

Optimasi *Hyperparameter Deep Learning* untuk Deteksi X-Ray Paru-Paru Menggunakan *Bayesian Optimization*

Fayza Shahira¹, Benny Sukma Negara², Febi Yanto³, Suwanto Sanjaya⁴

^{1,4} Program Studi Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau

¹12150121579@students.uin-suska.ac.id

²bsnegara@uin-suska.ac.id

³febiyanto@uin-suska.ac.id

⁴suwantosanjaya@uin-suska.ac.id

Abstrak— Penyakit paru-paru, seperti pneumonia dan COVID-19, merupakan ancaman serius terhadap kesehatan masyarakat, terutama jika diagnosisnya mengalami keterlambatan. Pendekatan deteksi dini melalui citra X-ray dada banyak digunakan, namun akurasi sangat bergantung pada kemampuan sistem klasifikasi. Penelitian ini bertujuan untuk meningkatkan performa klasifikasi citra X-ray paru-paru dengan mengimplementasikan metode *deep learning* menggunakan arsitektur ResNet-101 yang dioptimasi menggunakan teknik *Bayesian Optimization*. Dataset yang digunakan dalam penelitian ini terdiri dari tiga kelas yaitu Normal, Pneumonia, dan COVID-19, masing-masing sejumlah 1.000 citra. Kinerja model hasil optimasi dibandingkan dengan model *baseline* pada tiga skenario *split data* yaitu 90:10, 80:20, 70:30. Hasil penelitian mengindikasikan bahwa model yang telah dioptimasi mampu meningkatkan performa pada seluruh metrik evaluasi mencakup akurasi, presisi, recall, spesifisitas, dan *F1-score*. Akurasi tertinggi tercatat sebesar 93,83% pada skenario 80:20, melampaui akurasi *baseline* yang sebesar 91,83. Selain itu, kurva akurasi dan *loss* menunjukkan proses *training* yang stabil dan konvergen secara cepat tanpa indikasi *overfitting* yang signifikan. Penerapan *Bayesian Optimization* terbukti efektif dalam menemukan konfigurasi *hyperparameter* optimal yang berdampak pada peningkatan dalam tiap metrik evaluasi.

Kata Kunci— Deep Learning, Bayesian Optimization, Citra X-Ray Dada, Optimasi Hyperparameter, Deteksi Penyakit Paru-Paru

I. PENDAHULUAN

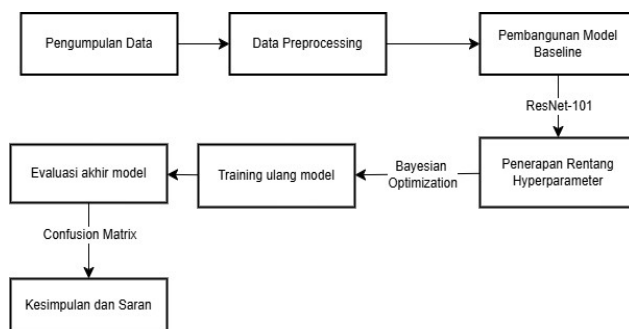
Paru-paru adalah organ vital dalam sistem pernapasan manusia yang berperan dalam proses pertukaran oksigen dan karbondioksida di dalam darah. Memburuknya kualitas udara di berbagai belahan dunia telah meningkatkan risiko terjadinya penyakit paru-paru. Salah satunya adalah paru-paru basah atau pneumonia. Penyakit ini merupakan penyebab utama kematian akibat infeksi pada anak-anak di seluruh dunia [1]. Selain pneumonia, pada Desember 2019, di kota Wuhan Provinsi Hubei, Cina, telah ditemukan pertama kali wabah penyakit coronavirus disease 2019 atau yang lebih dikenal sebagai Covid-19. Kasus positif dan angka kematian akibat Covid-19 meningkat dengan cepat setiap harinya, menyebar secara merata ke berbagai belahan dunia [2]. Karena itu, pemeriksaan dan diagnosis yang cepat dapat membantu mengurangi dampak penyakit yang dapat mengancam jiwa dan meningkatkan kualitas hidup pasien. Salah satu cara yang dapat dilakukan adalah menggunakan rontgen dada (chest x-ray/chest radiograph) [3]. Namun meskipun sering digunakan, pembacaan foto rontgen dada

(chest radiograph) memiliki keterbatasan, seperti kesulitan dalam mendeteksi penyakit, yang menyebabkan waktu yang lama bagi pasien atau tenaga medis untuk mendiagnosis kondisi yang diderita. Salah satu solusi untuk mengatasi masalah ini adalah dengan mengklasifikasikan citra rontgen dada ke dalam kelas tertentu menggunakan teknik machine learning [4]. Klasifikasi citra x-ray dada menggunakan machine learning sendiri sudah pernah dilakukan oleh (Fitriyarsari, 2022) [5] yang mengimplementasikan algoritma backpropagation sebagai classifier. Dalam penelitian tersebut, penulis membagi citra x-ray dada menjadi 154 citra data latih, dan 41 citra data uji. Penelitian ini memperoleh akurasi tertinggi sebesar 87,8%. Selain machine learning, deteksi penyakit pada citra x-ray dada juga dapat dilakukan menggunakan deep learning. Penelitian deteksi penyakit pneumonia dan COVID-19 pada citra x-ray dada menggunakan deep learning pernah dilakukan oleh (Yudistira., 2020) [6] yang membandingkan arsitektur-arsitektur deep learning, dalam penelitian tersebut, arsitektur ResNet50 mencapai akurasi tertinggi yaitu sebesar 90,8%, dengan rincian akurasi COVID-19 mencapai 95,17%, paru-paru normal 95,72% dan paru-paru pneumonia 81,5%. Penelitian-penelitian dalam deep learning sendiri masih memiliki tantangan, terutama dalam memilih arsitektur dan hyperparameter yang tepat untuk mendapatkan akurasi yang maksimal. Meskipun *deep learning* telah menunjukkan potensi yang baik dalam analisis, pemilihan hyperparameter dalam suatu model arsitektur deep learning masih dilakukan secara manual. Kelemahan ini dapat dikompensasi dengan teknik *hyperparameter tuning* berbasis algoritma *bayesian optimization* dalam membantu memilih hyperparameter yang tepat untuk meningkatkan performa model [7]. Salah satu penelitian mengenai hyperparameter tuning dilakukan oleh (Mae, K., 2024) [8]. yang membandingkan akurasi suatu model ketika sebelum dan sesudah diterapkan teknik hyperparameter tuning memperoleh hasil bahwa modelnya yang menggunakan metode random forest mencapai akurasi 98.16%, dan setelah diterapkan metode grid search mencapai 98.22%, sedangkan ketika penerapan Bayesian optimization meningkat menjadi 98.34%. penelitian yang membahas tentang peningkatan akurasi menggunakan Bayesian optimization sebagai tuning hyperparameter juga sudah dilakukan. Seperti penelitian oleh (Loey dkk., 2022) [9] dimana metode convolutional neural network (CNN) digunakan untuk mengekstrak fitur. Kemudian dilanjutkan dengan optimasi menggunakan Bayesian optimization untuk mengatur hyperparameter CNN tersebut sehingga akurasi

akhir mencapai angka 96%, akurasi ini diperoleh pada pembagian data dengan 80% data training, 10% data validation, dan 10% data testing. Penelitian yang melibatkan Bayesian optimization dan deep learning juga dilakukan oleh (Miranda dkk., 2021) [10] yang menggunakan metode transfer learning berbasis convolutional neural network (CNN) dan optimasi hyperparameter Bayesian. Penelitian ini berfokus pada eksplorasi otomatis berbagai konfigurasi hyperparameter. Model terbaik yang dihasilkan pada penelitian ini mendapatkan akurasi sebesar 98.67% dengan 300 data, dan akurasi terendah di angka 94.74% dengan data berjumlah 150. Hasil penelitian-penelitian tersebut menunjukkan bahwa optimasi yang terotomatisasi dapat menghasilkan model yang terbaik. Berdasarkan masalah-masalah yang telah diuraikan, maka penelitian ini menggunakan metode deep learning dengan fokus utama pada optimasi hyperparameter menggunakan metode Bayesian optimization yang secara otomatis akan mengeksplorasi aturan-aturan hyperparameter agar meningkatkan performa model deteksi penyakit paru-paru pada citra x-ray dada.

II. METODE

Penelitian ini dilakukan dengan menggunakan dataset citra X-Ray paru-paru yang bersumber dari Mendeley Data yang bersifat open source. Dataset tersebut kemudian diklasifikasikan ke dalam tiga kelas yaitu Normal, Pneumonia, dan COVID-19 menggunakan salah satu arsitektur *deep learning* yaitu ResNet-101 yang dioptimalkan dengan metode *Bayesian Optimization*. Tahapan penelitian meliputi pengumpulan dataset, *preprocessing*, pengembangan model, penerapan *Bayesian optimization*, *training model*, serta evaluasi model. Penelitian ini dilakukan menggunakan bahasa pemrograman *python* pada *google colab* dengan bantuan library TensorFlow dan Keras. Alur penelitian dapat dilihat pada gambar 1.

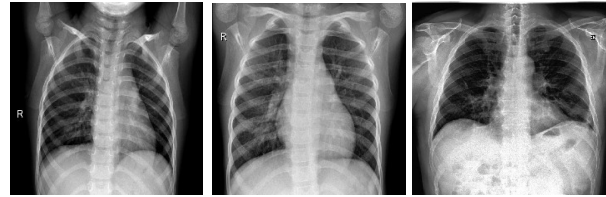


Gbr. 1 Alur Penelitian.

A. Pengumpulan Data

Pada tahap ini dilakukan pengumpulan dataset. Dataset yang digunakan pada penelitian ini bersumber dari Mendeley Data Kumar, Sachin (2022) [11]. Yang berjudul "Covid-19-Pneumonia-Normal Chest X-Ray Images". Dataset ini terdiri dari 3 kelas yaitu Normal, Pneumonia, dan COVID-19. Dengan

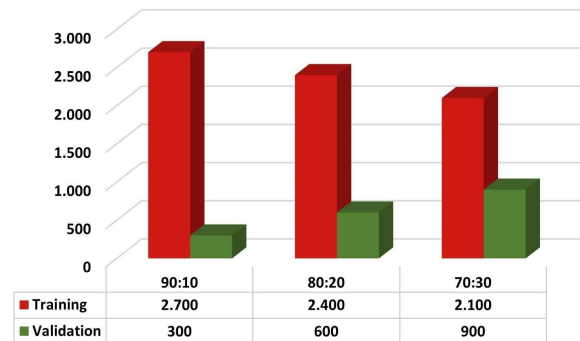
masing-masing kelas yang digunakan dalam penelitian ini berisi 1.000 data citra. Dataset ini digunakan untuk melatih dan menguji model dalam mendeteksi penyakit paru-paru menggunakan teknik deep learning. Contoh dataset yang digunakan pada penelitian disajikan pada gambar 2, berikut:



Gbr. 2 Citra X-ray paru-paru Normal, Pneumonia, COVID-19.

B. Data Preprocessing

Tahap preprocessing dilakukan dengan mengolah gambar x-ray paru-paru agar sesuai dengan model ResNet-101, yaitu dilakukan *resizing* untuk format gambar sehingga menjadi ukuran 224 x 224 piksel [12]. Selanjutnya, data dibagi (split data) menjadi data training dan data validation ke dalam 3 skenario pembagian dengan rasio 90:10, 80:20 dan 70:30 guna mengevaluasi performa model secara luas. Visualisasi distribusi pembagian data *training* dan data *validation* dapat dilihat pada gambar 3.

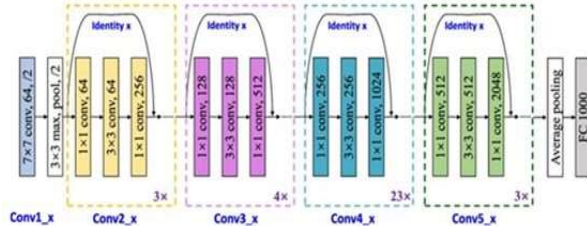


Gbr. 3 Data Splitting.

C. Pembangunan Model Baseline

Model *baseline* dibangun menggunakan arsitektur ResNet-101, yang merupakan salah satu arsitektur *Convolutional Neural Network* (CNN), dioptimalkan menggunakan *Adam Optimizer* dan fungsi aktivasi *LeakyReLU*. Hyperparameter yang digunakan mengikuti *default* dari ResNet-101 dan *Adam Optimizer*, dengan jumlah *epoch* yang ditetapkan oleh penulis sebanyak 50. ResNet-101 merupakan arsitektur yang terdiri dari 101 lapisan dan telah dilatih pada lebih dari satu juta gambar pada *ImageNet*, yang memungkinkan jaringan untuk mengklasifikasikan gambar ke dalam 1.000 kategori objek [13]. Dalam penelitian yang dilakukan oleh Viesrisyah dkk, (2023) [13] yang melakukan perbandingan 5 arsitektur berbeda yaitu *VGG16*, *VGG19*, *ResNet-101*, *ResNet-50*, dan *Xception* untuk mengklasifikasikan citra paru-paru, diperoleh hasil bahwa ResNet-101 menghasilkan kinerja tertinggi dengan akurasi 93.4%, sensitivitas 90,2% dan spesifisitas 95.1%. Berikut

gambar 4 merupakan visualisasi dari arsitektur ResNet-101 [13].



Gbr. 4 Arsitektur ResNet-101.

D. Hyperparameter Tuning

Pengaturan hyperparameter merupakan langkah penting dalam pengembangan model machine learning maupun deep learning untuk mencapai performa yang optimal [13]. Untuk membuat model yang optimal itu sendiri, diperlukan proses untuk menemukan konfigurasi hyperparameter yang optimal untuk digunakan disebut dengan hyperparameter tuning [14]. Hyperparameter merupakan nilai yang harus diterapkan secara manual sebelum tahap pelatihan model. Proses penyesuaian hyperparameter yang disebut optimasi berperan penting dalam meningkatkan kinerja algoritma machine learning. Optimasi hyperparameter bertujuan untuk menemukan kombinasi nilai terbaik untuk mencapai hasil optimal. Proses ini melibatkan eksperimen dengan berbagai konfigurasi hyperparameter serta pelatihan beberapa model untuk menentukan yang paling optimal [15]. Proses pencarian konfigurasi hyperparameter terbaik ini dikenal sebagai Hyperparameter Optimziation (HPO). HPO bertujuan untuk menemukan kombinasi hyperprameter yang menghasilkan performa model terbaik dengan cara yang lebih sistematis dan efisien dibandingkan pencarian manual [16].

1): Penerapan Bayesian Optimization: Bayesian Optimization adalah metode terkemuka dalam tuning hyperparameter, menawarkan strategi global tanpa gradien untuk menemukan nilai optimal dengan jumlah iterasi minimal. Metode ini digunakan untuk secara otomatis menemukan parameter optimal pada CNN. Bayesian Optimization bertugas mengidentifikasi kombinasi parameter terbaik dengan tujuan mengurangi waktu yang diperlukan untuk mencapai model yang akurat [15]. Bayesian Optimization (BO) merupakan algoritma iteratif yang banyak digunakan dalam masalah hyperparameter optimization (HPO), BO menentukan titik evaluasi selanjutnya berdasarkan hasil yang telah diperoleh sebelumnya. BO menyeimbangkan eksplorasi dan eksploitasi untuk mendeteksi area yang paling mungkin menghasilkan konfigurasi optimal, sekaligus menghindari kehilangan solusi yang lebih baik di area yang belum dijelajahi[16].

E. Training Model

Setelah diperoleh kombinasi *hyperparameter* optimal melalui proses *Bayesian Optimization*, tahap selanjutnya adalah melakukan pelatihan model menggunakan arsitektur ResNet-101. Proses *training* dilakukan secara terpisah pada tiga skenario pembagian data, yaitu rasio 90:10, 80:20, dan 70:30, untuk mengevaluasi pengaruh konfigurasi data terhadap performa model. Pada tahap ini, model dilatih menggunakan konfigurasi *hyperparameter* hasil optimasi, seperti *learning rate*, jumlah *epoch*, *batch size*, dan *kernel size*. Selama proses *training*, metrik akurasi dan loss dipantau setiap epoch untuk memastikan model mengalami konvergensi yang stabil tanpa *overfitting*. Pelatihan ini bertujuan untuk membangun model akhir yang mampu mengklasifikasikan citra X-ray paru-paru secara lebih akurat berdasarkan hasil *tuning hyperparameter* sebelumnya.

F. Evaluasi Model

Setelah model dilatih menggunakan data training, model kemudian akan dievaluasi menggunakan data validation untuk mengukur performa selama proses pelatihan. Data testing terdiri dari data citra X-ray paru-paru baru yang tidak terlibat dalam proses pelatihan maupun validasi, dan digunakan untuk menguji kemampuan model dalam mengklasifikasikan gambar ke dalam kelas yang sesuai. Seluruh data untuk training, validation, dan testing adalah dataset yang berbeda, untuk menjaga validitas dan akurasi hasil evaluasi. Pada tahap evaluasi, digunakan confusion matrix untuk mengukur kinerja model. Metrik yang digunakan meliputi accuracy, precision, recall, specificity, F1-score. Nilai dari metrik-metrik tersebut dapat dihitung dengan persamaan berikut[17]:

Accuracy

$$\frac{TP+TN}{TP + TN + FP + FN} \quad (2)$$

Precision

$$\frac{TP}{TP + FP} \quad (3)$$

Recall

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

Specificity

$$\frac{TN}{TN + FP} \quad (4)$$

F1-Score

$$\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

Dimana nilai dari:

True Positive (TP): Data positif yang diprediksi benar oleh model

True Negative (TN): Data negatif yang diprediksi benar oleh model

False Positive (FP): Data negatif yang diprediksi positif

False Negative (FN): Data positif yang diprediksi negative

III. HASIL DAN PEMBAHASAN

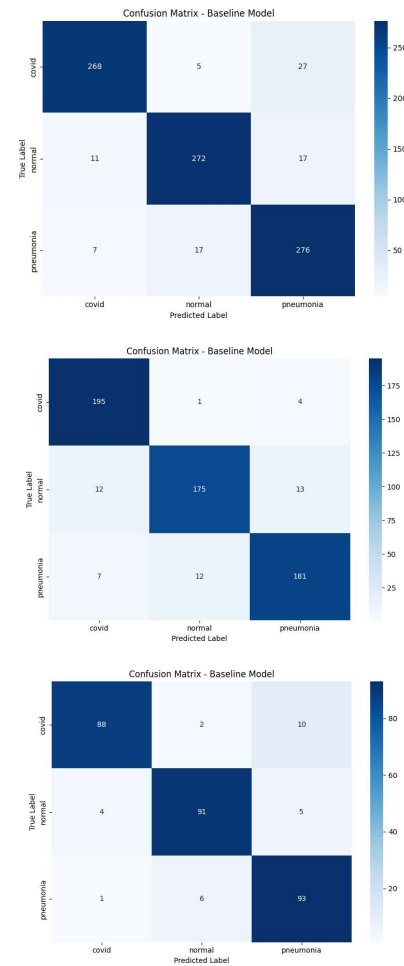
A. Model Baseline

Model *baseline* dibangun menggunakan arsitektur ResNet-101. Model dikonfigurasi dengan *hyperparameter* standar dan dilatih selama 50 *epoch*. Tujuan dari evaluasi model *baseline* adalah untuk membangun titik acuan kinerja sebelum dilakukan proses optimasi. Pengujian dilakukan pada tiga skenario pembagian data yang berbeda untuk melatih dan mengevaluasi model, yaitu rasio 90:10, 80:20, dan 70:30. Rangkuman kinerja model ResNet-101 baseline pada ketiga skenario pembagian data dapat dilihat pada tabel 1.

TABEL I
KINERJA MODEL BASELINE DENGAN RESNET-101

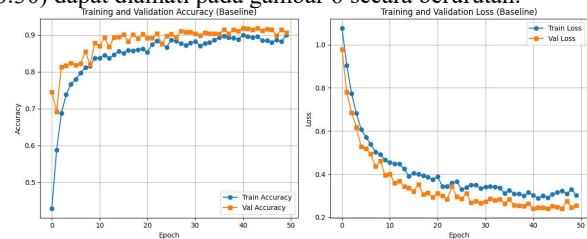
Metrik Evaluasi	90:10	80:20	70:30
<i>Accuracy</i>	90.67%	91.83%	90.67%
<i>Precision</i>	91%	92%	91%
<i>Recall</i>	91%	92%	91%
<i>Spesificity</i>	95.33%	95.92%	95.33%
<i>F1-Score</i>	91%	92%	91%

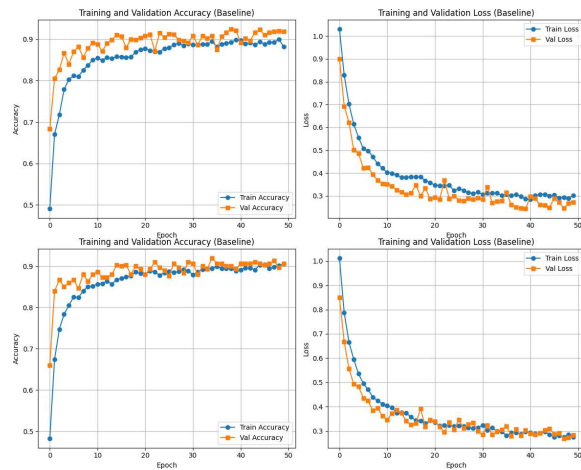
Berdasarkan tabel 1, dapat diamati bahwa kinerja model ResNet-101 baseline menunjukkan variasi yang bergantung pada skenario pembagian data yang digunakan. Skenario 80:20 menghasilkan kinerja terbaik di hampir semua metrik, dengan akurasi mencapai 91.83%, presisi 92%, *recall* 92%, spesifisitas 95.92%, dan F1-Score 92%. Di sisi lain, skenario 90:10 dan 70:30 menunjukkan kinerja yang serupa di sebagian besar metrik. Kedua skenario ini mencatat akurasi sebesar 90.67%, presisi 91%, *recall* 91%, dan F1-Score 91%. Untuk spesifisitas, skenario 90:10 dan 70:30 keduanya mencapai 95.33%. Secara keseluruhan, model baseline ResNet-101 menunjukkan kemampuan klasifikasi yang solid dengan kinerja di atas 90% untuk sebagian besar metrik pada semua skenario pembagian data. Peningkatan kinerja terlihat jelas pada skenario 80:20, yang mengindikasikan bahwa pembagian data 80% untuk pelatihan dan 20% untuk pengujian mungkin merupakan konfigurasi yang paling optimal untuk dataset ini dalam kondisi baseline. Hasil ini akan menjadi dasar penting untuk mengevaluasi efektivitas proses optimasi yang akan dilakukan selanjutnya



Gbr. 5 Confusion Matrix Model Baseline

Untuk memberikan gambaran yang lebih komprehensif mengenai perilaku dan kinerja model baseline selama pelatihan dan pengujian, evaluasi dilakukan tidak hanya melalui *confusion matrix*, tetap juga dengan mengamati grafik kurva akurasi dan *loss*. *confusion matrix* pada figure 5 untuk setiap skenario pembagian data. Serta kurva akurasi dan *loss* selama 50 *epoch* untuk setiap skenario (90:10, 80:20, dan 70:30) dapat diamati pada gambar 6 secara berurutan.





Gbr. 6 Plot Akurasi dan Loss

B. Bayesian Optimization

Setelah evaluasi model baseline, dilakukan proses *hyperparameter tuning* (HPO) untuk meningkatkan kinerja model ResNet-101. dalam penelitian ini, proses *hyperparameter tuning* (HPO) dilakukan menggunakan metode *Bayesian Optimization*. Range pencarian untuk setiap *hyperparameter* yang dioptimasi pada penelitian ini mencakup *learning rate*, *epoch*, *batch size*, dan *kernel size* dengan nilai yang telah didefinisikan berdasarkan penelitian oleh Balseca (2020), [18], yang dituliskan pada tabel 2 berikut:

TABEL 2
RANGE BAYESIAN OPTIMIZATION

Hyperparameter	Range
Learning Rate	0.00001 - 0.0003
Epoch	200 - 250
Batch Size	1 - 3
Kernel Size	1 - 5

Berdasarkan tabel 2, *learning rate* dioptimasi dalam rentang nilai yang digunakan adalah 0.0005 hingga 0.0003. Jumlah *epoch* yang digunakan untuk pelatihan model dalam setiap iterasi optimasi berkisar antara 200 hingga 250. *Batch size* diatur untuk mencari nilai optimal antara 1 hingga 3, dan *kernel size* dari lapisan konvolusi dioptimasi dalam rentang 1 hingga 5. Penentuan rentang ini mengikuti panduan dan hasil yang terbukti efektif dalam penelitian sebelumnya yang dilakukan oleh Balseca, (2020) [18]. Proses bayesian optimization sendiri dilakukan sebanyak 5 kali iterasi oleh penulis, dengan hasil rekomendasi *hyperparameter* yang diberikan oleh *bayesian optimization* dapat dilihat pada tabel 3 berikut:

TABEL 3
HYPERPARAMETER OPTIMAL HASIL BAYESIAN OPTIMIZATION

Hyperparameter	90:10	80:20	70:30
Learning Rate	0.000027	0.000222	0.000287
Epoch	208	248	249
Batch Size	1	2	3

Kernel Size	3	3	3
-------------	---	---	---

Tabel 3 menunjukkan bahwa *Bayesian Optimization* merekomendasikan set *hyperparameter* yang berbeda untuk setiap skenario pembagian data. Misalnya, pada skenario 90:10, *learning rate* optimal adalah 0.000027 sementara pada skenario 80:20, *learning rate* optimalnya adalah 0.000222. Perbedaan ini mengindikasikan bahwa rasio pembagian data dapat memengaruhi konfigurasi *hyperparameter* yang paling efektif untuk mencapai kinerja optimal. Nilai-nilai *hyperparameter* ini kemudian digunakan untuk melatih model ResNet-101 akhir pada masing-masing skenario.

C. Evaluasi Model Akhir

Model ResNet-101 yang telah dilatih ulang menggunakan *hyperparameter* optimal kemudian dievaluasi pada dataset pengujianya masing-masing. Perbandingan kinerja antara model baseline dan model akhir yang telah dioptimasi disajikan pada table 8.

TABEL 4
PERBANDINGAN KINERJA MODEL BASELINE DAN MODEL AKHIR

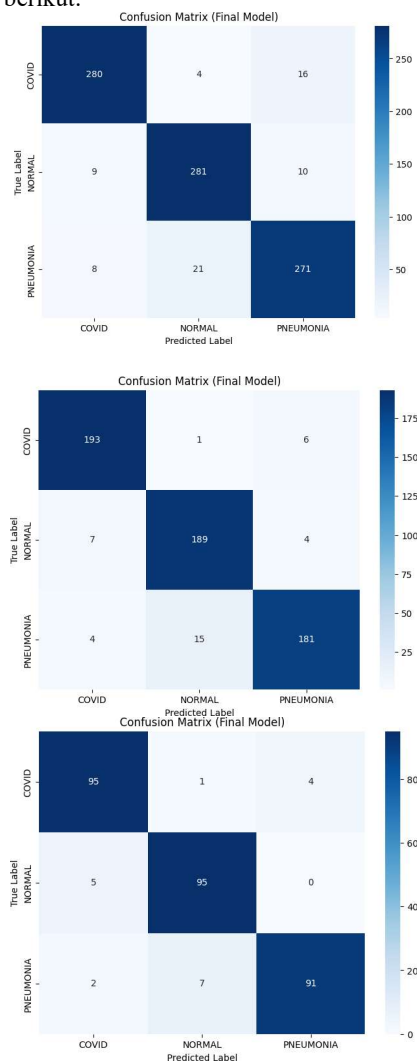
Metrik Evaluasi	Skenario	Model Baseline	Model Akhir	Peningkatan
Accuracy	90:10	90.67%	93.67%	3.31%
	80:20	91.83%	93.83%	2.18%
	70:30	90.67%	92.44%	1.95%
Precision	90:10	91%	94%	3.30%
	80:20	92%	94%	2.17%
	70:30	91%	92%	1.10%
Recall	90:10	91%	94%	3.30%
	80:20	92%	94%	2.17%
	70:30	91%	92%	1.10%
Spesificity	90:10	95.33%	96.83%	1.57%
	80:20	95.92%	95.92%	0.00%
	70:30	95.33%	96.22%	0.93%
F1-Score	90:10	91%	94%	3.30%
	80:20	92%	94%	2.17%
	70:30	91%	92%	1.10%

Berdasarkan tabel 4, model ResNet-101 yang dioptimasi dengan metode *bayesian optimization* menunjukkan peningkatan kinerja yang signifikan pada sebagian besar metrik evaluasi di semua skenario pembagian data dibandingkan dengan model baseline.

Skenario 90:10 menunjukkan peningkatan akurasi sebesar 3.00% (dari 90.67% menjadi 93.67%), serta peningkatan presisi, *recall*, dan F1-Score sebesar 3% (dari 91% menjadi 94%). Spesifisitas juga meningkat 1.50% (dari 95.33% menjadi 96.83%). Selanjutnya untuk skenario 80:20 yang sebelumnya merupakan skenario terbaik di baseline, akurasi meningkat 2.00% (dari 91.83% menjadi 93.83%), dan metrik presisi, *recall*, serta F1-Score juga meningkat 2% (dari 92% menjadi 94%). Meskipun spesifisitas tetap pada 95.92% (tidak ada peningkatan), nilai ini sudah sangat tinggi. Lalu pada skenario

70:30 juga menunjukkan peningkatan kinerja, dengan akurasi meningkat 1.77% (dari 90.67% menjadi 92.44%), dan metrik presisi, *recall*, serta *F1-Score* meningkat 1% (dari 91% menjadi 92%). Spesifisitas pada skenario ini juga meningkat 0.89% (dari 95.33% menjadi 96.22%). Peningkatan kinerja yang konsisten di seluruh metrik dan skenario ini secara jelas mengindikasikan efektivitas Bayesian Optimization dalam menemukan *hyperparameter* yang lebih optimal. Hal ini berdampak positif pada kemampuan generalisasi dan klasifikasi model, memungkinkan model untuk mencapai hasil yang lebih akurat dan robust dalam mendeteksi kelas-kelas yang berbeda.

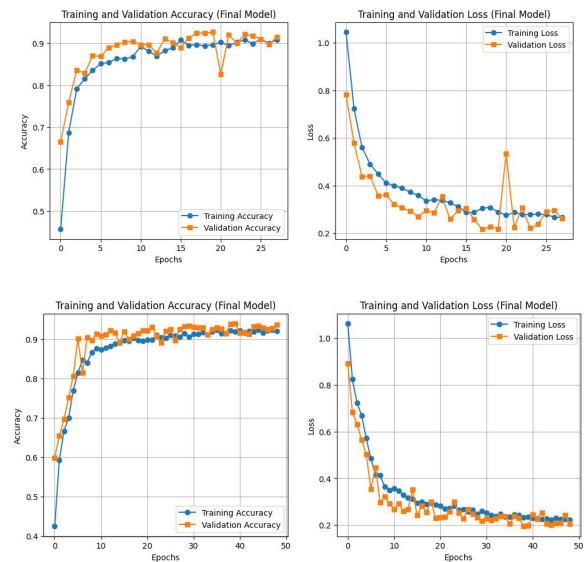
Confusion matrix dari model akhir yang telah dioptimasi disajikan untuk menganalisis kinerja klasifikasi per kelas secara mendalam. *Confusion matrix* untuk skenario 90:10, 80:20, dan 70:30 masing-masing secara berurutan ditunjukkan pada Gambar 7 berikut:

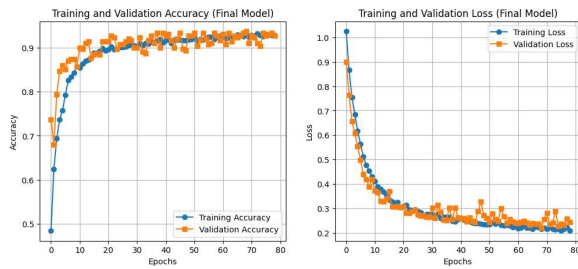


Gbr. 7 *Confusion Matrix Model Akhir*

Pada skenario 70:30, model menunjukkan performa sangat baik dengan tingkat akurasi tinggi pada kelas COVID-19 dan Normal, meskipun terdapat misklasifikasi terbesar pada Pneumonia yang diprediksi sebagai Normal (21 kasus) dan COVID-19 sebagai Pneumonia (16 kasus). Dibandingkan baseline, model berhasil menurunkan false positive dan false negative secara signifikan. Selanjutnya, pada skenario 80:20, model efektif dalam mengklasifikasikan COVID-19 dan Pneumonia dengan jumlah misklasifikasi terbanyak terjadi pada Pneumonia yang diprediksi sebagai Normal (15 kasus) dan Normal sebagai COVID-19 (7 kasus). Akurasi meningkat terutama pada kelas COVID-19, dengan pengurangan signifikan misklasifikasi dibanding baseline. Lalu, pada skenario 90:10, model mengklasifikasikan sebagian besar kasus secara tepat dengan kesalahan minimal. Misklasifikasi terbanyak adalah Normal yang diprediksi sebagai COVID-19 (5 kasus) dan Pneumonia sebagai Normal (7 kasus). Jika dibandingkan baseline, model menunjukkan peningkatan akurasi dan presisi pada semua kelas.

Selain evaluasi berdasarkan *confusion matrix*, analisis performa model juga diperkuat melalui pengamatan terhadap plot kurva akurasi dan *loss* selama proses pelatihan dan validasi. visualisasi ini memberikan gambaran mendalam mengenai stabilitas pelatihan, kecepatan konvergensi, serta indikasi adanya *overfitting* atau tidak, pada masing masing skenario pembagian data. Plot kurva pada dilihat pada Gambar 8.





Pada skenario 90:10, model menunjukkan peningkatan akurasi pelatihan dan validasi yang cepat dan stabil, dengan akurasi validasi mencapai lebih dari 90% dan loss turun konsisten hingga sekitar 0.2–0.25. Tidak terdapat indikasi overfitting, dan model menunjukkan konvergensi lebih optimal dibandingkan baseline. Sementara pada skenario 80:20, akurasi validasi stabil di atas 90% setelah 10–15 epoch, dengan loss pelatihan dan validasi yang turun mulus hingga di bawah 0.25. Model menunjukkan generalisasi yang sangat baik, dengan performa yang jelas lebih baik dibanding baseline, berkat optimalisasi hyperparameter. Sedangkan, pada skenario 70:30, akurasi dan loss menunjukkan tren positif, meskipun terdapat sedikit fluktuasi yang lebih nyata dan cukup signifikan pada kurva validasi (sekitar *epoch* 20–25). Namun akurasi validasi tetap mencapai lebih dari 90%, dan *loss* berada di level rendah, meskipun kurang stabil dibanding skenario lain. Meski begitu, performa tetap meningkat signifikan dibanding baseline, dan tidak menghambat kemampuan model untuk mencapai kinerja tinggi.

Observasi dari ketiga grafik ini secara kolektif menunjukkan bahwa model yang dioptimasi dengan *bayesian optimization* mencapai konvergensi yang lebih cepat dan stabil, serta akurasi validasi yang lebih tinggi dan *loss* yang lebih rendah dibandingkan dengan model baseline di semua skenario. Hal ini mengindikasikan bahwa hyperparameter optimal berhasil meminimalkan *overfitting* dan meningkatkan kemampuan generalisasi model secara signifikan. Kinerja paling stabil dan optimal terlihat pada skenario 80:20, diikuti oleh 90:10 dan 70:30 menunjukkan kinerja yang baik meskipun dengan sedikit banyak fluktuasi.

IV. KESIMPULAN

Penelitian ini secara komprehensif membuktikan bahwa optimasi hyperparameter menggunakan metode *Bayesian Optimization* pada arsitektur *deep learning* ResNet-101 secara signifikan meningkatkan performa klasifikasi citra x-ray paru-paru untuk deteksi normal, pneumonia, dan COVID-19. Peningkatan kinerja model pasca-optimasi diamati secara konsisten pada seluruh metrik evaluasi (akurasi, presisi, *recall*, spesifisitas, dan *F1-Score*) dibandingkan dengan model *baseline* pada ketiga skenario pembagian data (90:10, 80:20, dan 70:30). Akurasi tertinggi model akhir, sebesar 93,83% pada skenario 80:20, mengindikasikan bahwa konfigurasi pembagian data tersebut, dikombinasikan dengan

hyperparameter hasil optimasi, merupakan kombinasi paling efektif.

Analisis kurva akurasi dan *loss* menegaskan stabilitas dan kecepatan konvergensi model yang dioptimasi, dengan minimnya indikasi *overfitting* yang signifikan. Hal ini secara implisit menunjukkan bahwa Bayesian Optimization berhasil menjelajahi ruang *hyperparameter* secara efisien, menemukan kombinasi optimal yang mengoptimalkan keseimbangan antara *bias* dan *variance* model. Efektivitas *Bayesian Optimization* dalam mengidentifikasi *hyperparameter* optimal ini tidak hanya berujung pada peningkatan metrik kuantitatif, tetapi juga secara fundamental memperkuat kemampuan generalisasi dan *robustness* model dalam mengklasifikasikan kasus-kasus penyakit paru-paru yang beragam. Oleh karena itu, penelitian ini memberikan kontribusi krusial dalam pengembangan sistem deteksi otomatis berbasis *deep learning* yang lebih akurat dan andal di bidang radiologi medis, berpotensi mempercepat diagnosis dan mendukung penanganan penyakit paru-paru secara dini.

REFERENSI

- [1] A. D. Deva, F. Firdaus, S. Hasyim, B. Yanto, R. M. Candra *et al.*, “Klasifikasi prediksi penyakit paru-paru normal dengan pneumonia berdasarkan citra image x-ray dengan optimasi adam convolutional neural network (cnn),” *RJOCS (Riau Journal of Computer Science)*, vol. 10, no. 2, pp. 146–155, 2024
- [2] M. N. Achmadiah, J. M. Hasan, and N. Setyawan, “Convolutional neural network (cnn) sebagai metode pendeteksi penderita covid-19 pada x-ray paru-paru manusia,” *CYCLOTRON*, vol. 5, no. 2, 2022.
- [3] A. D. Azzumzumi, M. Hanafi, and W. M. P. Dhuhita, “Klasifikasi Penyakit Paru-Paru Berdasarkan Peningkatan Kualitas Kontras dan EfficientNet Menggunakan Gambar X-Ray,” *Teknika*, vol. 13, no. 2, pp. 293–300, Jul. 2024.
- [4] L. A. Andika, H. Pratiwi, and S. S. Handajani, “Klasifikasi penyakit pneumonia menggunakan metode convolutional neural network dengan optimasi adaptive momentum,” *Indonesian Journal of Statistics and Its Applications*, vol. 3, no. 3, pp. 331–340, 2019
- [5] M. Fitriyarsari, “Deteksi covid-19 pada citra x-ray dada menggunakan machine learning,” *Jurnal INSTEK (Informatika Sains dan Teknologi)*, vol. 7, no. 1, pp. 1–10, 2022
- [6] N. Yudistira, A. W. Widodo, and B. Rahayudi, “Deteksi covid-19 pada citra sinar-x dada menggunakan deep learning yang efisien,” *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 7, no. 6, pp. 1289–1296, 2020.
- [7] S. Vedhanayaki and V. Indragandhi, “A bayesian optimized deep learning approach for accurate state of charge estimation of lithium ion batteries used for electric vehicle application,” *IEEE Access*, vol. 12, pp. 43 308–43 327, 2024

- [8] K. Mae, "Perbandingan grid search dan bayesian optimization pada klasifikasi collection intensity scoring dan prediksi channel recommendation (studi kasus: P2p lending)," 2024, skripsi, Universitas Multimedia Nusantara.
- [9] M. Loey, S. El-Sappagh, and S. Mirjalili, "Bayesian-based optimized deep learning model to detect COVID-19 patients using chest X-ray image data," *Comput Biol Med*, vol. 142, Mar. 2022.
- [10] M. Miranda, K. Valeriano *et al.*, "A detailed study on the choice of hyperparameters for transfer learning in covid-19 image datasets using bayesian optimization," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 4, 2021.
- [11] S. Kumar, "Covid19-pneumonia-normal chest x-ray images," <https://doi.org/10.17632/dvntn9yhd2.1>, 2022, accessed: 2024-05-01.
- [12] M. Ahsan, M. A. Based, J. Haider, and M. Kowalski, "Covid-19 detection from chest x-ray images using feature fusion and deep learning," *Sensors*, vol. 21, no. 4, p. 1480, 2021.
- [13] A. Vierisyah, R. M. Fajri *et al.*, "Klasifikasi kanker paru paru menggunakan cnn dengan 5 arsitektur," *Journal of Intelligent Networks and IoT Global*, vol. 1, no. 2, pp. 84–91, 2023.
- [14] J. Khatib Sulaiman, J. Agung Nurcahyo, T. Bayu Sasongko, U. Amikom Yogyakarta, and K. Kunci, "Hyperparameter Tuning Algoritma Supervised Learning untuk Klasifikasi Keluarga Penerima Bantuan Pangan Beras," *Indonesian Journal of Computer Science*. Vol 12, no. 3, 2023.
- [15] K. E. S. Wijaya, G. A. Pradipta, and D. Hermawan, "Optimisasi parameter vggnet melalui bayesian optimization untuk klasifikasi nodul paru," in *Seminar Hasil Penelitian Informatika dan Komputer (SPINTER)* Institut Teknologi dan Bisnis STIKOM Bali, 2024, pp. 882–887.
- [16] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, 2020.
- [17] M. Ait Amou, K. Xia, S. Kamhi, and M. Mouhafid, "A novel mri diagnosis method for brain tumor classification based on cnn and bayesian optimization," in *Healthcare*, vol. 10, no. 3. MDPI, 2022, p. 494.
- [18] J. Balseca, M. Cruz, and M. G. Baldeón, "Covid-19 resnet: Residual neural network for covid-19 classification with three-step bayesian optimization," *ACI Avances en Ciencias e Ingenierías*, no. 1, 2020.