	0.662	0.699
nodule/mass	0.590	0.614	0.646
consolidation	0.786	0.759	0.835
infiltration	0.713	0.704	0.764

Untuk mengoptimalkan kinerja EfficientNetB0 lebih lanjut, klasifikasi kardiomegali dapat ditingkatkan melalui pretraining berbasis segmentasi serta penambahan jumlah data berlabel khusus untuk kasus kardiomegali dalam dataset. Sementara itu, deteksi nodul/massa

dapat ditingkatkan dengan menerapkan ekstraksi fitur multi-skala dan mekanisme atensi guna memperkuat kemampuan model dalam mengenali lesi kecil secara lebih akurat.

4.3.4 Analisis Macro-Average pada Semua Metrik

Analisis komparatif ini adalah mengevaluasi performa keseluruhan dari MobileNetV2, ShuffleNetV2, dan EfficientNetB0 berdasarkan nilai rata-rata makro untuk metrik akurasi, presisi, recall, F1-score, dan AUC-ROC. Seperti yang ditunjukkan pada Tabel 4. 8, EfficientNetB0 secara konsisten menunjukkan performa terbaik di semua metrik, menjadikannya model yang paling sesuai untuk deteksi kelainan toraks.

Tabel 4. 8 Macro-Average Kinerja Model

Label	Accuracy	Precision	Recall	F1-Score	ROC-AUC
MobileNetV2	0.734	0.483	0.506	0.494	0.719
ShuffleNetV2	0.751	0.518	0.450	0.481	0.713
EfficientNetB0	0.772	0.562	0.551	0.556	0.765

EfficientNetB0 mencatatkan akurasi (0,772), presisi (0,562), recall (0,551), dan F1-score (0,556) tertinggi, mengonfirmasi keseimbangan klasifikasi yang unggul dari model ini. Nilai recall yang tinggi menunjukkan lebih sedikit kasus abnormal yang terlewat, yang sangat penting dalam pencitraan medis, sementara nilai presisi yang tinggi membantu mengurangi jumlah *false positives*. Sebaliknya, MobileNetV2 (F1-score 0,494; AUC-ROC 0,719) dan ShuffleNetV2 (F1-score 0,481; AUC-ROC 0,713) menunjukkan performa klasifikasi yang lebih lemah secara keseluruhan. ShuffleNetV2 menjadi model dengan performa paling rendah, terutama karena recall yang rendah (0,450), yang menyebabkan peningkatan jumlah *false negative*.

Di antara MobileNetV2 dan ShuffleNetV2, MobileNetV2 sedikit lebih unggul karena memiliki nilai recall dan F1-score yang lebih baik, yang menunjukkan kemampuannya dalam menangkap lebih banyak kasus positif meskipun presisinya sedikit lebih rendah dibandingkan ShuffleNetV2. Namun, kedua model ini masih belum mampu menyaingi EfficientNetB0, yang menunjukkan keseimbangan terbaik antara sensitivitas (recall) dan spesifisitas (precision). Dengan nilai AUC-ROC tertinggi (0,765), EfficientNetB0 menjadi model yang paling andal dalam membedakan antara kasus normal dan abnormal, menjadikannya pilihan paling tepat untuk diterapkan dalam konteks klinis.

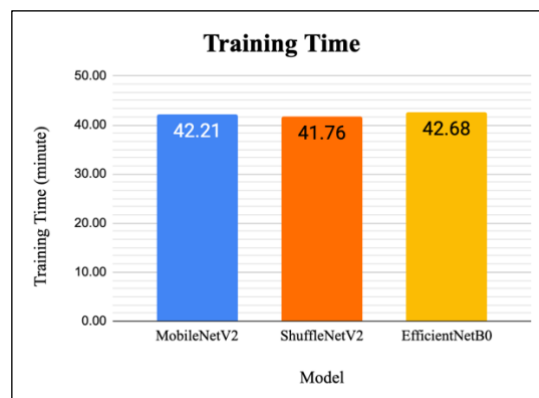
Secara keseluruhan, EfficientNetB0 menempati peringkat pertama. Meskipun MobileNetV2 dapat dijadikan alternatif yang cukup layak, nilai presisi dan recall yang lebih rendah membuatnya kurang efektif dibandingkan EfficientNetB0. Sementara itu, meskipun ShuffleNetV2 memiliki presisi yang lebih tinggi, jumlah false negative yang tinggi menjadikannya kurang dapat diandalkan untuk penggunaan medis. Oleh karena itu, EfficientNetB0 merupakan kandidat terbaik untuk diterapkan dalam aplikasi pencitraan medis di dunia nyata.

4.4 Analisis Performa Model secara Keseluruhan

Selain beberapa metrik yang dibahas sebelumnya (misalnya: akurasi), efisiensi komputasi dan kebutuhan sumber daya dari model pembelajaran mendalam (*deep learning*) sangat penting dalam menentukan kelayakan implementasi di dunia nyata, khususnya di lingkungan dengan keterbatasan sumber daya seperti layanan kesehatan di daerah terpencil atau pada perangkat mobile. Bagian ini menyajikan analisis menyeluruh terhadap setiap model dengan mempertimbangkan waktu pelatihan, waktu inferensi, jumlah parameter yang dapat dilatih atau kompleksitas model, dan akurasi klasifikasi. Pembahasan dibagi menjadi dua fase utama: fase pelatihan dan fase pengujian.

4.4.1 Performa Efisiensi Pelatihan

Efisiensi dalam proses pelatihan merupakan aspek penting, terutama pada lingkungan dengan keterbatasan perangkat keras atau kebutuhan pelatihan ulang secara berkala, seperti pada sistem pembelajaran berkelanjutan dalam konteks medis. Dalam penelitian ini, durasi pelatihan untuk tiga model CNN—MobileNetV2, ShuffleNetV2, dan EfficientNetB0—menggunakan dataset berisi 10.000 (data pelatihan dan validasi) citra rontgen dada ditunjukkan pada Gambar 4. 7.



Gambar 4. 7 Waktu Pelatihan (dalam menit) untuk Setiap Model

Meskipun selisih waktu pelatihan di antara model hanya sekitar satu menit, perbedaan ini mencerminkan implikasi dari desain arsitektural masing-masing. ShuffleNetV2 menunjukkan waktu pelatihan tercepat, sejalan dengan desainnya yang ringan dan efisien melalui teknik channel *shuffling* dan operasi konvolusi yang disederhanakan.

MobileNetV2 menggunakan pendekatan *depthwise separable convolution* untuk mengurangi beban komputasi, sementara EfficientNetB0, meski sedikit lebih lama waktu pelatihannya, mengandalkan strategi *compound scaling* yang mengatur secara bersamaan kedalaman, lebar, dan resolusi jaringan untuk mencapai efisiensi komputasi tanpa mengorbankan kualitas ekstraksi fitur.

Secara keseluruhan, ketiga model menunjukkan efisiensi pelatihan yang relatif setara. Oleh karena itu, dalam konteks implementasi nyata, efisiensi pelatihan saja tidak cukup sebagai penentu utama pemilihan model. Faktor lain seperti waktu inferensi, tingkat akurasi, dan kompleksitas model menjadi lebih relevan dalam menentukan model yang paling sesuai untuk digunakan di lingkungan klinis.

4.4.2 Performa Pengujian secara Menyeluruh

Evaluasi performa setiap model selama fase pengujian dilakukan berdasarkan tiga hal utama yang ditunjukkan pada Tabel 4. 9, yaitu: waktu inferensi pada 2.500 citra X-ray, jumlah parameter yang dapat dilatih atau kompleksitas model, dan akurasi klasifikasi akhir.

Tabel 4. 9 Akurasi, Waktu Pengujian, dan Kompleksitas Model

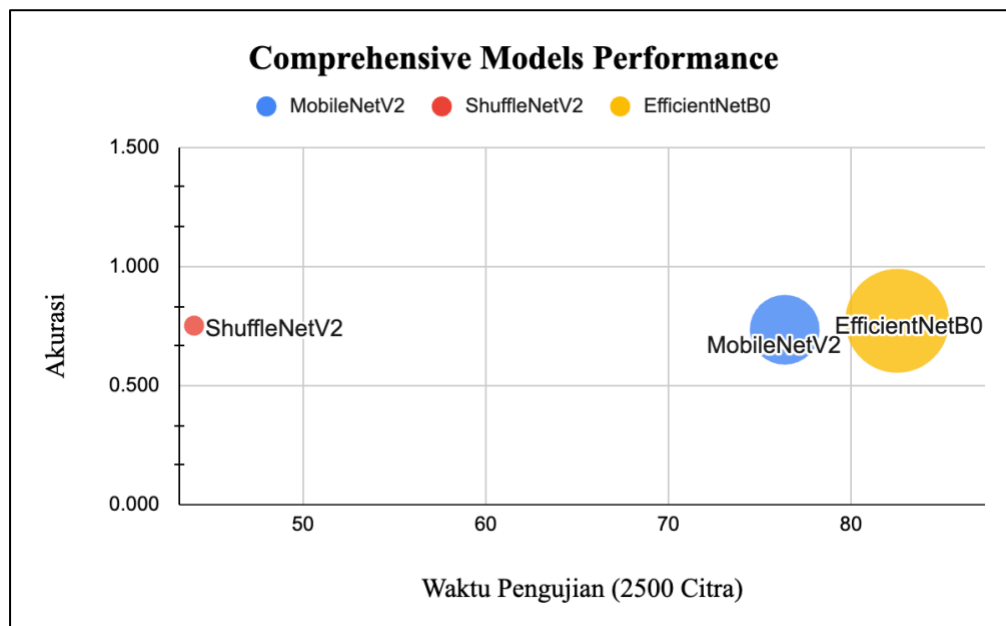
Model	Akurasi	Waktu Pengujian (s)	Waktu Pengujian per Citra (ms)	Jumlah Parameter Latih
MobileNetV2	0.734	76.37	~30.5	2,230,277
ShuffleNetV2	0.751	44.04	~17.6	1,258,729
EfficientNetB0	0.772	82.53	~33.0	4,013,953

Hasil pengujian menunjukkan bahwa EfficientNetB0 mencapai akurasi tertinggi sebesar 0,772, menandakan kapasitas generalisasi dan representasi fitur yang kuat. Namun, pencapaian ini disertai dengan waktu inferensi paling lama (82,53 detik) dan jumlah parameter terbanyak (lebih dari 4 juta), yang dapat menjadi kendala pada lingkungan dengan keterbatasan komputasi, seperti perangkat edge atau sistem dengan daya rendah.

Di sisi lain, ShuffleNetV2 tampil sebagai model paling efisien, dengan waktu inferensi tercepat (44,04 detik) dan parameter paling sedikit (~1,26 juta). Meski ringan, model ini tetap menghasilkan akurasi yang kompetitif (0,751), bahkan melampaui MobileNetV2 (0,734), yang memiliki parameter lebih banyak (~2,23 juta) dan waktu inferensi lebih lama (76,37 detik). Hal ini menunjukkan bahwa efisiensi model tidak selalu harus dikompromikan dengan performa klasifikasi, asalkan arsitektur dirancang dengan baik.

Jika dilihat dari rata-rata waktu inferensi per gambar, ShuffleNetV2 membutuhkan waktu hanya sekitar 17,6 milidetik, jauh lebih cepat dibanding EfficientNetB0 (33,0 ms/citra) dan MobileNetV2 (30,5 ms/citra). Kecepatan ini menjadikan ShuffleNetV2 sangat cocok untuk aplikasi yang membutuhkan respon cepat, seperti skrining massal atau deteksi penyakit di lokasi terpencil dengan komputasi perangkat yang terbatas.

Untuk memberikan pemahaman visual terhadap hubungan antara efisiensi dan akurasi model, digunakan diagram bubble chart yang ditunjukkan oleh Gambar 4. 8, di mana sumbu X merepresentasikan waktu inferensi (semakin ke kiri berarti semakin cepat), sumbu Y menunjukkan akurasi (semakin ke atas semakin baik), dan ukuran gelembung melambangkan kompleksitas model berdasarkan jumlah parameternya. Dalam visualisasi ini, EfficientNetB0 menempati posisi kanan atas dengan gelembung terbesar, menggambarkan akurasi tinggi namun kompleksitas tinggi pula. ShuffleNetV2 berada di kiri bawah dengan gelembung terkecil, mencerminkan efisiensi tinggi dengan akurasi tetap solid. Sementara itu, MobileNetV2 berada di tengah, menunjukkan karakteristik moderat pada ketiga aspek.



Gambar 4. 8 Kinerja Model secara Komprehensif dalam *Bubble Chart*

4.4.3 Pengujian dan Analisis Deployment pada Perangkat Edge (Android)

Sebagai tindak lanjut dari batasan masalah yang menyatakan bahwa model deep learning akan diimplementasikan pada perangkat dengan keterbatasan sumber daya, dilakukan pengujian terhadap deployment model pada perangkat mobile menggunakan Android emulator. Pengujian ini bertujuan untuk mengetahui kelayakan dan efisiensi operasional model dalam kondisi nyata yang terbatas secara komputasi, seperti yang ditemukan pada sistem embedded, perangkat genggam, atau lokasi dengan infrastruktur terbatas.

Pengujian dilakukan menggunakan aplikasi Android yang dikembangkan dengan integrasi langsung terhadap model hasil pelatihan, dan dijalankan pada Android emulator dengan spesifikasi yang ditunjukkan pada Tabel 4. 10.

Tabel 4. 10 Spesifikasi Perangkat Android Emulator untuk *Deployment*

Komponen	Spesifikasi
Nama Perangkat	Medium Phone (Emulator)
Android Version	Android 14 (API Level 36)
CPU Cores	1 Core
RAM	2 GB
Internal Storage	6 GB
Screen Resolution	1080 × 2400 px
Screen Density	420 dpi
ABI	arm64-v8a

Hasil pengujian pada perangkat emulator menunjukkan bahwa efisiensi waktu inferensi sangat dipengaruhi oleh arsitektur model yang digunakan, hal ini dapat dilihat pada Tabel 4. 11. ShuffleNetV2 secara konsisten mencatat waktu inferensi tercepat, dengan rata-rata hanya 103 milidetik untuk memproses satu citra, jauh lebih efisien dibandingkan MobileNetV2 (3822 ms) dan EfficientNetB0 (3950 ms). Hal ini memperkuat klaim bahwa ShuffleNetV2 sangat ideal untuk diterapkan pada aplikasi real-time di perangkat dengan sumber daya terbatas. Meskipun secara absolut terdapat perbedaan signifikan antara performa inference di emulator dan di workstation (GPU A100), pola perbandingan relatif antar model tetap konsisten. Artinya, keunggulan efisiensi arsitektural yang diamati selama pengujian di workstation masih relevan dan tercermin pada lingkungan edge seperti perangkat mobile.

EfficientNetB0 memang tetap menjadi model dengan akurasi tertinggi (0,772), namun waktu inferensi yang jauh lebih lambat pada perangkat dengan sumber daya terbatas menjadi hambatan nyata untuk penerapan dalam konteks waktu-nyata.

Dari sisi validitas hasil, karena model yang dijalankan pada perangkat Android merupakan hasil konversi langsung dari model terlatih di lingkungan workstation—tanpa adanya kuantisasi atau modifikasi struktur—maka dapat disimpulkan bahwa akurasi dan hasil prediksi tidak mengalami perubahan. Dengan demikian, eksperimen ini memberikan justifikasi bahwa ketiga model dapat dijalankan secara fungsional di lingkungan edge tanpa mengorbankan integritas prediksi yang telah diuji sebelumnya di lingkungan pelatihan.

Tabel 4. 11 Pengujian Waktu Inferensi pada Android Emulator

Testing Time on Mobile Device				
Models	Time Inference (ms)			Avg. Time
	Image 1	Image 2	Image 3	
MobileNetV2	2916	5200	3351	3822
ShuffleNetV2	113	113	82	103
EfficientNetB0	4684	3261	3906	3950

Kesimpulannya, penelitian ini menyoroti efisiensi tiga model deep learning—EfficientNetB0, ShuffleNetV2, dan MobileNetV2—dalam klasifikasi citra X-ray, dengan mempertimbangkan akurasi, waktu inferensi, dan kompleksitas model. Meskipun EfficientNetB0 unggul dalam akurasi, ShuffleNetV2 lebih efisien untuk perangkat dengan keterbatasan sumber daya, sementara MobileNetV2 menawarkan solusi seimbang. Penting dicatat bahwa ketiga model ini, berdasarkan studi literatur sebelumnya, termasuk dalam kategori mobile model yang dirancang untuk berjalan secara optimal pada perangkat kecil seperti smartphone atau sistem tertanam (Tobiasz et al. 2023). Hal ini menjadikan ketiganya relevan untuk implementasi diagnostik berbasis AI yang membutuhkan efisiensi tinggi, portabilitas, dan kemampuan beroperasi dalam kondisi infrastruktur terbatas.

4.4.4 Pertimbangan Pemilihan Arsitektur Model

Sebagai pembanding terhadap model ringan yang digunakan dalam studi ini, beberapa baseline arsitektur non-mobile telah terbukti memberikan performa klasifikasi yang tinggi dalam mendeteksi kelainan toraks pada citra X-ray. Misalnya, (Nam et al. 2021) dalam

pengembangan sistem DLAD-10 menggunakan arsitektur ResNet34 dan melatih model tersebut pada dataset besar yang terdiri atas 146.717 radiograf dari lebih dari 100.000 pasien. Hasilnya, model ini mampu mencapai nilai AUC-ROC sebesar 0.997, menunjukkan akurasi yang sangat tinggi dalam klasifikasi multi-label abnormalitas toraks.

Namun, performa tersebut dicapai dengan mengorbankan kompleksitas model yang tinggi, baik dari segi kompleksitas model, waktu pelatihan, maupun kebutuhan memori dan komputasi saat inferensi. Model seperti ResNet34 umumnya tidak dirancang untuk kebutuhan deployment pada perangkat dengan keterbatasan sumber daya seperti smartphone atau edge device.

Selain itu, (Wang et al. 2017) menggunakan arsitektur ResNet50 dalam studi ChestX-ray8, yaitu sebuah dataset publik yang berisi 108.948 citra frontal-view X-ray dari 32.717 pasien. ResNet50 dalam eksperimen mereka mencatat nilai AUC tertinggi sebesar 0.8141 pada label Cardiomegaly dan terendah sebesar 0.5609 pada label Mass. Hal ini menunjukkan bahwa meskipun arsitektur yang lebih dalam digunakan, performa klasifikasi belum tentu optimal pada semua jenis penyakit, dan penggunaan model besar tetap memiliki keterbatasan skalabilitas dalam praktik klinis.

Sebaliknya, model ringan (*mobile models*) seperti MobileNetV2, ShuffleNetV2, dan EfficientNet-B0, yang digunakan dalam studi ini, memiliki jumlah parameter jauh lebih kecil serta kecepatan inferensi yang tinggi. Meski nilai AUC yang diperoleh hanya sekitar ~0.73 pada dataset ChestX-ray8 yang telah direduksi, performa tersebut masih tergolong kompetitif dan layak digunakan dalam konteks perangkat edge. Dalam skenario seperti pra-skrining massal, triase cepat di IGD, klinik dengan keterbatasan radiolog, atau integrasi dengan aplikasi kesehatan mobile, kecepatan dan efisiensi model menjadi faktor yang lebih dominan dibandingkan sedikit peningkatan akurasi.

Untuk memperjelas perbandingan antara model baseline dan model ringan, berikut disajikan perbandingan performa dan kompleksitas beberapa arsitektur CNN yang ditampilkan pada Tabel 4. 12.

Tabel 4. 12 Perbandingan performa dan kompleksitas beberapa arsitektur CNN dalam klasifikasi kelainan toraks.

Model	Dataset	AUC (Terbaik)	Kompleksitas	Cocok untuk Mobile
ResNet34	DLAD-10 (Nam et al., 2021)	0.997	Tinggi	Tidak
ResNet50	ChestX-ray8 (Wang et al., 2017)	0.8141	Sangat tinggi	Tidak
<i>Mobile Models</i>	ChestX-ray8 (tereduksi)	~0.73	Sangat rendah	Ya

Dalam praktik klinis, waktu interpretasi citra oleh radiolog bervariasi tergantung pada kompleksitas kasus. Berdasarkan studi oleh Nam et al. (2021), waktu interpretasi rata-rata per citra radiograf dada oleh radiolog dalam kondisi normal adalah sekitar 20 detik. Ini menunjukkan bahwa meskipun waktu tersebut terbilang cepat, kehadiran volume pasien tinggi di instalasi gawat darurat atau rumah sakit besar dapat mengakumulasi beban kerja secara signifikan, menyebabkan keterlambatan pelaporan.

Model ShuffleNetV2 dalam studi ini memiliki waktu inferensi rata-rata sekitar 103 milidetik per citra, yang berarti dapat memproses satu citra dalam waktu ~0,1 detik, atau ratusan kali lebih cepat dibandingkan interpretasi manual oleh radiolog (perlu diingat bahwa alat ini ‘membantu’, bukan ‘menggantikan’ peran radiolog). Hal ini menjadi sangat relevan untuk pra-skrining massal atau triase cepat (misalnya saat wabah penyakit menular), sistem telemedisin atau rumah sakit dengan radiolog terbatas, dan pengurangan waktu tunggu pelaporan pada kasus kritis, yang secara langsung dapat meningkatkan prioritas penanganan pasien.

4.4.5 Peran AI dalam Mengurangi Waktu Interpretasi Radiograf Dada dan Beban Kerja

Model ShuffleNetV2 dalam studi ini memiliki waktu inferensi rata-rata sekitar 103 milidetik per citra, yang berarti dapat memproses satu citra dalam waktu ~0,1 detik, atau ratusan kali lebih cepat dibandingkan interpretasi manual oleh radiolog (perlu diingat bahwa alat ini

‘membantu’, bukan ‘menggantikan’ peran radiolog). Hal ini menjadi sangat relevan untuk pra-skrining massal atau triase cepat (misalnya saat wabah penyakit menular), sistem telemedisin atau rumah sakit dengan radiolog terbatas, dan pengurangan waktu tunggu pelaporan pada kasus kritis, yang secara langsung dapat meningkatkan prioritas penanganan pasien.

Dalam praktik klinis, waktu interpretasi citra radiograf dada oleh radiolog sangat bergantung pada kompleksitas temuan dan beban kerja saat itu. Studi oleh (Nam et al. 2021) menunjukkan bahwa waktu interpretasi rata-rata radiograf dada oleh radiolog adalah sekitar 20 detik per citra, tetapi ini belum termasuk waktu tambahan untuk kasus kompleks, dokumentasi, dan pengambilan keputusan.

Namun, peningkatan volume pasien, khususnya di instalasi gawat darurat atau rumah sakit rujukan, dapat memperparah keterlambatan pelaporan akibat beban kerja radiolog yang berlebih. Dalam konteks ini, pemanfaatan algoritma pembelajaran mendalam (deep learning) seperti DLAD-10 terbukti meningkatkan sensitivitas deteksi kelainan dan secara signifikan mempercepat waktu pelaporan untuk kasus kritis dan urgensi tinggi. DLAD-10, dalam simulasi pembacaan oleh enam radiolog, menurunkan waktu rata-rata pelaporan kasus kritis dari 3371 detik menjadi 640 detik, dan kasus urgensi dari 2127 detik menjadi 1840 detik (semuanya $p < 0,01$), serta mengurangi waktu interpretasi per citra dari 23,5 detik menjadi 20,5 detik.

Studi lain oleh Stanford University (CheXNeXt) mengungkapkan bahwa sebuah algoritma AI mampu mengevaluasi 14 jenis kelainan dalam sekitar 90 detik untuk 420 citra, dibandingkan tiga jam waktu total yang dibutuhkan oleh para radiolog (Armitage 2018). Untuk 11 dari 14 patologi, performa AI menyamai atau bahkan melampaui radiolog manusia. Ini menggambarkan potensi AI sebagai alat triase digital berkualitas tinggi, terutama di wilayah dengan akses radiolog yang terbatas.

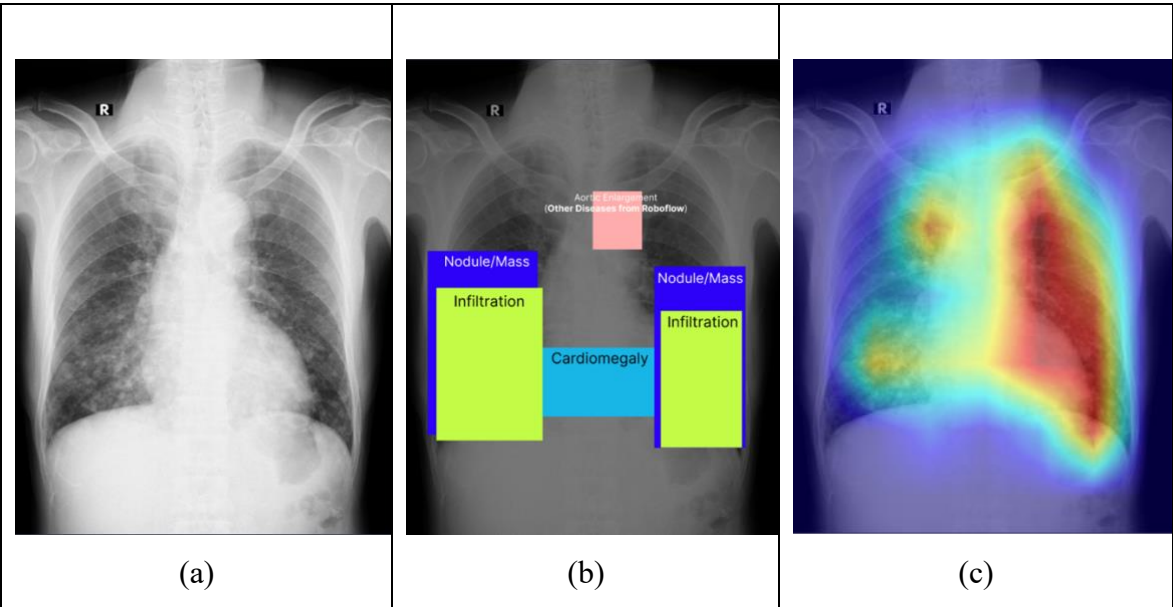
Lebih lanjut, studi terbaru dari Yonsei University (Shin et al., 2023) mengevaluasi penggunaan klinis AI komersial (Insight CXR v3) dalam interpretasi radiograf dada (Morton 2023). Hasilnya menunjukkan bahwa AI dapat mempercepat waktu pembacaan untuk citra normal dari 13,1 detik menjadi 10,8 detik. Total waktu interpretasi seluruh citra juga berkurang dari 14,8 detik menjadi 13,3 detik ($p < 0,001$). Namun, menariknya, saat AI mendeteksi kelainan, waktu interpretasi malah sedikit bertambah. Hal ini kemungkinan disebabkan oleh radiolog yang perlu memverifikasi hasil AI dan mendokumentasikan lebih detail.

Dalam studi Kurmukov et al. (2024), juga dijelaskan bahwa integrasi AI dapat mengurangi kelelahan mental radiolog, meningkatkan akurasi, serta menyediakan prioritas otomatis terhadap kasus-kasus kritis, yang secara sistemik berdampak positif terhadap efisiensi layanan radiologi (Kurmukov et al. 2024).

Dalam konteks penelitian ini, penggunaan model ringan seperti ShuffleNetV2 menghasilkan waktu inferensi rata-rata hanya 103 milidetik per citra, yang berarti ratusan kali lebih cepat dibandingkan interpretasi manual oleh manusia. Meskipun tidak menggantikan peran radiolog, kecepatan ini sangat relevan untuk pra-skrining massal, triase cepat di fasilitas dengan sumber daya terbatas, serta pengurangan waktu tunggu pelaporan pada kasus gawat darurat.

4.5 Analisis Visualisasi grad-CAM

Berdasarkan Gambar 4. 9 dan Tabel 4. 13, visualisasi Grad-CAM dari EfficientNetB0 memperlihatkan baik keunggulan maupun keterbatasannya dalam mendeteksi kelainan toraks. Label dan lokasi ground-truth pada analisis ini diperoleh dari dataset beranotasi milik Roboflow (*Roboflow 2022*), yang menyediakan kotak pembatas (bounding box) akurat untuk berbagai jenis kelainan. Dataset ini menjadi acuan utama untuk mengevaluasi kemampuan model dalam mengidentifikasi dan melokalisasi penyakit toraks secara tepat.



Gambar 4. 9 Citra Paru dengan Visualisasi Grad-CAM: (a) Citra ground-truth, (b) label asli dan lokasi abnormalitas berdasarkan data dari Roboflow (*Roboflow 2022*), (c) hasil prediksi Grad-CAM oleh EfficientNetB0.

Tabel 4. 13 Hasil Prediksi dari 1 Citra pada Analisis Visualisasi Grad-CAM

Label	<i>Actual Label</i>	EfficientNetB0	<i>is match?</i>
pneumothorax	No	0.93 - Yes	No
cardiomegaly	Yes	0.06 - No	No
nodule/mass	Yes	0.01 - No	No
consolidation	No	0.00 - No	No
infiltration	Yes	0.70 - Yes	Yes

Namun, model secara keliru memprediksi adanya pneumotoraks dengan probabilitas tinggi (0,93), yang kemungkinan besar disebabkan oleh kebingungan antara infiltrasi dan paru yang kolaps. Di sisi lain, infiltrasi berhasil diidentifikasi dengan benar (0,70). Meskipun anotasi ground-truth menunjukkan adanya kardiomegali dan nodul/massa, EfficientNetB0 hanya memberikan probabilitas yang sangat rendah terhadap kedua kondisi tersebut (masing-masing 0,06 dan 0,01), yang menandakan sensitivitas deteksi yang masih rendah terhadap kondisi tersebut.

Heatmap Grad-CAM menunjukkan aktivasi yang kuat di area paru kanan, sejalan dengan salah deteksi pneumotoraks, sementara aktivasi yang lemah di area jantung menjelaskan kegagalan model dalam mengenali kardiomegali. Hal ini mengindikasikan adanya kelemahan dalam membedakan jenis kelainan yang mirip serta keterbatasan dalam ekstraksi fitur untuk kondisi yang lebih halus seperti pembesaran jantung dan nodul kecil.

Untuk meningkatkan akurasi, diperlukan pretraining berbasis segmentasi khusus untuk kardiomegali, penerapan ekstraksi fitur multi-skala untuk deteksi nodul, serta pelatihan model dengan data yang lebih seimbang. Meskipun EfficientNetB0 menunjukkan kemampuan pelokalan yang baik untuk opasitas paru, model ini masih perlu disempurnakan lebih lanjut untuk mengurangi kesalahan deteksi (false positive) dan meningkatkan sensitivitas terhadap patologi yang kurang mencolok, khususnya pada kasus kardiomegali dan nodul/massa.

BAB 5

Kesimpulan dan Saran

5.1 Kesimpulan

Penelitian ini mengevaluasi kinerja MobileNetV2, ShuffleNetV2, dan EfficientNetB0 untuk deteksi otomatis kelainan toraks pada citra rontgen dada, sebagai respons terhadap keterbatasan jumlah radiolog dan tantangan diagnosis di Indonesia. Menggunakan dataset ChestX-ray8, model dievaluasi berdasarkan akurasi, presisi, recall, F1-score, dan AUC-ROC, dengan pemilihan model terbaik ditentukan melalui nilai F1-score tertinggi pada setiap fold validasi.

EfficientNetB0 muncul sebagai model dengan performa terbaik, mencatatkan F1-score rata-rata makro sebesar 0,556 dan AUC-ROC sebesar 0,765, melampaui MobileNetV2 dan ShuffleNetV2. EfficientNetB0 unggul dalam mendeteksi pneumotoraks, konsolidasi, dan infiltrasi, namun masih lemah pada deteksi kardiomegali dan nodul/massa. MobileNetV2 menunjukkan kinerja sedang dengan risiko overfitting, sementara ShuffleNetV2 paling efisien secara komputasi, tetapi memiliki recall terendah, yang berisiko tinggi dalam konteks klinis. Ketiga model juga berhasil dijalankan di perangkat Android dengan konfigurasi minimal, di mana ShuffleNetV2 mencatat waktu inferensi tercepat (103 ms), menguatkan posisinya sebagai model paling ringan dan responsif untuk deployment pada perangkat dengan sumber daya terbatas.

Analisis Grad-CAM menunjukkan bahwa model ini mampu melakukan pelokalan yang baik pada kasus pneumotoraks dan konsolidasi, namun masih menunjukkan kesalahan klasifikasi pada deteksi kardiomegali dan nodul/massa, yang menggarisbawahi tantangan dalam membedakan abnormalitas dengan karakteristik halus.

5.2 Saran

Temuan ini menunjukkan potensi EfficientNetB0 dalam mendukung diagnosis berbasis AI di lingkungan layanan kesehatan dengan sumber daya terbatas. Namun, hasil tersebut juga menekankan pentingnya pretraining berbasis segmentasi, ekstraksi fitur multi-skala, serta peningkatan interpretabilitas model dan pengurangan false positive untuk memastikan aplikasi klinis yang andal. Selain itu, berdasarkan penurunan akurasi yang teramati pada model MobileNetV2 dan EfficientNetB0 setelah epoch ke-10, disarankan untuk menerapkan teknik *early stopping* guna mencegah *overfitting* dan menjaga performa model agar tetap optimal dalam proses pelatihan.

Daftar Pustaka

- Agarap, Abien Fred. 2018. "Deep Learning Using Rectified Linear Units (ReLU)." <https://arxiv.org/abs/1803.08375v2>.
- Akter, Shamima, F. M. Javed Mehedi Shamrat, Sovon Chakraborty, Asif Karim, and Sami Azam. 2021. "COVID-19 Detection Using Deep Learning Algorithm on Chest X-Ray Images." *Biology* 10(11):1174. doi:10.3390/biology10111174.
- Almutairi, Tahani, Mohamed Maher, and Ouiem Bchir. 2022. "X-Ray Based COVID-19 Classification Using Lightweight EfficientNet." *Journal on Artificial Intelligence* 4(3):167–87. doi:10.32604/jai.2022.032974.
- Alzubaidi, Laith, Jinglan Zhang, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, J. Santamaría, Mohammed A. Fadhel, Muthana Al-Amidie, and Laith Farhan. 2021. "Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions." *Journal of Big Data* 8(1):53. doi:10.1186/s40537-021-00444-8.
- Amin, Hina, and Waqas J. Siddiqui. 2025. "Cardiomegaly." in *StatPearls*. Treasure Island (FL): StatPearls Publishing.
- An, Le, Kexin Peng, Xing Yang, Pan Huang, Yan Luo, Peng Feng, and Biao Wei. 2022. "E-TBNet: Light Deep Neural Network for Automatic Detection of Tuberculosis with X-Ray DR Imaging." *Sensors* 22(3):821. doi:10.3390/s22030821.
- An, Qiuyu, Wei Chen, and Wei Shao. 2024. "A Deep Convolutional Neural Network for Pneumonia Detection in X-Ray Images with Attention Ensemble." *Diagnostics (Basel, Switzerland)* 14(4):390. doi:10.3390/diagnostics14040390.
- Andarini, Sita, Elisna Syahrudin, Nathaniel Aditya, Jamal Zaini, Ferry Dwi Kurniawan, Sabrina Ermayanti, Noni Novisari Soeroso, Sri Melati Munir, Andreas Infianto, Ana Rima, Ungky Agus Setyawan, Laksmi Wulandari, Haryati Haryati, Ida Ayu Jasminarti, and Arif Santoso. 2023. "Indonesian Society of Respiriology (ISR) Consensus Statement on Lung Cancer Screening and Early Detection in Indonesia." *Jurnal Respirologi Indonesia* 43(2):144–50. doi:10.36497/jri.v43i2.455.
- Anderson, Pamela G., Hannah Tarder-Stoll, Mehmet Alpaslan, Nora Keathley, David L. Levin, Srivas Venkatesh, Elliot Bartel, Serge Sicular, Scott Howell, Robert V. Lindsey, and Rebecca M. Jones. 2024. "Deep Learning Improves Physician Accuracy in the Comprehensive Detection of Abnormalities on Chest X-Rays." *Scientific Reports* 14(1):25151. doi:10.1038/s41598-024-76608-2.

- Archana, R., and P. S. Eliahim Jeevaraj. 2024. "Deep Learning Models for Digital Image Processing: A Review." *Artificial Intelligence Review* 57(1):11. doi:10.1007/s10462-023-10631-z.
- Armitage, Hanae. 2018. "Artificial Intelligence Rivals Radiologists in Screening X-Rays for Certain Diseases." <https://med.stanford.edu/news/all-news/2018/11/ai-outperformed-radiologists-in-screening-x-rays-for-certain-diseases.html>.
- Bhatta, Saurav. 2023. "Empowering Rural HealthCare: Mobilenet- Driven Learning for Early Diabetic Retinopathy Detection in Nepal." 5(4).
- Blantz, Eric. 2010. "4 Key Challenges and Solutions to ICT Deployments for Rural Healthcare." <https://www.ictworks.org/4-key-challenges-and-solutions-ict-deployments-rural-healthcare/>.
- Bomanji, Jamshed B., Narainder Gupta, Parveen Gulati, and Chandan J. Das. 2015. "Imaging in TuberculosisImaging in Tuberculosis." *Cold Spring Harbor Perspectives in Medicine* 5(6):a017814. doi:10.1101/cshperspect.a017814.
- Bradley, Stephen H., Sarah Abraham, Matthew Ej Callister, Adam Grice, William T. Hamilton, Rocio Rodriguez Lopez, Bethany Shinkins, and Richard D. Neal. 2019. "Sensitivity of Chest X-Ray for Detecting Lung Cancer in People Presenting with Symptoms: A Systematic Review." *The British Journal of General Practice: The Journal of the Royal College of General Practitioners* 69(689):e827–35. doi:10.3399/bjgp19X706853.
- Brakohiapa, Edmund Kwadwo Kwakye, Benard Ohene Botwe, Benjamin Dabo Sarkodie, Eric Kwesi Ofori, and Jerry Coleman. 2017. "Radiographic Determination of Cardiomegaly Using Cardiothoracic Ratio and Transverse Cardiac Diameter: Can One Size Fit All? Part One." *The Pan African Medical Journal* 27:201. doi:10.11604/pamj.2017.27.201.12017.
- Breve, Fabricio Aparecido. 2022. "COVID-19 Detection on Chest X-Ray Images: A Comparison of CNN Architectures and Ensembles." *Expert Systems with Applications* 204:117549. doi:10.1016/j.eswa.2022.117549.
- Bursa, Sevda Özge, Özlem Durmaz İncel, and Gülfem Işıklar Alptekin. 2023. "Building Lightweight Deep Learning Models with TensorFlow Lite for Human Activity Recognition on Mobile Devices." *Annals of Telecommunications* 78(11):687–702. doi:10.1007/s12243-023-00962-x.
- cancer.org. 2024. "Lung Cancer Early Detection, Diagnosis, and Staging." <https://www.cancer.org/content/dam/CRC/PDF/Public/8705.00.pdf>.

- Cheng, Ting, Huanying Wan, Qijian Cheng, Yi Guo, Yanrong Qian, Liang Fan, Yun Feng, Yanyan Song, Min Zhou, Qingyun Li, Guochao Shi, and Shaoguang Huang. 2016. "Computed Tomography Manifestation of Acute Exacerbation of Chronic Obstructive Pulmonary Disease: A Pilot Study." *Experimental and Therapeutic Medicine* 11(2):519–29. doi:10.3892/etm.2015.2930.
- Christen, Peter, David J. Hand, and Nishadi Kirielle. 2023. "A Review of the F-Measure: Its History, Properties, Criticism, and Alternatives." *ACM Comput. Surv.* 56(3):73:1-73:24. doi:10.1145/3606367.
- clevelandclinic. 2022. "Enlarged Heart (Cardiomegaly): What It Is, Symptoms & Treatment." <https://my.clevelandclinic.org/health/diseases/21490-enlarged-heart-cardiomegaly>.
- clevelandclinic. 2023. "Pneumothorax (Collapsed Lung)." <https://my.clevelandclinic.org/health/diseases/15304-collapsed-lung-pneumothorax>.
- Cundrle, Ivan, Lyle J. Olson, and Bruce D. Johnson. 2019. "Pulmonary Limitations in Heart Failure." *Clinics in Chest Medicine* 40(2):439–48. doi:10.1016/j.ccm.2019.02.010.
- Davila, Ana, Jacinto Colan, and Yasuhisa Hasegawa. 2024. "Comparison of Fine-Tuning Strategies for Transfer Learning in Medical Image Classification." *Image and Vision Computing* 146:105012. doi:10.1016/j.imavis.2024.105012.
- Dayne, Suzanna. 2021. "Protecting Children from the Most-Deadly Infectious Disease in Indonesia | UNICEF Indonesia." <https://www.unicef.org/indonesia/health/stories/protecting-children-most-deadly-infectious-disease-indonesia>.
- Du, Ke-Lin, and M. N. S. Swamy. 2019. "Multilayer Perceptrons: Architecture and Error Backpropagation." Pp. 97–141 in *Neural Networks and Statistical Learning*, edited by K.-L. Du and M. N. S. Swamy. London: Springer.
- Dubey, Shiv Ram, Satish Kumar Singh, and Bidyut Baran Chaudhuri. 2022. "Activation Functions in Deep Learning: A Comprehensive Survey and Benchmark."
- Dumoulin, Vincent, and Francesco Visin. 2018. "A Guide to Convolution Arithmetic for Deep Learning."
- Eisen, Lewis A., Jeffrey S. Berger, Abhijith Hegde, and Roslyn F. Schneider. 2006. "Competency in Chest Radiography." *Journal of General Internal Medicine* 21(5):460–65. doi:10.1111/j.1525-1497.2006.00427.x.
- Elharrouss, Omar, Younes Akbari, Noor Almaadeed, and Somaya Al-Maadeed. 2022. "Backbones-Review: Feature Extraction Networks for Deep Learning and Deep

- Reinforcement Learning Approaches.” *arXiv.Org*. doi:10.1016/j.cosrev.2024.100645.
- Gaillard, Frank. 2024. “Consolidation | Radiology Reference Article | Radiopaedia.Org.” https://radiopaedia.org/articles/consolidation?lang=us&utm_source=chatgpt.com.
- Gakhar, Manav, and Apeksha Aggarwal. 2022. “ThoraciNet: Thoracic Abnormality Detection and Disease Classification Using Fusion DCNNs.” *Physical and Engineering Sciences in Medicine* 45(3):961–70. doi:10.1007/s13246-022-01137-z.
- Gerdes, Anne. 2024. “The Role of Explainability in AI-Supported Medical Decision-Making.” *Discover Artificial Intelligence* 4(1):29. doi:10.1007/s44163-024-00119-2.
- Gu, Chanhoe, and Minhyeok Lee. 2024. “Deep Transfer Learning Using Real-World Image Features for Medical Image Classification, with a Case Study on Pneumonia X-Ray Images.” *Bioengineering* 11(4):406. doi:10.3390/bioengineering11040406.
- Guo, Lili, Changsheng Zhou, Jingxu Xu, Chencui Huang, Yizhou Yu, and Guangming Lu. 2024. “Deep Learning for Chest X-Ray Diagnosis: Competition Between Radiologists with or Without Artificial Intelligence Assistance.” *Journal of Imaging Informatics in Medicine* 37(3):922–34. doi:10.1007/s10278-024-00990-6.
- Howard, Andrew G., Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.”
- Inui, Shohei, Wataru Gono, Ryo Kurokawa, Yudai Nakai, Yusuke Watanabe, Keita Sakurai, Masanori Ishida, Akira Fujikawa, and Osamu Abe. 2021. “The Role of Chest Imaging in the Diagnosis, Management, and Monitoring of Coronavirus Disease 2019 (COVID-19).” *Insights into Imaging* 12(1):155. doi:10.1186/s13244-021-01096-1.
- Iqbal, Hammad, Arshad Khan, Narayan Nepal, Faheem Khan, and Yeon-Kug Moon. 2024. “Deep Learning Approaches for Chest Radiograph Interpretation: A Systematic Review.” *Electronics* 13(23):4688. doi:10.3390/electronics13234688.
- Jeong, Yeon Joo, and Kyung Soo Lee. 2008. “Pulmonary Tuberculosis: Up-to-Date Imaging and Management.” *American Journal of Roentgenology* 191(3):834–44. doi:10.2214/AJR.07.3896.
- Jewell, Tim. 2023. “Chest X-Rays in Tuberculosis Diagnosis.” <https://www.healthline.com/health/tuberculosis-x-ray>.

- Juba, Brendan, and Hai S. Le. 2019. "Precision-Recall versus Accuracy and the Role of Large Data Sets." *Proceedings of the AAAI Conference on Artificial Intelligence* 33(01):4039–48. doi:10.1609/aaai.v33i01.33014039.
- Kansal, Kajal, Tej Bahadur Chandra, and Akansha Singh. 2024. "ResNet-50 vs. EfficientNet-B0: Multi-Centric Classification of Various Lung Abnormalities Using Deep Learning." *Procedia Computer Science* 235:70–80. doi:10.1016/j.procs.2024.04.007.
- Keskar, Nitish Shirish, Dheevatsa Mudigere, Jorge Nocedal, Mikhail Smelyanskiy, and Ping Tak Peter Tang. 2017. "On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima."
- King, Paul T. 2015. "Inflammation in Chronic Obstructive Pulmonary Disease and Its Role in Cardiovascular Disease and Lung Cancer." *Clinical and Translational Medicine* 4:26. doi:10.1186/s40169-015-0068-z.
- Kingma, Diederik P., and Jimmy Ba. 2017. "Adam: A Method for Stochastic Optimization."
- Kolonne, Shammi, Chamodi Fernando, Hashara Kumarasinghe, and Dulani Meedeniya. 2021. "MobileNetV2 Based Chest X-Rays Classification." Pp. 57–61 in *2021 International Conference on Decision Aid Sciences and Application (DASA)*. Sakheer, Bahrain: IEEE.
- Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. 2012. "ImageNet Classification with Deep Convolutional Neural Networks." in *Advances in Neural Information Processing Systems*. Vol. 25. Curran Associates, Inc.
- Kshatri, Sapna Singh, and Deepak Singh. 2023. "Convolutional Neural Network in Medical Image Analysis: A Review." *Archives of Computational Methods in Engineering* 30(4):2793–2810. doi:10.1007/s11831-023-09898-w.
- Kufel, Jakub, Michał Bielówka, Marcin Rojek, Adam Mitreğa, Piotr Lewandowski, Maciej Cebula, Dariusz Krawczyk, Marta Bielówka, Dominika Kondoł, Katarzyna Bargieł-Łączek, Iga Paszkiewicz, Łukasz Czogalik, Dominika Kaczyńska, Aleksandra Wocław, Katarzyna Gruszczyńska, and Zbigniew Nawrat. 2023. "Multi-Label Classification of Chest X-Ray Abnormalities Using Transfer Learning Techniques." *Journal of Personalized Medicine* 13(10):1426. doi:10.3390/jpm13101426.
- Kurmukov, Anvar, Valeria Chernina, Regina Gareeva, Maria Dugova, Ekaterina Petrash, Olga Aleshina, Maxim Pisov, Boris Shirokikh, Valentin Samokhin, Vladislav Proskurov, Stanislav Shimovolos, Maria Basova, Mikhail Goncahrov, Eugenia Soboleva, Maria Donskova, Farukh Yaushev, Alexey Shevtsov, Alexey Zakharov,

- Talgat Saparov, Victor Gombolevskiy, and Mikhail Belyaev. 2024. “The Impact of Deep Learning Aid on the Workload and Interpretation Accuracy of Radiologists on Chest Computed Tomography: A Cross-over Reader Study.”
- Kvak, Daniel, Anna Chromcová, Marek Bíroš, Robert Hrubý, Karolína Kvaková, Marija Pajdaković, and Petra Ovesná. 2023. “Chest X-Ray Abnormality Detection by Using Artificial Intelligence: A Single-Site Retrospective Study of Deep Learning Model Performance.” *BioMedInformatics* 3(1):82–101. doi:10.3390/biomedinformatics3010006.
- Le, Nhung, Benjamin Männel, Mihaela Jarema, Thach Thanh Luong, Luyen K. Bui, Hai Quoc Vy, and Harald Schuh. 2024. “K-Fold Cross-Validation: An Effective Hyperparameter Tuning Technique in Machine Learning on GNSS Time Series for Movement Forecast.” Pp. 377–82 in *Recent Research on Geotechnical Engineering, Remote Sensing, Geophysics and Earthquake Seismology, Advances in Science, Technology & Innovation*, edited by A. Çiner, Z. A. Ergüler, M. Bezzeghoud, M. Ustuner, M. Eshagh, H. El-Askary, A. Biswas, L. Gasperini, K.-G. Hinzen, M. Karakus, C. Comina, A. Karrech, A. Polonia, and H. I. Chaminé. Cham: Springer Nature Switzerland.
- Lee, Kyung Soo, Joungho Han, Man Pyo Chung, and Yeon Joo Jeong. 2013. “Consolidation.” *Radiology Illustrated: Chest Radiology* 221–33. doi:10.1007/978-3-642-37096-0_22.
- van Leeuwen, Kicky G., Maarten de Rooij, Steven Schalekamp, Bram van Ginneken, and Matthieu J. C. M. Rutten. 2022. “How Does Artificial Intelligence in Radiology Improve Efficiency and Health Outcomes?” *Pediatric Radiology* 52(11):2087–93. doi:10.1007/s00247-021-05114-8.
- Li, Hang. 2018. “Deep Learning for Natural Language Processing: Advantages and Challenges.” *National Science Review* 5(1):24–26. doi:10.1093/nsr/nwx110.
- Lin, Min, Qiang Chen, and Shuicheng Yan. 2014. “Network In Network.”
- Litjens, Geert, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A. W. M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. 2017. “A Survey on Deep Learning in Medical Image Analysis.” *Medical Image Analysis* 42:60–88. doi:10.1016/j.media.2017.07.005.

- Liu, Hou-I., Marco Galindo, Hongxia Xie, Lai-Kuan Wong, Hong-Han Shuai, Yung-Hui Li, and Wen-Huang Cheng. 2024. "Lightweight Deep Learning for Resource-Constrained Environments: A Survey."
- Ma, Ningning, Xiangyu Zhang, Hai-Tao Zheng, and Jian Sun. 2018. "ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design."
- Mahmoud, Maha, Ahmed Arafa, Shaimaa Abdelsattar Mohammad, and George Ezzat Elkess Yacoub. 2023. "Role of Artificial Intelligence in CXR Interpretation in Pediatric Intensive Care Unit Ain Shams University." *QJM: An International Journal of Medicine* 116(Supplement_1):hcad069.723. doi:10.1093/qjmed/hcad069.723.
- McKnight, Catherine L., and Bracken Burns. 2025. "Pneumothorax." in *StatPearls*. Treasure Island (FL): StatPearls Publishing.
- Melo, Francisco. 2013. "Area under the ROC Curve." Pp. 38–39 in *Encyclopedia of Systems Biology*, edited by W. Dubitzky, O. Wolkenhauer, K.-H. Cho, and H. Yokota. New York, NY: Springer.
- Metkus, Thomas S. 2024. "Pulmonary Edema: MedlinePlus Medical Encyclopedia." <https://medlineplus.gov/ency/article/000140.htm>.
- Morton, Will. 2023. "AI Could Save Radiologists Chest X-Ray Interpretation Time." <https://www.auntminnie.com/imaging-informatics/artificial-intelligence/article/15633357/ai-could-save-radiologists-chest-xray-interpretation-time>.
- Muharram, Farizal Rizky, Chaq El Chaq Zamzam Multazam, Ali Mustofa, Wigaviola Socha, Andrianto, Santi Martini, Leopold Aminde, and Chung Yi-Li. 2024. "The 30 Years of Shifting in The Indonesian Cardiovascular Burden—Analysis of The Global Burden of Disease Study." *Journal of Epidemiology and Global Health* 14(1):193–212. doi:10.1007/s44197-024-00187-8.
- Mzoughi, Hiba, Ines Njeh, Mohamed Ben Slima, and Ahmed BenHamida. 2023. "Deep Efficient-Nets with Transfer Learning Assisted Detection of COVID-19 Using Chest X-Ray Radiology Imaging." *Multimedia Tools and Applications* 82(25):39303–25. doi:10.1007/s11042-023-15097-3.
- Nabulsi, Zaid. 2021. "Detecting Abnormal Chest X-Rays Using Deep Learning." <http://research.google/blog/detecting-abnormal-chest-x-rays-using-deep-learning/>.
- Naidu, Gireen, Tranos Zuva, and Elias Mmbongeni Sibanda. 2023. "A Review of Evaluation Metrics in Machine Learning Algorithms." Pp. 15–25 in *Artificial Intelligence*

- Application in Networks and Systems*, edited by R. Silhavy and P. Silhavy. Cham: Springer International Publishing.
- Nam, Ju Gang, Minchul Kim, Jongchan Park, Eui Jin Hwang, Jong Hyuk Lee, Jung Hee Hong, Jin Mo Goo, and Chang Min Park. 2021. "Development and Validation of a Deep Learning Algorithm Detecting 10 Common Abnormalities on Chest Radiographs." *European Respiratory Journal* 57(5):2003061. doi:10.1183/13993003.03061-2020.
- nela. 2023. "Indonesia Faces Shortage of Specialized Doctors, Hindering Healthcare Services." <https://www.theonlinecitizen.com/2023/07/05/indonesia-faces-shortage-of-specialized-doctors-hindering-healthcare-services/>.
- O'Shea, Keiron, and Ryan Nash. 2015. "An Introduction to Convolutional Neural Networks."
- Päpper, Marc. 2021. "Depthwise Separable Convolutions in PyTorch :: Päpper's Machine Learning Blog — This Blog Features State of the Art Applications in Machine Learning with a Lot of PyTorch Samples and Deep Learning Code. You Will Learn about Neural Network Optimization and Potential Insights for Artificial Intelligence for Example in the Medical Domain." <https://www.paepper.com/blog/posts/depthwise-separable-convolutions-in-pytorch/>.
- Patrício, Cristiano, João C. Neves, and Luís F. Teixeira. 2022. "Explainable Deep Learning Methods in Medical Diagnosis: A Survey." *arXiv.Org*. <https://arxiv.org/abs/2205.04766v1>.
- Pochepnia, Svitlana, Elzbieta Magdalena Grabczak, Emma Johnson, Fusun Oner Eyuboglu, Onno Akkerman, and Helmut Prosch. 2024. "Imaging in Pulmonary Infections of Immunocompetent Adult Patients." *Breathe* 20(1). doi:10.1183/20734735.0186-2023.
- PricewaterhouseCoopers. 2023. "How Can Technology Accelerate the Digitisation of the Indonesian Healthcare Sector?" <https://www.pwc.com/id/en/media-centre/press-release/2023/english/how-can-technology-accelerate-the-digitisation-of-the-indonesian-healthcare-sector.html>.
- Purwins, Hendrik, Bo Li, Tuomas Virtanen, Jan Schlüter, Shuo-yiin Chang, and Tara Sainath. 2019. "Deep Learning for Audio Signal Processing." *arXiv.Org*. doi:10.1109/JSTSP.2019.2908700.

- Rafferty, Jim. 2005. "Qualifications for Interpreting/Classifying Chest Roentgenograms and Maintenance of Interpretation Forms. | Occupational Safety and Health Administration." https://www.osha.gov/laws-regs/standardinterpretations/2005-11-30-0?utm_source=chatgpt.com.
- Rajaraman, Sivaramakrishnan, Incheol Kim, and Sameer K. Antani. 2020. "Detection and Visualization of Abnormality in Chest Radiographs Using Modality-Specific Convolutional Neural Network Ensembles." *PeerJ* 8:e8693. doi:10.7717/peerj.8693.
- Rajpurkar, Pranav, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, Matthew P. Lungren, and Andrew Y. Ng. 2017. "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning."
- Rangelov, Bojidar A., Alexandra L. Young, Joseph Jacob, Anthony P. Cahn, Sarah Lee, Frederick J. Wilson, David J. Hawkes, and John R. Hurst. 2020. "Thoracic Imaging at Exacerbation of Chronic Obstructive Pulmonary Disease: A Systematic Review." *International Journal of Chronic Obstructive Pulmonary Disease* 15:1751–87. doi:10.2147/COPD.S250746.
- Refaeilzadeh, Payam, Lei Tang, and Huan Liu. 2009. "Cross-Validation." Pp. 532–38 in *Encyclopedia of Database Systems*, edited by L. Liu and M. T. Özsu. Boston, MA: Springer US.
- Ridley, Lloyd. 2019. "Chest Radiograph Signs Suggestive of Pericardial Disease." https://www.acc.org/Latest-in-Cardiology/Articles/2019/09/09/10/46/http%3a%2f%2fwww.acc.org%2fLatest-in-Cardiology%2fArticles%2f2019%2f09%2f10%2f46%2fChest-Radiograph-Signs-Suggestive-of-Pericardial-Disease%3futm_source%3dchatgpt.com.
- Roboflow. 2022. "Chest Abnormalities > Browse." <https://universe.roboflow.com/xray-annotations/chest-abnormalities>.
- Rosenbaum, Robert. 2022. "On the Relationship between Predictive Coding and Backpropagation." *PLOS ONE* 17(3):e0266102. doi:10.1371/journal.pone.0266102.
- Rybczak, Monika, and Krystian Kozakiewicz. 2024. "Deep Machine Learning of MobileNet, Efficient, and Inception Models." *Algorithms* 17(3):96. doi:10.3390/a17030096.
- Sahin, Emre, Azer Celikten, Semih Demirel, Andac Akpulat, Kadir Budak, and Hakan Karatas. 2024. "Pneumonia Detection in Chest X - Ray Images Using Deep Learning

- and Grad-CAM++ Visualization.” Pp. 1–5 in *2024 Innovations in Intelligent Systems and Applications Conference (ASYU)*. Ankara, Turkiye: IEEE.
- Sandler, Mark, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. 2018. “MobileNetV2: Inverted Residuals and Linear Bottlenecks.” Pp. 4510–20 in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT: IEEE.
- Sarkar, Shriharsha. 2021. “Overcoming Hospital Challenges with Enhanced Technology Solutions in Indonesia.”
- Sarvamangala, D. R., and Raghavendra V. Kulkarni. 2022. “Convolutional Neural Networks in Medical Image Understanding: A Survey.” *Evolutionary Intelligence* 15(1):1–22. doi:10.1007/s12065-020-00540-3.
- Selvaraju, Ramprasaath R., Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2020. “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization.” *International Journal of Computer Vision* 128(2):336–59. doi:10.1007/s11263-019-01228-7.
- Shah Gupta, Rikisha, Ardita Koteci, Ann Morgan, Peter M. George, and Jennifer K. Quint. 2023. “Incidence and Prevalence of Interstitial Lung Diseases Worldwide: A Systematic Literature Review.” *BMJ Open Respiratory Research* 10(1):e001291. doi:10.1136/bmjresp-2022-001291.
- Shiri, Farhad Morteza pour, Thinagaran Perumal, Norwati Mustapha, and Raihani Mohamed. 2024. “A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU.” doi:https://doi.org/10.32604/jai.2024.054314.
- Smithuis, Robin, and Otto van Delden. 2022. “The Radiology Assistant : Chest X-Ray - Basic Interpretation.” <https://radiologyassistant.nl/chest/chest-x-ray/basic-interpretation>.
- Statista. 2024. “Indonesia: COPD Projection 2017-2024.” <https://www.statista.com/statistics/1052634/indonesia-copd-projection/>.
- Suara, Subhashis, Aayush Jha, Pratik Sinha, and Arif Ahmed Sekh. 2024. “Is Grad-CAM Explainable in Medical Images?” Pp. 124–35 in Vol. 2009.
- Susanto, Prastyo Eko, Arrie Kurniawardhan, Dhomas Hatta Fudholi, and Ridho Rahmadi. 2022. “A Mobile Deep Learning Model on Covid-19 CT-Scan Classification.” *International Journal of Artificial Intelligence Research* 6(2). doi:10.29099/ijair.v6i1.257.

- Taha, Bahauddin, Fahmida Rahman Liza, Md Moniruzzaman, and Md Samsuzzaman. 2022. "Comparative Analysis of Fine-Tuned MobileNet Versions to Detect Brain Tumors from MRI Images." Pp. 1–6 in *2022 4th International Conference on Sustainable Technologies for Industry 4.0 (STI)*. Dhaka, Bangladesh: IEEE.
- Talaei Khoei, Tala, Hadjar Ould Slimane, and Naima Kaabouch. 2023. "Deep Learning: Systematic Review, Models, Challenges, and Research Directions." *Neural Computing and Applications* 35(31):23103–24. doi:10.1007/s00521-023-08957-4.
- Tan, Mingxing, and Quoc V. Le. 2020. "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks."
- Terven, Juan, Diana M. Cordova-Esparza, Alfonso Ramirez-Pedraza, Edgar A. Chavez-Urbiola, and Julio A. Romero-Gonzalez. 2024. "Loss Functions and Metrics in Deep Learning."
- Tian, Yingjie, Yuqi Zhang, and Haibin Zhang. 2023. "Recent Advances in Stochastic Gradient Descent in Deep Learning." *Mathematics* 11(3):682. doi:10.3390/math11030682.
- Tobiasz, R., G. Wilczyński, P. Graszka, N. Czechowski, and S. Łuczak. 2023. "Edge Devices Inference Performance Comparison." *Journal of Computing Science and Engineering* 17(2):51–59. doi:10.5626/JCSE.2023.17.2.51.
- Understanding Convolutional Neural Network: A Complete Guide. 2023. <https://learnopencv.com/understanding-convolutional-neural-networks-cnn/>.
- utswmed. 2025. "Lung Nodules | Condition | UT Southwestern Medical Center." <http://utswmed.org/conditions-treatments/lung-nodules/>.
- Velu, Shubashini. 2023. "An Efficient, Lightweight MobileNetV2-Based Fine-Tuned Model for COVID-19 Detection Using Chest X-Ray Images." *Mathematical Biosciences and Engineering* 20(5):8400–8427. doi:10.3934/mbe.2023368.
- Wang, Haohan, and Bhiksha Raj. 2017. "On the Origin of Deep Learning."
- Wang, Xiaosong, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M. Summers. 2017. "ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases." Pp. 3462–71 in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Wulan, Dinda Rakhma. 2024. "Kegiatan Puncak Hari Tuberkulosis Sedunia 2024: Gerakan Indonesia Akhiri Tuberkulosis." <https://tbindonesia.or.id/puncak-htbs-2024/>.

- Yunus, Reyhan Eddy. 2024. "Radiology Loading and Coverage Hours in Indonesia." *Korean Journal of Radiology* 25(7):597–99. doi:10.3348/kjr.2024.0267.
- Zhan, Qishi, Dan Sun, Erdi Gao, Yuhan Ma, Yaxin Liang, and Haowei Yang. 2024. "Advancements in Feature Extraction Recognition of Medical Imaging Systems Through Deep Learning Technique."
- Zhang, Chaoning, Philipp Benz, Dawit Mureja Argaw, Seokju Lee, Junsik Kim, Francois Rameau, Jean-Charles Bazin, and In So Kweon. 2020. "ResNet or DenseNet? Introducing Dense Shortcuts to ResNet."
- Zhang, Huanhuan, and Yufei Qie. 2023. "Applying Deep Learning to Medical Imaging: A Review." *Applied Sciences* 13(18):10521. doi:10.3390/app131810521.
- Zhang, Xiangyu, Xinyu Zhou, Mengxiao Lin, and Jian Sun. 2017a. "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices."
- Zhang, Xiangyu, Xinyu Zhou, Mengxiao Lin, and Jian Sun. 2017b. "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices."
- Zhao, Xia, Limin Wang, Yufei Zhang, Xuming Han, Muhammet Deveci, and Milan Parmar. 2024. "A Review of Convolutional Neural Networks in Computer Vision." *Artificial Intelligence Review* 57(4):99. doi:10.1007/s10462-024-10721-6.