



Penerapan Saliency Maps dalam Explainable AI Untuk Deteksi Penyakit Paru-Paru pada Citra X-Ray Dada dengan Deep Learning

Wahyu Reinaldy, Benny Sukma Negara*, Muhammad Irsyad, Muhammad Affandes, Surya Agustian

Fakultas Sains dan Teknologi, Prodi Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ¹12150111610@students.uin-suska.ac.id, ^{2,*}bsnegara@uin-suska.ac.id, ³irsyadtech@uin-suska.ac.id,

⁴affandes@uin-suska.ac.id, ⁵surya.agustian@uin-suska.ac.id

Email Penulis Korespondensi: bsnegara@uin-suska.ac.id

Abstrak—Identifikasi dini penyakit paru-paru sangat penting agar tenaga medis dapat dengan cepat memberikan pertolongan pertama dan mempelajari lebih lanjut kondisi pasien. Dalam penelitian ini, dikembangkan sebuah model untuk mengklasifikasikan gambar X-ray dada pada paru-paru menggunakan arsitektur VGG16. Gambar X-ray dada ini dikategorikan menjadi tiga kelompok: covid-19, paru-paru normal, dan pneumonia. Campuran hyperparameter, termasuk *learning rate* 0,001, 50 epoch, dan *batch size* 16, digunakan untuk melatih model, menghasilkan akurasi sebesar 96%. Beberapa metrik ukuran, termasuk *precision*, *recall*, *f1-score*, dan *confusion matrix*, digunakan untuk mengevaluasi model. Selain itu, metode *saliency maps* digunakan untuk menafsirkan secara visual output prediksi model dan menampilkan area gambar X-ray dada yang paling mempengaruhi pengambilan keputusan model. Temuan visualisasi *saliency maps* menunjukkan bahwa model memfokuskan prediksinya pada wilayah paru-paru yang terkait dengan penyakit, yang membantu dalam memahami proses pengambilan keputusan algoritma.

Kata Kunci: Penyakit Paru-Paru; Gambar X-Ray Dada; VGG16; Saliency Maps; Visual

Abstract—Early identification of lung diseases is very important so that medical personnel can quickly provide first aid and further study the patient's condition. In this study, a model was developed to classify chest X-ray images of the lungs using the VGG16 architecture. These chest X-ray images were categorized into three groups: COVID-19, normal lungs, and pneumonia. A combination of hyperparameters, including a learning rate of 0.001, 50 epochs, and a batch size of 16, was used to train the model, achieved an accuracy of 96%. Several evaluation metrics, including precision, recall, f1-score, and confusion matrix, were used to assess the model. In addition, saliency map methods were used to visually interpret the model's prediction output and display the areas of the chest X-ray images that most influenced the model's decision-making. The saliency map visualization findings show that the model focuses its predictions on regions of the lungs associated with the disease, which helps in understanding the algorithm's decision-making process.

Keywords: Lung Diseases; Chest X-Ray Images; VGG16; Saliency Maps; Visually Interpret

1. PENDAHULUAN

Salah satu masalah kesehatan yang dapat mengakibatkan rawat inap atau bahkan kematian adalah penyakit paru-paru yang menyebar melalui sistem pernapasan. Pneumonia dan covid-19 adalah dua dari kondisi paru-paru yang dapat menimbulkan risiko besar jika tidak segera dan tepat ditangani. SARS-CoV-2 adalah penyebab covid-19, suatu penyakit yang dapat menyebar melalui pernapasan dan menampilkan gejala berupa demam, batuk, dan sesak napas (Jain et al., 2020). Pneumonia, di sisi lain, adalah penyakit yang menyerang jaringan paru-paru (alveoli) dan disebabkan oleh berbagai mikroorganisme, seperti bakteri, jamur, dan virus. Di antara beberapa penyakit lainnya, pneumonia menyumbang 14,5 persen dari angka kematian pada bayi berusia antara 29 hari hingga 11 bulan pada tahun 2020 (Ma'ruf, 2020). Pemeriksaan X-ray dada adalah salah satu cara untuk menilai penyakit paru-paru karena mudah didapat dan sangat efektif dalam mengidentifikasi kelainan pada paru-paru (Satari & Banda, 2024). Pemeriksaan ini dapat dilakukan menggunakan mesin X-ray tanpa membedah tubuh pasien (Azizah & Juniati, 2021). Namun, interpretasi manual gambar X-ray dada biasanya hanya bergantung pada pengetahuan dokter atau radiolog, yang dapat mengakibatkan subjektivitas dan variasi dalam diagnosis antar pemeriksa.

Deep learning telah berkembang secara signifikan dalam bidang pemrosesan citra medis. Berkat penggunaan CNN (*Convolutional Neural Network*), yang dapat secara otomatis mengekstrak dan mengenali informasi visual dari gambar. *Deep learning* dalam perawatan kesehatan telah secara signifikan mengubah cara gambar medis diperiksa dan dipahami (M.Kranthi et al., 2024). Dalam sebuah studi berbasis CNN, tingkat akurasi pelatihan model adalah 99,9% (Karar et al., 2021). VGG16 adalah salah satu jenis CNN yang populer yang memiliki struktur dalam dan dapat menghasilkan representasi fitur yang kuat. Model CNN telah menunjukkan tingkat akurasi yang tinggi dalam membedakan antara penyakit paru-paru normal dan abnormal pada gambar X-ray dada. Namun, penggunaan *deep learning* di domain medis memerlukan kinerja yang tinggi serta transparansi dalam proses pengambilan keputusan model.

Fakta bahwa model *deep learning* sering berubah menjadi "black box" adalah salah satu hambatan utama dalam menggunakannya untuk diagnosis medis. Meskipun model *deep learning* dapat menghasilkan hasil yang benar, sifat *black box* mereka membuat sulit untuk memahami bagaimana algoritma membuat penilaian (Rahman et al., 2021).

Ketika digunakan di lingkungan klinis, situasi ini mengurangi kepercayaan terhadap keamanan dan keandalan model, terutama dalam hal mengidentifikasi penyakit seperti pneumonia dan covid-19. Sulit untuk menentukan apakah model benar-benar memfokuskan prediksinya pada wilayah paru-paru yang relevan atau pada aspek gambar yang tidak terkait dengan penyakit tersebut tanpa adanya interpretasi yang jelas. Oleh karena itu, dibutuhkan metode XAI yang dapat menjelaskan proses pengambilan keputusan model selain mengutamakan akurasi (Arsenault et al., 2025).

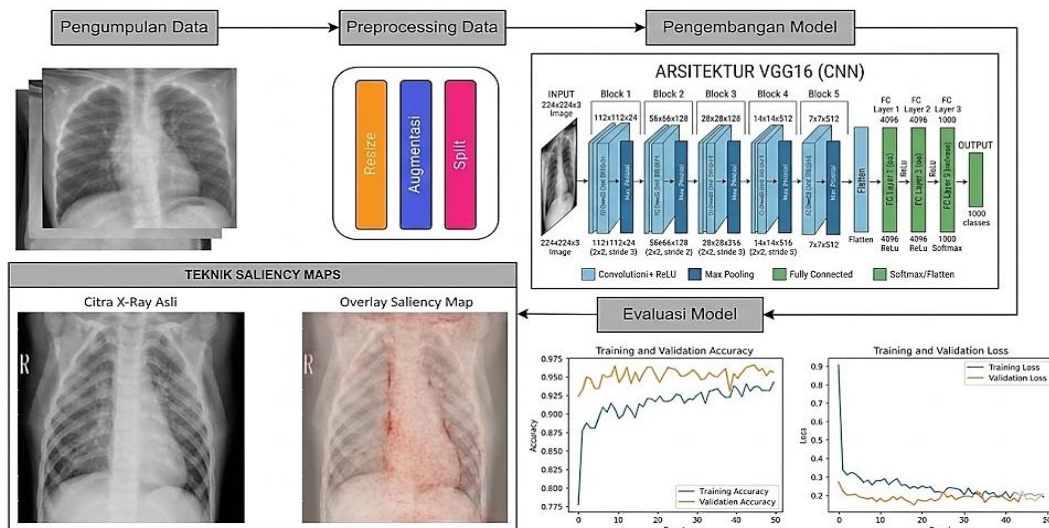
Menurut studi sebelumnya, model CNN yang dibangun berdasarkan arsitektur VGG16 dapat secara efektif mengkategorikan gambar X-ray dada (Idhom et al., 2023; Shahzad et al., 2022). Sebagian besar dari studi ini tidak secara menyeluruh meneliti faktor-faktor interpretabilitas, sebaliknya mereka lebih fokus pada peningkatan kinerja model. Selain itu, model *deep learning* pada citra medis telah diinterpretasikan menggunakan beberapa teknik XAI, seperti Grad-CAM dan SHAP (Muliani et al., 2025; Sofiyah et al., 2025). Namun, proses yang terlibat dalam penerapan dan interpretasinya biasanya lebih rumit. Untuk membantu memahami proses pengambilan keputusan dari model klasifikasi gambar X-Ray Dada, metode *saliency maps* digunakan dalam penelitian ini untuk memberikan representasi yang mudah dipahami dari area penting dalam citra berdasarkan gradien model.

Mengingat masalah-masalah ini, penelitian ini menyarankan penggunaan arsitektur VGG16 bersamaan dengan metode *saliency maps* untuk mengidentifikasi gambar X-ray dada sebagai bagian dari strategi XAI (*Explainable Artificial Intelligence*). *Saliency maps* dapat menarik perhatian pada aspek-aspek yang relevan dari proses pengambilan keputusan model (Wollek et al., 2023). Selain itu, telah ditunjukkan bahwa berbagai jenis *saliency maps* dapat menginterpretasikan secara visual dari keluaran model (Arun et al., 2021). Pemeriksaan visual terhadap area yang dilihat model untuk setiap jenis penyakit, yang membuat keputusan model lebih jelas dan mudah dipahami, adalah yang membuat penelitian ini istimewa. Dengan menggunakan *saliency maps*, penelitian ini bertujuan untuk menunjukkan bahwa model tidak hanya membuat prediksi yang tepat tetapi juga menarik perhatian pada area paru-paru yang relevan, meningkatkan kepercayaan pada hasil kategorisasi.

Tujuan utama dari studi ini adalah untuk membuat dan menilai model *deep learning* berbasis VGG16 untuk mengkategorikan gambar X-ray dada menjadi tiga kelompok: covid-19, paru-paru normal, dan pneumonia. Selain itu, studi ini meneliti seberapa baik model menggunakan *saliency maps* untuk memberikan penjelasan yang jelas. Ruang lingkup studi ini dibatasi pada penggunaan dataset gambar X-ray Dada, mengevaluasi kinerja model menggunakan metrik *accuracy*, *precision*, *recall*, dan *f1-score*, serta menggunakan *saliency maps* untuk memvisualisasikan area yang menjadi fokus model. Penciptaan model klasifikasi gambar X-ray dada berbasis VGG16 bersamaan dengan teknik *saliency maps* untuk meningkatkan interpretabilitas model dalam mengidentifikasi covid-19, paru-paru normal, dan pneumonia merupakan kontribusi utama dari studi ini. Studi ini menawarkan visualisasi wilayah citra yang memengaruhi proses pengambilan keputusan model selain memusatkan perhatian pada kinerja klasifikasi. Harapan penelitian ini adalah untuk membantu menciptakan sistem pendukung keputusan medis yang andal, transparan, dan akurat.

2. METODOLOGI PENELITIAN

Metodologi studi ini adalah sebagai berikut. Dua tahap metodologi tersebut adalah klasifikasi gambar X-ray dada menggunakan *deep learning* dan penerapan *saliency maps* untuk menganalisis kinerja model. Gambar 1 menggambarkan alur pendekatan penelitian yang digunakan dalam penyelidikan ini.



Gambar 1. Metodologi Penelitian

2.1 Pengumpulan Data

Gambar X-ray dada yang menunjukkan masalah pada paru-paru akan menjadi dataset yang digunakan dalam penelitian ini. Informasi ini disediakan oleh situs web Mendeley, yang tersedia untuk umum (Shastri et al., 2022). Dataset ini dibagi menjadi tiga kategori: covid-19, paru-paru normal, dan pneumonia. Terdapat 3006 foto dalam seluruh dataset. Setelah itu, dataset dipisahkan menjadi data pelatihan, yang terdiri dari 2400 gambar, data validasi, yang terdiri dari 600 gambar, dan enam lagi data gambar dari sumber yang sama yang belum dilatih digunakan untuk membuat *saliency*

maps. Gambar 2 menunjukkan contoh gambar X-ray dada untuk kategori covid-19 (a), paru-paru normal (b), dan pneumonia (c).



(a)

(b)

(c)

Gambar2. Gambar X-Ray Dada (a) Covid-19, (b) Paru-paru Normal, (c) Pneumonia

2.2 Preprocessing Data

Prosedur *preprocessing* kemudian diselesaikan sebelum prosedur pelatihan model. Langkah-langkah *preprocessing* yang diselesaikan adalah sebagai berikut:

2.2.1 Resize

Untuk memenuhi persyaratan input dari arsitektur VGG16, prosedur penskalaan digunakan untuk menstandarisasi ukuran gambar X-ray dada. Setiap gambar kini diskalakan menjadi 224 x 224 piksel. Model VGG16 hanya dapat menerima input dengan dimensi tertentu, sehingga menstandarisasi ukuran gambar sangat penting. Ukuran gambar yang konsisten memungkinkan jaringan saraf konvolusional mempelajari fitur dengan lebih efisien dan teratur.

2.2.2 Normalisasi

Tujuan normalisasi gambar adalah untuk mengurangi rentang nilai piksel menjadi ukuran yang lebih konsisten dan lebih kecil. Dalam penyelidikan ini, nilai piksel maksimum digunakan untuk membagi nilai intensitas piksel, yang awalnya berkisar dari 0 hingga 255, menjadi rentang 0 hingga 1. Dengan meminimalkan perbedaan skala antar fitur, tahap ini membantu meningkatkan stabilitas proses pelatihan dan mempercepat konvergensi model.

2.2.3 Split Data

20% dari data digunakan untuk validasi, sementara 80% sisanya digunakan untuk pelatihan. Data pelatihan dan validasi dibagi dengan rasio 80:20. Rasio ini dipilih berdasarkan penelitian yang menunjukkan bahwa rasio 80:20 adalah standar yang sering diterapkan dalam studi serupa karena telah terbukti menghasilkan kinerja model yang stabil dan ideal dalam berbagai skenario (Nguyen et al., 2021). Tabel 1 menunjukkan jumlah dataset yang digunakan pada penelitian ini untuk masing-masing kategori.

Tabel 1. Pembagian dataset

Dataset Paru-Paru	Covid	Paru-paru normal	Pneumonia
Training	800	800	800
Testing	200	200	200

2.2.4 Augmentasi

Data pelatihan menjalani prosedur augmentasi data untuk meningkatkan kapasitas model dalam melakukan generalisasi dan mengurangi kecenderungannya untuk *overfitting*. Tujuan dari augmentasi data adalah meningkatkan variasi gambar tanpa mengubah label kelas, yang memungkinkan model untuk mengenali beragam pola lebih banyak. Diharapkan perubahan gambar yang dihasilkan oleh augmentasi dapat secara akurat menggambarkan kondisi nyata yang mungkin terjadi selama pemindaian X-ray dada. Untuk mempertahankan objektivitas hasil evaluasi model, prosedur ini hanya diterapkan pada data pelatihan, sementara data validasi tetap tidak diubah.

2.3 Proses Deep Learning

Teknik utama yang digunakan dalam penelitian ini untuk membangun sistem klasifikasi gambar X-ray dada adalah *deep learning*. Metode ini dipilih karena tidak memerlukan prosedur ekstraksi fitur manual dan dapat mempelajari pola dan representasi fitur yang kompleks secara langsung dari data gambar. Model dapat mengenali ciri visual yang terkait dengan masalah paru-paru pada gambar paru-paru normal maupun yang menunjukkan adanya penyakit berkat teknik *deep learning*.

Sejumlah langkah penting terlibat dalam proses *deep learning* terapan, seperti pelatihan model dan evaluasi kinerja dari hasil klasifikasi. Arsitektur CNN, yang sering digunakan dalam pemrosesan dan analisis gambar digital

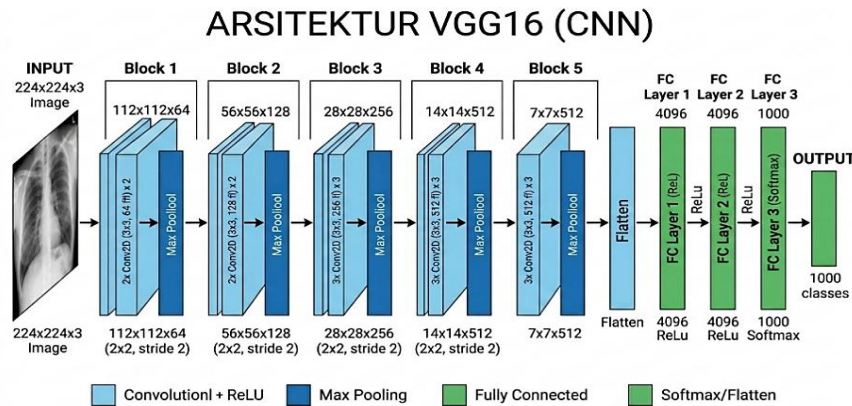
serta menjadi dasar untuk pembuatan model VGG16, digunakan dalam penelitian ini. Oleh karena itu, untuk menjelaskan proses kategorisasi gambar X-Ray Dada yang digunakan dalam penelitian ini, pemahaman dasar tentang CNN diperlukan.

2.3.1 Convolutional Neural Network

Salah satu arsitektur *deep learning* untuk menangani data gambar adalah CNN (*Convolutional Neural Network*). Melalui operasi konvolusi, CNN dapat secara otomatis mengekstrak karakteristik visual, memungkinkan untuk mengidentifikasi pola penting seperti tepi, tekstur, dan bentuk objek. Akibatnya, CNN sangat berguna untuk identifikasi, klasifikasi, dan pengenalan item yang lebih akurat (Gupta et al., 2022). *Convolutional layer*, *activation function*, *pooling layer*, dan *fully connected layer* adalah empat arsitektur utama CNN. *Convolutional Layer* membuat peta fitur dengan melakukan operasi konvolusi pada gambar input menggunakan filter atau kernel. Untuk menambahkan karakteristik non-linear pada model, keluaran dari konvolusi ini kemudian dijalankan melalui fungsi aktivasi, seperti ReLU (*Rectified Linear Unit*). *Pooling layer* kemudian mengurangi dimensi peta fitur sambil mempertahankan informasi penting, yang menurunkan biaya komputasi dan mengurangi kemungkinan *overfitting*. *Fully connected layer* di akhir jaringan digunakan untuk melakukan kategorisasi menggunakan fitur yang dikumpulkan. CNN sangat berguna untuk analisis citra medis karena kemampuannya untuk menangkap fitur spasial dari gambar.

2.3.2 Arsitektur VGG16

Universitas Oxford melalui VGG (*Visual Geometry Group*) mengembangkan jaringan CNN VGG16. *Convolutional layer* dan *fully connected layer* termasuk di antara 16 lapisan dengan parameter yang membentuk desain ini. Dengan struktur yang sederhana dan konsisten, VGG16 menggunakan filter konvolusi berukuran kecil 3x3 yang diatur secara bertahap untuk mengekstraksi fitur secara mendalam. Fitur yang diekstraksi kemudian digunakan untuk melakukan klasifikasi. CNN sangat berguna untuk analisis citra medis karena kemampuan mereka menangkap fitur spasial dari gambar. VGG16 digunakan sebagai model dasar dalam penelitian ini dengan menggunakan teknik *transfer learning*, bobot awal model diperoleh melalui pelatihan pada dataset ImageNet. Pendekatan ini dipilih karena dapat mempercepat pelatihan dan meningkatkan kinerja model, terutama dalam situasi di mana data pelatihan terbatas. Lapisan klasifikasi pada bagian terakhir VGG16 kemudian diubah untuk mencerminkan tiga klasifikasi yang digunakan dalam penelitian ini: covid-19, paru-paru normal, dan pneumonia. VGG16 dipilih sebagai model dasar untuk penelitian ini karena stabilitasnya dalam sejumlah percobaan klasifikasi gambar medis dan kemampuannya untuk mengekstraksi fitur visual yang baik. Gambar 3 menunjukkan alur kerja arsitektur VGG16 yang digunakan dalam proses klasifikasi.



Gambar 3. Arsitektur VGG16

2.3.3 Konfigurasi Pelatihan Model

Untuk mendapatkan hasil terbaik, sejumlah parameter kunci ditentukan selama proses pelatihan model. Model VGG16 yang sebelumnya dilatih pada dataset *ImageNet*, digunakan dalam penelitian ini untuk menggunakan metode *transfer learning*. Dengan memanfaatkan tugas-tugas yang serupa dari model yang telah dilatih dalam studi sebelumnya, seperti *ImageNet*, *transfer learning* bertujuan untuk meningkatkan kinerja model (H. E. Kim et al., 2022). Ini meningkatkan kemampuan model untuk mengklasifikasikan gambar X-ray dada dan mempercepat pelatihan. Konfigurasi pelatihan mencakup *Learning rate* sebesar 0,001, 50 epoch pelatihan, *batch size* sebesar 16. *Learning rate* yang dipilih berusaha untuk menyeimbangkan antara kecepatan belajar dan stabilitas model sehingga model dapat memahami pola data secara efisien tanpa mengalami gangguan selama proses pembelajaran. Tabel 2 menunjukkan kombinasi hyperparameter yang digunakan.

Tabel 2. Hyperparameter

Parameter	Nilai
Pembagian Data	80:20
Optimizer	Adam



Parameter	Nilai
Activation Function	ReLu
Epoch	50
Batch Size	16
Learning Rate	0,001

2.4 Evaluasi Model

Kemampuan model untuk mengklasifikasikan gambar X-ray dada dengan benar dan berhasil ke dalam kategori yang telah ditentukan dievaluasi. Penilaian ini bertujuan untuk menentukan akurasi dan konsistensi model dalam memprediksi hasil data uji. Untuk memahami sepenuhnya kemampuan model, studi ini mencakup sejumlah metrik penilaian, termasuk *accuracy*, *precision*, *recall*, *f1-score*, dan *confusion matrix*.

Ukuran evaluasi sebagian besar dihitung menggunakan *confusion matrix*. Dengan membandingkan label yang diharapkan dan yang sebenarnya, metrik ini merangkum hasil prediksi model. Nilai dari temuan uji dapat dipelajari dengan bantuan *confusion matrix* (S. C. Kim & Cho, 2022). *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) adalah empat komponen utama dari *confusion matrix*. Berikut ini adalah *accuracy*, *precision*, *recall*, dan *f1-score*:

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

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

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

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

Persamaan tersebut menjelaskan komponen dalam *confusion matrix* yang digunakan untuk mengevaluasi kinerja model klasifikasi. True Positive (TP) merupakan data positif yang berhasil diprediksi dengan benar sebagai positif, sedangkan True Negative (TN) adalah data negatif yang diprediksi benar sebagai negatif. False Positive (FP) menunjukkan data negatif yang keliru diprediksi sebagai positif, sementara False Negative (FN) adalah data positif yang salah diprediksi sebagai negatif oleh model. Evaluasi model dalam penelitian ini dapat memberikan penjelasan yang lebih menyeluruh mengenai kemampuan model dengan menggabungkan berbagai metrik ukuran, termasuk *accuracy*, *precision*, *recall*, dan *f1-score* bersama dengan temuan dari analisis *confusion matrix*. Selain mempertimbangkan akurasi keseluruhan dari prediksi, pendekatan evaluasi ini membantu dalam mengidentifikasi pola yang salah dalam mengklasifikasikan setiap kelas, yang dapat digunakan sebagai panduan untuk analisis lebih lanjut dan perbaikan model.

2.5 Implementasi Saliency Maps

Tahap berikutnya dari penelitian ini adalah menggunakan metode *saliency maps* bersamaan dengan metodologi XAI (*Explainable Artificial Intelligence*) untuk menjelaskan model *deep learning* yang telah dilatih. Untuk meningkatkan akurasi dan keandalan model dalam analisis citra medis, teknik ini berupaya untuk merepresentasikan secara grafis bagian-bagian dari gambar X-ray dada yang memiliki dampak substansial terhadap hasil klasifikasi yang dihasilkan oleh model.

Gradien dari skor kelas output model dalam kaitannya dengan nilai setiap piksel pada gambar input adalah bagaimana *saliency maps* bekerja. Dengan metode ini, kita dapat menentukan sejauh mana setiap piksel mempengaruhi prediksi kelas tertentu, piksel dengan nilai gradien yang lebih tinggi dianggap memiliki pengaruh lebih besar terhadap proses pengambilan keputusan model. Simonyan et al., “*class saliency maps can be obtained by computing the gradient of the class score with respect to the input image*” (Simonyan et al., 2014). Menurut pernyataan ini, informasi gradien dapat digunakan sebagai referensi untuk menunjukkan bagaimana model memusatkan perhatian pada area tertentu dari gambar. Penggunaan kode *saliency maps* dalam studi ini didasarkan pada referensi yang dimodifikasi dari artikel serupa, dengan modifikasi yang dilakukan pada struktur model dan kebutuhan dataset (Subhash, 2022). *Saliency maps*, dapat dirumuskan sebagai turunan pertama dari skor kelas $S_c(I)$ terhadap citra input I , yang bisa dinyatakan sebagai berikut:

$$M = \left| \frac{\partial S_c(I)}{\partial(I)} \right| \quad (5)$$

M merupakan *saliency maps*, $S_c(I)$ adalah skor prediksi untuk kelas c dan I akan mempresentasikan citra input, Nilai absolut dari gradien digunakan untuk menunjukkan seberapa penting setiap piksel, tanpa memandang arah perubahan gradien tersebut. Gambar X-ray dada asli ditampilkan di sebelah *heatmap* yang mewakili hasil dari proses tersebut. Bagian paru-paru yang menjadi fokus model selama klasifikasi apakah pada covid-19, paru-paru normal, atau pneumonia ditampilkan dalam grafik ini. Oleh karena itu, *saliency maps* merupakan alat yang berguna baik untuk memvalidasi secara klinis hasil prediksi maupun menjelaskan proses pengambilan keputusan model *deep learning*.

3. HASIL DAN PEMBAHASAN

Temuan eksperimen dan diskusi dari penggunaan model *deep learning* berbasis arsitektur VGG16 dalam klasifikasi gambar X-ray dada disajikan dalam bagian ini. Hasil pengujian model dengan berbagai kombinasi hyperparameter, penilaian kinerja model menggunakan berbagai metrik pengujian, dan interpretasi keputusan model menggunakan *saliency maps* semuanya termasuk dalam hasil yang diberikan.

3.1 Pelatihan Model

Hasil studi yang menggunakan berbagai kombinasi hyperparameter untuk menentukan konfigurasi pelatihan model yang optimal ditunjukkan dalam subbagian ini. Dengan rasio pembagian data tetap 80:20 antara data pelatihan dan validasi, eksperimen hyperparameter melibatkan perubahan pada *learning rate*, jumlah epoch, dan *batch size*. Tabel 3 menunjukkan hasil beberapa percobaan kombinasi hyperparameter yang dilakukan.

Tabel 3. Kombinasi Hyperparameter

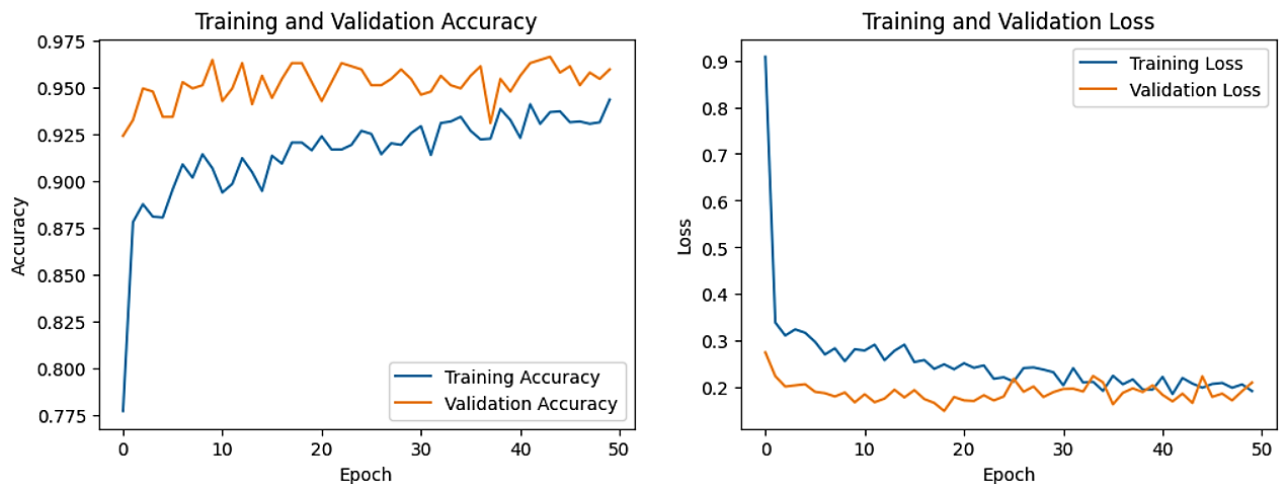
No	Rasio	Learning Rate	Epoch	Batch Size	Accuracy
1	80:20	0,0001	50	32	95%
2	80:20	0,001	25	32	94%
3	80:20	0,001	50	16	96%
4	80:20	0,01	25	8	33%
5	80:20	0,001	25	8	95%
6	80:20	0,01	50	8	33%
7	80:20	0,001	50	8	95%

Kombinasi hyperparameter dengan *learning rate* 0,001, 50 epoch, dan *batch size* 16 menghasilkan *accuracy* maksimum sebesar 96%, menurut hasil eksperimen yang ditampilkan pada Tabel 3. Temuan ini menunjukkan bahwa memilih set hyperparameter yang tepat dapat memengaruhi kemampuan model untuk mengklasifikasikan gambar X-ray dada seefektif mungkin. Berdasarkan studi sebelumnya yang menghasilkan hasil klasifikasi terbaik dalam tugas serupa, beberapa kombinasi hyperparameter ini dipilih (Sofiyah et al., 2025).

Menurut hasil uji, sejumlah kombinasi hyperparameter dapat menghasilkan nilai *accuracy* terbaik, yang berkisar antara 94% hingga 96%. Namun, konfigurasi optimal ditentukan sebagai kombinasi hyperparameter dengan *learning rate* 0,001, 50 epoch, *batch size* 16 karena menunjukkan pola pelatihan dan validasi yang relatif stabil dan konsisten selain menghasilkan *accuracy* tertinggi. Persyaratan ini menunjukkan bahwa model dapat secara efektif mempelajari fitur-fitur data dan menggeneralisasi pada data yang belum pernah dilihat sebelumnya.

3.2 Hasil Evaluasi Model

Bagian ini meneliti seberapa baik model *deep learning* yang dibangun berdasarkan arsitektur VGG16 dapat mengkategorikan gambar X-ray dada ke dalam tiga kelompok: covid-19, paru-paru normal, dan pneumonia. Penilaian ini dilakukan dengan melacak perkembangan *training accuracy*, *validation accuracy*, *training loss*, dan *validation loss* selama proses pelatihan model selama 50 epoch. Gambar 4 menunjukkan grafik hasil evaluasi model.



Gambar 4. Visualisasi evaluasi kinerja model

Terlihat dari grafik di atas bahwa dalam beberapa epoch pertama, nilai *accuracy* pada data pelatihan meningkat secara signifikan. *Training accuracy* dimulai sekitar 0,77 dan secara bertahap naik hampir mencapai 0,95 pada epoch terakhir. Pola ini menunjukkan bahwa seiring berjalannya pelatihan, model mampu mempelajari karakteristik kunci dari gambar X-ray dada dengan lebih efektif. Nilai *validation accuracy* tetap berada dalam kisaran yang cukup sebanding dengan *accuracy* pelatihan dan menunjukkan tren naik yang stabil. *Validation accuracy* sering tetap di atas atau sangat



setara dengan akurasi pelatihan sepanjang fase pelatihan, akhirnya mencapai sekitar 0,96. Kemampuan generalisasi yang tinggi dari model dan tidak adanya *overfitting* yang parah ditunjukkan oleh kecocokan antara *training accuracy* dan *validation*. Selain itu, selama proses pelatihan, grafik *training* dan *validation Loss* menunjukkan penurunan nilai *loss* secara bertahap. Pada awal setiap epoch, nilai *training loss* turun dengan cepat. Setelah itu, nilai tersebut secara bertahap naik dan turun hingga mencapai sekitar 0,24 pada epoch terakhir. Pola penurunan ini menunjukkan bahwa proses pelatihan berjalan dengan baik dan model berhasil meminimalkan kesalahan dalam memprediksi data pelatihan. Meskipun terdapat variasi kecil pada fase pelatihan tertentu, nilai *validation loss* juga menunjukkan penurunan yang relatif stabil dan mengikuti pola *training loss*. Pada akhir tahapan, tidak ada peningkatan kehilangan yang dapat dideteksi, dan perubahan masih berada dalam kisaran normal. Ini menunjukkan bahwa model tidak mengalami *overfitting* dan dapat terus berkinerja baik pada data yang tidak terlihat selama pelatihan. Secara keseluruhan, temuan dari evaluasi kinerja model menunjukkan bahwa strategi pelatihan efektif dalam menciptakan model dengan kemampuan klasifikasi yang kuat dan dapat diandalkan. Kemampuan model untuk berfungsi dengan baik dalam berbagai skenario ditunjukkan oleh kejelasan antara angka dari tahapan pelatihan dan validasi, yang membuatnya cocok untuk tahapan verifikasi tambahan dan analisis kapasitas penjelasan model menggunakan teknik XAI seperti *saliency maps*.

3.3 Hasil Evaluasi Klasifikasi

Efektivitas model *deep learning* yang berbasis arsitektur VGG16 dalam mengklasifikasikan gambar X-ray dada ke dalam tiga kategori covid-19, paru-paru normal, dan pneumonia telah dinilai. Untuk memperoleh pemahaman yang menyeluruh tentang kemampuan model, evaluasi ini menggunakan sejumlah metrik, termasuk *accuracy*, *precision*, *recall*, *f1-score*, dan *confusion matrix*. Tabel 4 menunjukkan hasil evaluasi performa klasifikasi model menggunakan metrik *precision*, *recall*, *f1-score*, dan *accuracy*.

Tabel 4. Hasil Evaluasi Klasifikasi

	Precision	Recall	F1-Score
Covid-19	0,97	0,99	0,98
Normal	0,96	0,94	0,95
Pneumonia	0,94	0,94	0,94
Accuracy			0,96

Berdasarkan hasil evaluasi, model yang diusulkan berhasil mencapai tingkat *accuracy* sebesar 96% pada data uji. Tingkat *accuracy* yang tinggi ini menunjukkan bahwa, dari seluruh data uji yang digunakan, model mampu mengkategorikan sebagian besar gambar X-ray dada dengan akurat. Dengan *precision* sebesar 0,97, *recall* sebesar 0,99, dan *f1-score* sebesar 0,98, model menunjukkan kinerja yang sangat baik pada kelas kasus covid-19. Tingginya *recall* menunjukkan bahwa model dapat mengidentifikasi hampir semua kasus covid-19 dengan sangat sedikit *false negative*, yang sangat penting untuk diagnosis medis. Model memperoleh *f1-score* sebesar 0,95, skor *precision* sebesar 0,96, dan *recall* sebesar 0,94 untuk kelas paru-paru normal. Temuan ini menunjukkan bahwa, meskipun terdapat beberapa kesalahan klasifikasi kecil, model ini cukup efektif dalam membedakan antara penyakit paru-paru normal dan patologis. Sementara itu, model menghasilkan *f1-score* sebesar 0,94, *precision* sebesar 0,94, dan *recall* sebesar 0,94 untuk kelas pneumonia. Hasil ini menunjukkan bahwa model mampu mengklasifikasikan kelas pneumonia dengan kinerja yang cukup baik. Mayoritas data yang diprediksi sebagai pneumonia sesuai dengan label sebenarnya, menurut peringkat *precision* sebesar 94%. Dengan skor *recall* sebesar 94%, model dapat mengidentifikasi mayoritas data pneumonia dalam dataset. Selain itu, model dianggap memiliki kinerja klasifikasi yang konsisten dan efisien untuk kelas pneumonia karena *f1-score* sebesar 94% menunjukkan keseimbangan antara nilai *precision* dan *recall*. Tabel 5 menunjukkan hasil dari *confusion matrix*.

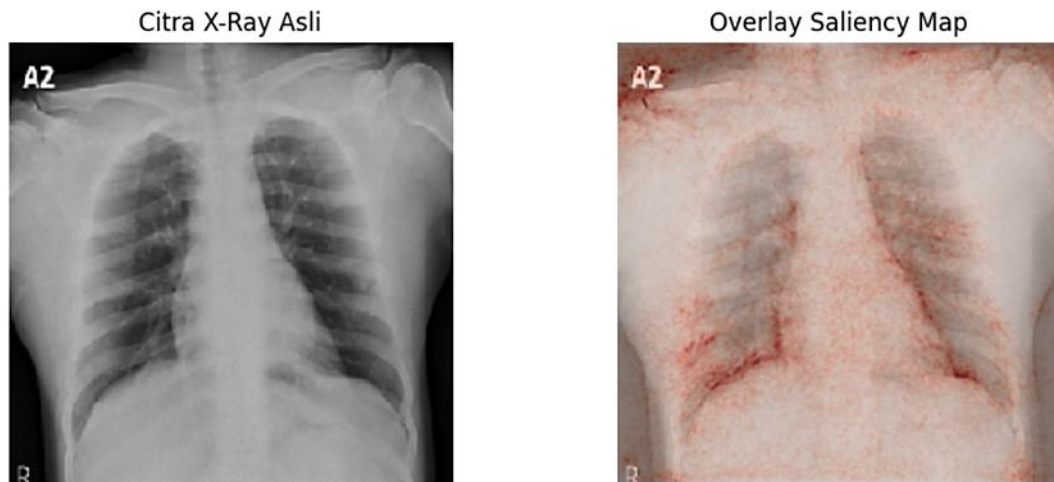
Tabel 5. Hasil Confusion Matrix

	Covid-19	Normal	Pneumonia
Covid-19	198	0	2
Normal	2	189	9
Pneumonia	4	7	189

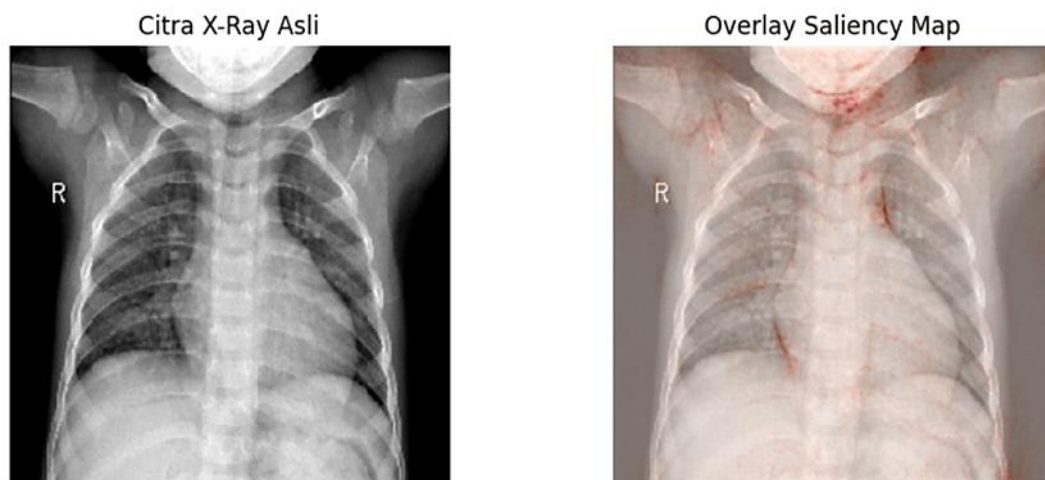
Confusion matrix digunakan untuk analisis tambahan guna lebih memahami pola kesalahan kategorisasi. Menurut *confusion matrix*, 198 dari 200 foto yang diuji pada kelas covid-19 diklasifikasikan dengan benar, sementara dua gambar salah dimasukkan ke dalam kelas pneumonia. Dari 200 foto di kelas normal, 189 didiagnosis dengan benar; gambar lainnya salah diberi label sebagai pneumonia atau covid-19. Dari 200 foto di kelas pneumonia, 189 diklasifikasikan dengan benar; gambar yang tersisa salah diklasifikasikan, sebagian besar ke kelas normal. Secara keseluruhan, hasil evaluasi klasifikasi menunjukkan bahwa model *deep learning* yang dikembangkan mampu mengklasifikasikan gambar X-ray dada dengan baik dan dapat diandalkan. Meskipun ada beberapa kesalahan, terutama pada kategori pneumonia dan normal, model ini tetap menunjukkan kemampuan yang cukup baik sebagai alat bantu diagnostik dini.

3.4 Hasil Implementasi Saliency Maps

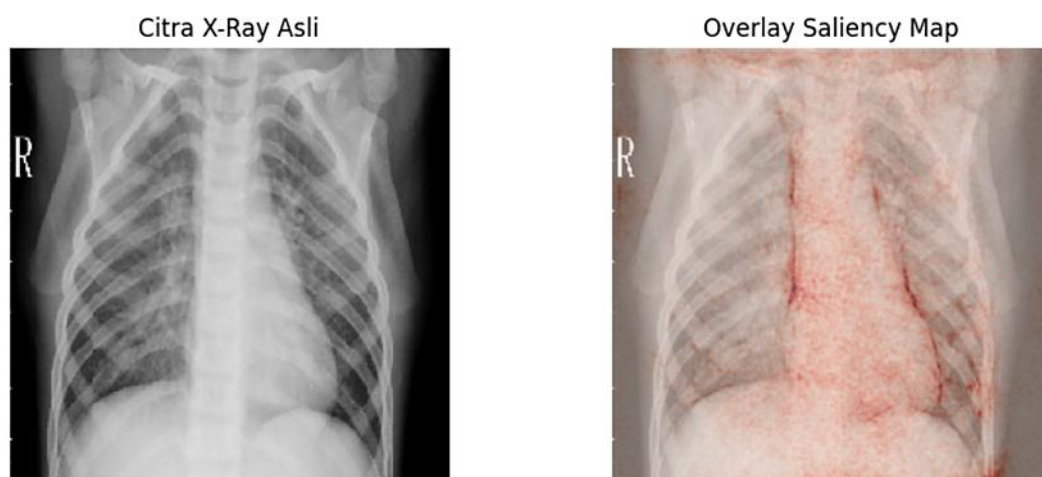
Untuk memberikan pemahaman tentang bagaimana model *deep learning* mengidentifikasi gambar X-ray dada, bab ini membahas hasil penggunaan pendekatan XAI dengan memanfaatkan *saliency maps*. *Saliency maps* digunakan untuk secara visual menggambarkan area gambar yang memengaruhi hasil prediksi model. Sebagai hasilnya, pengambilan keputusan model dievaluasi berdasarkan baik keterbacaannya maupun pemahamannya, serta hasil numeriknya. Gambar 5, gambar 6, dan gambar 7 menunjukkan hasil visualisasi *saliency maps* pada gambar X-Ray dada kategori covid-19, paru-paru normal, dan pneumonia secara berturut-turut.



Gambar 5. Hasil Visualisasi Saliency Maps pada gambar Paru-paru Covid-19



Gambar 6. Hasil Visualisasi Saliency Maps pada gambar Paru-paru Normal



Gambar 7. Hasil Visualisasi Saliency Maps pada gambar Paru-paru Pneumonia



Pendekatan *saliency maps* dapat menawarkan penjelasan visual yang jelas untuk penilaian model *deep learning*, berdasarkan temuan dari penerapan *saliency maps* pada gambar X-ray dada untuk kategori covid-19, paru-paru normal, dan pneumonia. Gambar visual yang dihasilkan menunjukkan bahwa model lebih fokus pada wilayah paru-paru yang signifikan secara medis, terutama yang menunjukkan masalah pada kasus covid-19 dan pneumonia.

Wilayah tengah hingga bawah paru-paru, yang biasanya terkait dengan munculnya opasitas, infiltrat, atau pola abnormal lainnya, menerima lebih banyak penekanan pada *saliency maps* dalam kelas covid-19 dan pneumonia. Ini menunjukkan bahwa model membuat keputusan berdasarkan karakteristik visual yang konsisten dengan pengetahuan medis selain mencapai hasil klasifikasi yang baik. Sebaliknya, model menunjukkan perhatian yang tersebar luas pada kelas paru-paru normal, tanpa satu lokasi pun menjadi fokus utama. Pola ini menunjukkan bahwa model dapat secara andal mengidentifikasi ketiadaan ciri penyakit, mendukung pilihan klasifikasi yang menunjukkan keadaan normal.

Secara keseluruhan, temuan visualisasi menunjukkan bahwa menggunakan *saliency maps* dapat meningkatkan interpretabilitas model saat mengklasifikasikan gambar X-ray dada. Memahami proses pengambilan keputusan model menjadi lebih mudah melalui visualisasi XAI yang dihasilkan, yang dapat menyoroti area penting dalam gambar yang sesuai dengan hasil prediksi model (Muliani et al., 2025). Nilai akurasi tinggi dari model tidak menjamin bahwa fitur yang relevan secara terapeutik digunakan dalam proses klasifikasi (Fauziyyah et al., 2025). Oleh karena itu, visualisasi *saliency maps* diperlukan untuk membantu memahami area gambar yang menjadi fokus model dalam membuat prediksi. Model ini tidak lagi menjadi misteri berkat representasi ini, yang juga mempermudah pemahaman proses pengambilan keputusan. Akibatnya, model ini dapat digunakan sebagai sistem pendukung untuk membantu diagnosis gangguan paru-paru.

4. KESIMPULAN

Jelas dari temuan studi bahwa pendekatan *deep learning* menggunakan arsitektur VGG16 melakukan pekerjaan yang sangat baik dalam mengklasifikasikan gambar X-ray dada menjadi tiga kelompok: covid-19, paru-paru normal, dan pneumonia. Model yang dibuat menunjukkan kemampuan beradaptasi yang baik terhadap data uji dan dapat mempelajari pola visual yang signifikan dari foto medis. Menurut hasil evaluasi model, semua kelas memiliki nilai *precision*, *recall*, dan *f1-score* yang baik, dengan tingkat *accuracy* sebesar 96%. Hal ini menunjukkan bahwa model dapat mengkategorikan setiap kelas dengan kesalahan yang relatif sedikit selain akurat secara keseluruhan. Meskipun masih terdapat beberapa kesalahan dalam membedakan antar kelas, terutama antara pneumonia dan covid-19 karena karakteristik visual kedua penyakit yang mirip, tabel *confusion matrix* yang dihasilkan menunjukkan bahwa sebagian besar data uji berhasil diidentifikasi dengan benar. Penelitian ini menggunakan *saliency maps* untuk menilai interpretabilitas model selain kemampuan klasifikasinya. Hasil visualisasi menunjukkan bahwa model lebih fokus pada daerah paru-paru yang relevan, terutama yang menunjukkan tanda-tanda covid-19 atau pneumonia. Sebaliknya, model membagi perhatiannya secara merata tanpa menekankan area tertentu pada gambar paru-paru yang sehat. Oleh karena itu, penelitian ini menunjukkan bahwa penggunaan pendekatan *accuracy* bersamaan dengan arsitektur VGG16 tidak hanya mampu mengembangkan model klasifikasi yang berkinerja baik, tetapi juga mampu meningkatkan transparansi dan kredibilitas model dalam analisis citra medis. Model yang dikembangkan berpotensi digunakan sebagai sistem pendukung diagnosis untuk membantu tenaga medis dalam melakukan identifikasi awal penyakit pernapasan. Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan. Visualisasi *saliency maps* yang dihasilkan belum sepenuhnya berfokus pada area paru-paru dan masih menunjukkan perhatian pada beberapa bagian citra yang tidak relevan. Selain itu, penelitian ini hanya menggunakan dataset tertentu sehingga performa model pada data medis yang lebih beragam masih perlu diuji lebih lanjut. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan dataset yang lebih bervariasi serta melakukan evaluasi interpretabilitas model secara lebih mendalam.

REFERENCES

- Arsenault, P. D., Wang, S., & Patenaude, J. M. (2025). A Survey of Explainable Artificial Intelligence (XAI) in Financial Time Series Forecasting. *ACM Computing Surveys*, 57(10). <https://doi.org/10.1145/3729531>
- Arun, N., Gaw, N., Singh, P., Chang, K., Aggarwal, M., Chen, B., Hoebel, K., Gupta, S., Patel, J., Gidwani, M., Adebayo, J., Li, M. D., & Kalpathy-Cramer, J. (2021). Assessing the trustworthiness of saliency maps for localizing abnormalities in medical imaging. *Radiology: Artificial Intelligence*, 3(6). <https://doi.org/10.1148/ryai.2021200267>
- Azizah, F. N., & Juniati, D. (2021). Analisis Jenis Penyakit Paru-Paru Berdasarkan Chest X-Ray Menggunakan Metode Fuzzy C-Means. *MATHunesa: Jurnal Ilmiah Matematika*, 9(2), 322–331. <https://doi.org/10.26740/mathunesa.v9n2.p322-331>
- Fauziyyah, L. N., Negara, B. S., Irsyad, M., Iskandar, I., & Yanto, F. (2025). Interpreting Lung Disease Detection from Chest X-rays Using Layer-wise Relevance Propagation (LRP). *Journal of Artificial Intelligence and Software Engineering*, 5(2), 697–708. <https://doi.org/10.30811/jaise.v5i2.7043>
- Gupta, J., Pathak, S., & Kumar, G. (2022). Deep Learning (CNN) and Transfer Learning: A Review. *Journal of Physics: Conference Series*, 2273(1). <https://doi.org/10.1088/1742-6596/2273/1/012029>
- Idhom, M., Prasetya, D. A., Riyantoko, P. A., Fahrudin, T. M., & Sari, A. P. (2023). Pneumonia Classification Utilizing VGG-16 Architecture and Convolutional Neural Network Algorithm for Imbalanced Datasets. *TIERS Information*



- Technology Journal*, 4(1), 73–82. <https://doi.org/10.38043/tiers.v4i1.4380>
- Jain, N., Choudhury, A., Sharma, J., Kumar, V., De, D., & Tiwari, R. (2020). A review of novel coronavirus infection (Coronavirus Disease-19). *Global Journal of Transfusion Medicine*, 5(1), 22. https://doi.org/10.4103/gjtm.gjtm_24_20
- Karar, M. E., Hemdan, E. E. D., & Shouman, M. A. (2021). Cascaded deep learning classifiers for computer-aided diagnosis of COVID-19 and pneumonia diseases in X-ray scans. *Complex and Intelligent Systems*, 7(1), 235–247. <https://doi.org/10.1007/s40747-020-00199-4>
- Kim, H. E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M. E., & Ganslandt, T. (2022). Transfer learning for medical image classification: a literature review. *BMC Medical Imaging*, 22(1), 1–13. <https://doi.org/10.1186/s12880-022-00793-7>
- Kim, S. C., & Cho, Y. S. (2022). Predictive System Implementation to Improve the Accuracy of Urine Self-Diagnosis with Smartphones: Application of a Confusion Matrix-Based Learning Model through RGB Semiquantitative Analysis. *Sensors*, 22(14). <https://doi.org/10.3390/s22145445>
- M.Kranthi, S.Sailaja, & E.V.N.Jyothi. (2024). Deep Learning Approaches for Medical Image Processing in the Big Data Era. *International Journal of Scientific Methods in Computational Science and Engineering*, 01(01), 24–31. <https://doi.org/10.58599/ijsmcse.2024.1108>
- Ma'ruf, M. drg. O. P. M. dr. A. (2020). *Profil Kesehatan Indonesia 2020*.
- Muliani, S., Sukma Negara, B., Irsyad, M., & Iskandar, I. (2025). Application of Shapley Additive Explanations (SHAP) in Deep Learning for Lung Disease Detection Using X-ray Images. *Journal of Artificial Intelligence and Software Engineering*, 5(2), 709–719. <https://doi.org/10.30811/jaise.v5i2.7044>
- Nguyen, Q. H., Ly, H. B., Ho, L. S., Al-Ansari, N., Van Le, H., Tran, V. Q., Prakash, I., & Pham, B. T. (2021). Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. *Mathematical Problems in Engineering*, 2021. <https://doi.org/10.1155/2021/4832864>
- Rahman, M. M., Matsuo, K., Matsuzaki, S., & Purushotham, S. (2021). DeepPseudo: Pseudo Value Based Deep Learning Models for Competing Risk Analysis. *35th AAAI Conference on Artificial Intelligence, AAAI 2021*, 1, 479–487. <https://doi.org/10.1609/aaai.v35i1.16125>
- Saturi, S., & Banda, S. (2024). Advanced Lung Disease Detection and Classification Using Ge-U-Net-ODLwith Gabor Filters and Entropy-Based Feature Selection. *Journal of Sensors, IoT & Health Sciences*, 2(2), 69–86. <https://doi.org/10.69996/jsihs.2024011>
- Shahzad, A., Arshed, M. A., Liaquat, F., Tanveer, M., Hussain, M., & Alamdar, R. (2022). Pneumonia Classification from Chest X-ray Images Using Pre-Trained Network Architectures. *VAWKUM Transactions on Computer Sciences*, 10(2), 34–44. <https://doi.org/10.21015/vtcs.v10i2.1271>
- Shastri, S., Kansal, I., Kumar, S., Singh, K., Popli, R., & Mansotra, V. (2022). CheXImageNet: a novel architecture for accurate classification of Covid-19 with chest x-ray digital images using deep convolutional neural networks. *Health and Technology*, 12(1), 193–204. <https://doi.org/10.1007/s12553-021-00630-x>
- Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). Deep inside convolutional networks: Visualising image classification models and saliency maps. *2nd International Conference on Learning Representations, ICLR 2014 - Workshop Track Proceedings*, 1–8.
- Sofiyah, W., Negara, B. S., Irsyad, M., Iskandar, I., & Yanto, F. (2025). Lung Disease Detection Using Gradient-Weighted Class Activation Mapping (Grad-CAM). *Journal of Artificial Intelligence and Software Engineering*, 5(2), 720–730. <https://doi.org/10.30811/jaise.v5i2.7041>
- Subhash, B. (2022). *Explainable AI: Saliency Maps*. Medium.
- Wollek, A., Graf, R., Čečátka, S., Fink, N., Willem, T., Sabel, B. O., & Lasser, T. (2023). Attention-based Saliency Maps Improve Interpretability of Pneumothorax Classification. *Radiology: Artificial Intelligence*, 5(2). <https://doi.org/10.1148/ryai.220187>