Journal of Artificial Intelligence and Software Engineering
Vol. 5, No. 2, Juni 2025, pp. 686~696
E-ISSN: 2777-001X, DOI: 10.30811/jaise.v5i2.7042

Applying Local Interpretable Model-agnostic Explanations (LIME) for Interpretable Deep Learning in Lung Disease Detection

Sherly Ananda[1], Benny Sukma Negara[2], Muhammad Irsyad[3], Jasril[4], Iwan Iskandar[5]
[1,2,3,4,5] Teknik Informatika, Fakultas Sains dan Teknologi, UIN Sultan Syarif Kasim Riau, Pekanbaru, 28293, Indonesia

Informasi Artikel

Diterima : 30 Mei 2025
Revisi : 06 Juni 2025
Publikasi : 20 Juni 2025

Kata Kunci:

Explainable AI
LIME
ResNet18
COVID-19
Pneumonia

ABSTRAK

Artificial Intelligence (AI) semakin banyak diterapkan dalam bidang kesehatan melalui model *Machine Learning* (ML) dan *Deep Learning* (DL). Namun, kompleksitas model *modern* yang bersifat *black-box* menimbulkan kebutuhan akan metode interpretasi yang transparan. *Explainable* AI (XAI) hadir untuk menjembatani hal tersebut, dengan memberikan pemahaman yang lebih baik terhadap kinerja model. Penelitian ini mengimplementasikan metode *Local Interpretable Model-agnostic Explanations* (LIME) untuk memvisualisasikan hasil klasifikasi model DL berbasis arsitektur ResNet18 terhadap citra *Chest X-ray* (CXR) pada tiga kelas: normal, COVID-19, dan pneumonia. Model mencapai *precision* 97%, *recall* 97%, dan *F1-score* 97%, serta *Accuracy* sebesar 98%. Visualisasi LIME menunjukkan area citra yang berkontribusi signifikan terhadap klasifikasi, serta mampu membedakan ketiga kelas dengan baik. Hasil dari penelitian ini menunjukkan bahwa penerapan XAI, khususnya LIME dengan model DL berbasis ResNet18 mampu memberikan interpretabilitas dalam tugas klasifikasi citra CXR.

ABSTRACT

Artificial Intelligence (AI) is increasingly being applied in the healthcare field through Machine Learning (ML) and Deep Learning (DL) models. However, the complexity of modern black-box models creates a need for transparent interpretation methods. Explainable AI (XAI) emerges to bridge this gap by providing better understanding of model performance. This study implements the Local Interpretable Model-agnostic Explanations (LIME) method to visualize the classification results of a DL model based on the ResNet18 architecture on Chest X-ray (CXR) images across three classes: normal, COVID-19, and pneumonia. The model achieved a precision of 97%, recall of 97%, and F1-score of 97%, with an accuracy of 98%. LIME visualizations highlight the image regions that significantly contribute to the classification and effectively distinguish among the three classes. The results of this study demonstrate that applying XAI specifically LIME with a ResNet18-based DL model can provide interpretability in CXR image classification tasks.

Penulis Koresponden

Email: bsnegara@uin-suska.ac.id

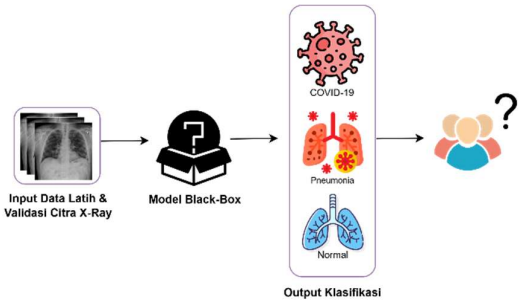
Cara sitasi IEEE::

S. Ananda, B.S. Negara, M. Irsyad, Jasril, I. Iskandar "*Applying Local Interpretable Model-agnostic Explanations* (LIME) *for Interpretable Deep Learning in Lung Disease Detection*," *Journal of Artificial*

Intelligence and Software Engineering (J-AISE), vol. 5, no. 2, p. 686-696, Juni 2025. doi:10.30811/jaise.v5i2.7042

1. PENDAHULUAN

Kecerdasan buatan telah mengalami pertumbuhan yang signifikan dalam berbagai bidang selama beberapa tahun terakhir. Dalam perkembangan tersebut, berbagai metode telah dilaporkan memanfaatkan model ML dan DL. Namun, mayoritas model-model tersebut bersifat kompleks dan tidak memberikan informasi yang jelas mengenai proses pengambilan keputusan yang dilakukan, sehingga sering disebut sebagai model *black-box*. Kondisi ini menimbulkan tantangan tersendiri, khususnya dalam konteks penerapan pada bidang-bidang yang bersifat kritikal seperti kesehatan, di mana interpretabilitas dan transparansi keputusan sistem menjadi aspek yang sangat penting untuk memastikan kepercayaan serta keselamatan pengguna akhir [1].



Gambar 1. Model *black-box* tidak transparan dalam proses pengambilan keputusan

Di bidang kesehatan, penyakit paru-paru memiliki berbagai variasi dengan gejala klinis yang sering kali serupa, sehingga menyulitkan masyarakat dalam mengidentifikasi jenis penyakit yang diderita secara akurat. Salah satu jenis penyakit paru-paru yang umum dijumpai adalah pneumonia, yaitu infeksi akut pada saluran pernapasan bagian bawah yang disebabkan oleh adanya peradangan pada jaringan dan alveolus (kantung udara) di paru-paru [2]. COVID-19, yang disebabkan oleh virus *SARS-CoV-2*, merupakan penyakit paru-paru yang menjadi sorotan global sejak pertama kali diidentifikasi di Wuhan, Tiongkok pada Desember 2019 [3].

Artificial *Neural Network* (ANN) sebagai salah satu terobosan utama ML, model yang terinspirasi dari cara kerja sistem saraf biologis, mampu melampaui performa pendekatan AI konvensional dalam berbagai tugas pembelajaran mesin. Salah satu arsitektur ANN yang paling menonjol adalah *Convolutional Neural Network* (CNN), yang secara luas digunakan dalam pengenalan pola visual, khususnya pada citra. CNN dikenal memiliki struktur yang efisien namun efektif, menjadikannya sebagai titik awal yang ideal dalam pemodelan jaringan saraf [4][5]. Salah satu varian arsitektur CNN yang paling banyak digunakan adalah *Residual Neural Network* (ResNet). Meskipun berasal dari prinsip dasar yang sama, arsitektur CNN seperti ResNet memiliki variasi dalam jumlah lapisan dan parameter yang digunakan [6]. Penelitian oleh Salih Sarp dkk [7] menunjukkan bahwa model ResNet mampu mencapai performa klasifikasi yang sangat tinggi, dengan skor F1 sebesar 98%. Temuan ini mengindikasikan bahwa ResNet merupakan salah satu metode paling efektif dan akurat dalam mendeteksi COVID-19 pada citra rontgen dada. Seiring dengan meningkatnya kompleksitas model ML modern yang bersifat *black-box*, kebutuhan akan metode interpretasi yang transparan menjadi semakin mendesak. Hal ini mendorong berkembangnya bidang XAI, yang bertujuan untuk menyediakan pemahaman yang lebih baik mengenai cara kerja model ML bagi pengguna dan pengambil keputusan [8]. XAI menjadi sangat relevan terutama pada model berarsitektur kompleks seperti jaringan saraf dalam, di mana transparansi dalam pengambilan keputusan merupakan aspek penting [9]. Salah satu pendekatan XAI yang paling dikenal adalah LIME, yang dirancang untuk meningkatkan interpretabilitas model *black-box*. LIME bekerja dengan membangun model lokal yang dapat ditafsirkan di sekitar prediksi tunggal, melalui pembuatan data sintetis dengan gangguan acak untuk mengidentifikasi fitur-fitur yang paling berpengaruh terhadap hasil prediksi [10].

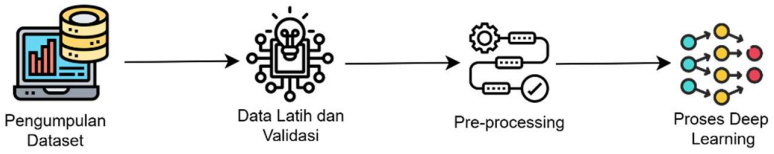
Pendekatan berbasis CNN dengan integrasi metode LIME telah berhasil diterapkan dalam berbagai studi. Penelitian Mesut Togacar dkk [11] menunjukkan bahwa kombinasi CNN dan LIME mampu membedakan COVID-19 dari pneumonia akibat infeksi bakteri dan virus secara akurat melalui analisis citra

medis. Selain itu, penelitian oleh Natasha Nigar dkk [12] mengembangkan model berbasis ResNet-18 yang dilatih menggunakan teknik *transfer learning* pada dataset ISIC 2019. Model tersebut mampu mengklasifikasikan delapan jenis lesi kulit secara akurat, dengan masing-masing nilai akurasi, presisi, *recall*, dan *F1-score* sebesar 94,47%, 93,57%, 94,01%, dan 94,45%. LIME digunakan dalam studi tersebut untuk memberikan visualisasi penjelasan yang mendukung pengambilan keputusan secara rasional, serta mengungkapkan generalisasi dan potensi bias yang dimiliki oleh model terhadap data citra yang digunakan.

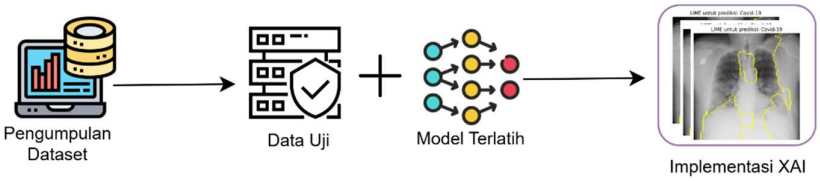
Penelitian ini memberikan sejumlah kontribusi penting dalam pengembangan sistem klasifikasi citra medis berbasis DL yang bersifat *interpretable*. Adapun kontribusi yang dapat diidentifikasi dari penelitian ini adalah penelitian ini menggunakan arsitektur ResNet18 untuk melakukan klasifikasi citra CXR ke dalam tiga kelas, yaitu normal, COVID-19, dan pneumonia. Penelitian ini mengintegrasikan metode LIME sebagai pendekatan XAI untuk memberikan interpretasi terhadap hasil klasifikasi model DL.

2. METODE

Pelaksanaan sebuah penelitian ilmiah menuntut adanya rancangan metodologis yang sistematis dan terarah. Metodologi penelitian tidak hanya berperan sebagai kerangka kerja konseptual, tetapi juga menjadi fondasi strategis dalam mengarahkan setiap tahapan penelitian secara berurutan. Pada penelitian ini, implementasi dilakukan menggunakan bahasa pemrograman *Python* dalam lingkungan sistem operasi *Windows* 11.



Gambar 2. Tahapan. klasifikasi

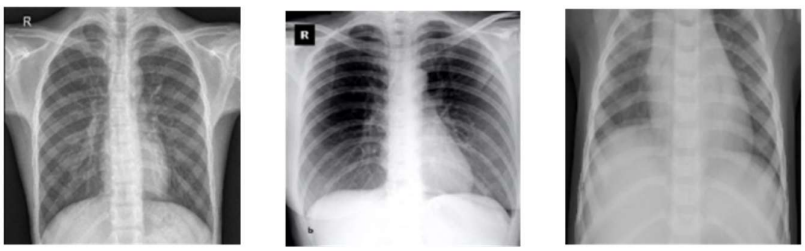


Gambar 3. Tahapan implementasi XAI

Gambar 2 dan 3 menunjukkan bahwa penelitian ini terdiri dari dua tahapan utama. Tahap pertama adalah proses klasifikasi, yang dimulai dengan pengumpulan data latih dan validasi, dilanjutkan dengan *pre-processing* citra CXR dan pelatihan model menggunakan *deep learning*. Tahap kedua adalah pengujian, di mana data uji diproses bersama model yang telah dilatih untuk menghasilkan prediksi, yang kemudian dijelaskan secara visual menggunakan metode *Explainable AI* (XAI).

2.1 Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan dengan mencari data yang bersifat *public* yang di dapat dari Sachin Kumar dkk [13]. Data yang dikumpulkan merupakan gambar citra CXR dada normal, pneumonia dan COVID-19 sebanyak 3009 citra yang mana 3000 untuk data latih dan validasi terbagi rata untuk setiap kelasnya. Setiap kelas terdiri dari sebagai berikut: normal 1000 gambar, pneumonia 1000 gambar dan COVID-19 1000 gambar. Kemudian 9 gambar dari 3 kelas yang berbeda untuk data uji yang digunakan untuk implementasi XAI. Berikut adalah contoh gambar dari 3 kelas citra CXR yang digunakan:



(a) (b) (c)

Gambar 4. Gambar CXR paru-paru (a) Normal, (b) COVID-19 dan (c) Pneumonia

2.2 *Pre-processing*

2.2.1 *Resize*

Pada tahap ini, dilakukan penyesuaian terhadap dimensi citra masukan menjadi 224 × 224 piksel. Penyesuaian ini bertujuan untuk memastikan kompatibilitas citra dengan arsitektur *deep learning* yang digunakan, sekaligus memfasilitasi efisiensi komputasi selama proses pelatihan dan validasi. Selain itu, perubahan ukuran ini dirancang untuk menyesuaikan format citra dengan standar *input* yang umum digunakan dalam berbagai model *deep learning*, khususnya arsitektur CNN, yang secara umum menerima citra berdimensi 224 × 224 piksel sebagai masukan *default*.

2.2.2 *Data Splitting*

Pembagian data (*data splitting*) merupakan langkah fundamental dalam proses pelatihan model pembelajaran mesin. Dataset biasanya dipisahkan menjadi dua subset utama: data pelatihan untuk membangun model dan data validasi untuk mengevaluasi kinerjanya. Rasio pembagian dapat bervariasi, seperti 90:10, 80:20, atau 70:30, tergantung pada kebutuhan dan ukuran dataset. Tujuan utama dari pembagian ini adalah untuk menghindari *overfitting* serta mengukur kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya [14].

2.3 Model *Deep Learning*

2.3.1 Pembangunan Model

Model yang dikembangkan dalam studi ini menggunakan salah satu arsitektur CNN yaitu ResNet18 yang diinisialisasi tanpa bobot pra-latih, sebagai upaya untuk memastikan bahwa proses pembelajaran dimulai dari kondisi netral tanpa pengaruh parameter eksternal. Adaptasi arsitektural dilakukan secara strategis dengan mengganti *layer fully connected* pada bagian akhir jaringan dengan *layer* linear yang disesuaikan dengan jumlah kelas target [15]. Selain itu, ditambahkan mekanisme *dropout* sebagai bentuk regularisasi untuk meminimalkan risiko *overfitting* dan meningkatkan kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya [16]. Proses pembangunan model dilakukan dengan menggunakan algoritma optimisasi Adam yang dipadukan dengan regularisasi bobot, bertujuan untuk menjaga stabilitas dan akurasi selama proses pembaruan parameter [17]. Sebagai strategi lanjutan untuk mengontrol dinamika pembelajaran, diterapkan penjadwalan *learning rate*, di mana *learning rate* awal ditetapkan dalam nilai tinggi guna mempercepat konvergensi awal, kemudian secara bertahap diturunkan untuk menyempurnakan parameter model. Pendekatan ini dirancang untuk menghasilkan model klasifikasi citra medis yang tidak hanya efisien, tetapi juga adaptif dan andal dalam menghadapi kompleksitas data visual klinis [18].

2.3.2 Pelatihan Model

Pelatihan model DL dalam klasifikasi citra CXR dilakukan melalui siklus *training* bertahap yang berulang selama sejumlah *epoch*. Pada awal setiap siklus, model diaktifkan dalam *mode* pelatihan untuk memproses seluruh *batch* citra, di mana setiap gambar dianalisis untuk menghasilkan prediksi awal[19] Jika terlalu banyak menggunakan *epoch* dapat menyebabkan *overfitting* dan terlalu sedikit menggunakan *epoch* dapat menyebabkan *underfitting*. Maka sangat penting untuk menentukan *epoch* yang sesuai [20]. Selama proses pelatihan, parameter model dioptimalkan secara iteratif guna meminimalkan fungsi *loss*. Nilai *loss* untuk data pelatihan dan validasi dievaluasi pada setiap *epoch* untuk memantau kinerja model. Model disimpan secara otomatis setiap kali terjadi penurunan nilai *loss* pada data validasi, sebagai indikasi peningkatan generalisasi terhadap data yang belum terlihat [21].

2.3.3 Evaluasi Model

		Actual/ Sebenarnya	
		Positif	Negatif
Prediksi	Positif	**TP** (True Positive)	**FP** (False Positive)
	Negatif	**TN** (True Negative)	**FN** (False Negative)

Gambar 5. Tabel *confusion matrix*

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

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

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

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

Confusion matrix merupakan alat evaluasi klasik yang telah lama digunakan dalam pengujian performa model klasifikasi, baik dalam ranah ilmiah maupun penerapan teknis. Matriks ini menjadi komponen esensial dalam berbagai disiplin seperti penglihatan komputer, pemrosesan bahasa alami, dan pengenalan suara. Dalam bentuk paling dasar, *confusion matrix* digunakan untuk menggambarkan hasil klasifikasi *biner* melalui tabel dua baris dua kolom, yang merepresentasikan distribusi empat kemungkinan keluaran model: *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), dan *False Negative* (FN). Masing-masing komponen ini memberikan informasi penting mengenai tingkat keberhasilan dan kegagalan model dalam membedakan antara kelas yang benar dan salah [22]

2.4 Implementasi XAI

XAI merupakan cabang penelitian yang berfokus pada upaya mengungkap dan menjelaskan proses pengambilan keputusan oleh sistem kecerdasan buatan [23]. Seiring meningkatnya kompleksitas arsitektur model, terutama jaringan saraf dalam, XAI muncul sebagai elemen krusial dalam ekosistem pembelajaran mesin modern, menjawab kebutuhan akan transparansi dan akuntabilitas dalam proses prediksi [9]. Beragam pendekatan XAI telah dikembangkan untuk menelusuri mekanisme internal model, dengan tujuan menginterpretasikan bagaimana jaringan saraf memproses dan merespons data masukan, sehingga memberikan wawasan yang dapat dipahami oleh pengguna maupun pengambil keputusan [24].

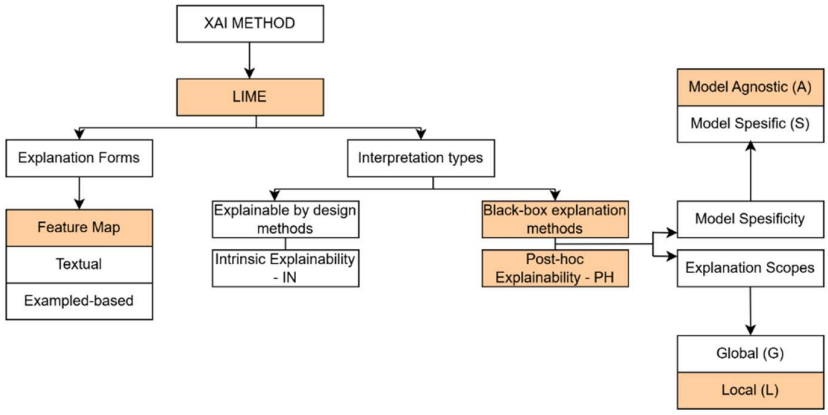
2.4.1 *Local Interpretable Model-agnostic Explanations* (LIME)

Local *Interpretable Model-Agnostic Explanations* (LIME) merupakan pendekatan yang menjelaskan prediksi model kompleks dengan membangun model sederhana di sekitar data yang dianalisis. Metode ini bersifat *model-agnostik*, sehingga dapat digunakan pada berbagai jenis model, termasuk yang tidak dapat diakses strukturnya atau *black-box*. LIME fokus pada penjelasan satu prediksi secara lokal, sehingga menghasilkan interpretasi visual yang mudah dipahami [25]. Untuk menghasilkan penjelasan terhadap suatu prediksi, LIME membangkitkan sejumlah data sintetis yang menyerupai contoh asli dan mengevaluasinya menggunakan model *black-box*. Hasil prediksi dari model tersebut kemudian digunakan sebagai data pelatihan bagi model linier sederhana yang dapat diinterpretasikan. Model linier ini bertindak sebagai penjelas lokal yang mendekati perilaku model kompleks di sekitar titik data tersebut. Dari model ini, LIME membentuk peta *saliency* yang menyoroti area pada gambar yang paling berpengaruh terhadap prediksi yang dihasilkan [26].

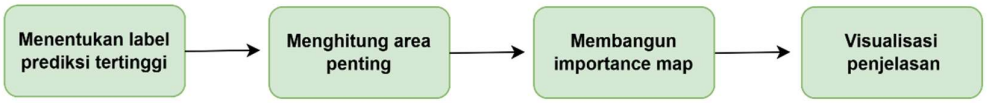
Penjelasan mengenai berbagai pendekatan interpretabilitas model kotak hitam dapat dirangkum sebagai berikut. Feature map memungkinkan sistem untuk mengidentifikasi dan menyoroti area-area spesifik dari input yang paling berpengaruh terhadap keputusan akhir [25]. Visualisasinya biasanya ditampilkan dalam bentuk citra asli yang dilapisi dengan peta keunggulan, yang merepresentasikan intensitas kontribusi tiap wilayah terhadap hasil prediksi. Sementara itu, pendekatan post-hoc bertujuan menginterpretasikan model berarsitektur kompleks, seperti jaringan saraf, dengan menggunakan teknik eksternal untuk memperoleh wawasan interpretatif terhadap model yang telah dilatih [8][25]. Kemudian, pendekatan model-agnostic berfungsi untuk menjelaskan berbagai jenis model kotak hitam tanpa memerlukan informasi tentang struktur internal model. Pendekatan ini juga termasuk dalam kategori post-hoc karena pemahamannya diperoleh melalui analisis output yang dihasilkan dari gangguan terkontrol pada data masukan [8][25][27].

Selanjutnya, pendekatan local digunakan untuk menjelaskan alasan keputusan model terhadap suatu instance tertentu. Pendekatan ini menganggap model sebagai entitas non-transparan dan menitikberatkan pada variabel-variabel lokal yang paling berpengaruh terhadap hasil prediksi [8][25]. Adapun metode saliency map bertujuan menginterpretasikan keputusan model dengan menetapkan skor relevansi pada setiap komponen input, yang divisualisasikan dalam bentuk heatmap atau peta probabilistik, menyoroti area input paling informatif [28]. Terakhir, pendekatan feature importance merupakan salah satu bentuk penjelasan paling umum

dalam interpretasi lokal, dengan memberikan nilai kepentingan pada setiap fitur yang menunjukkan seberapa besar pengaruh fitur tersebut terhadap prediksi yang sedang dianalisis [29].



Gambar 6. Metode XAI LIME



Gambar 7. Tahapan LIME

Prosedur LIME yaitu LIME mengidentifikasi label yang memiliki probabilitas tertinggi berdasarkan hasil prediksi model terhadap citra masukan. LIME mengestimasi bagian-bagian citra yang paling relevan terhadap keputusan model dengan membandingkan pengaruh segmen gambar terhadap *output*. Hasil estimasi tersebut dituangkan dalam bentuk peta visual yang menunjukkan kontribusi tiap bagian gambar terhadap klasifikasi. Peta kepentingan divisualisasikan dengan menampilkannya sebagai *overlay* pada gambar asli, sehingga bagian yang berpengaruh terhadap keputusan model dapat dikenali secara intuitif. Berikut di bawah ini merupakan perhitungan pada operasi LIME:

$$argmin_{\text{!}\epsilon G}\{L(f, g, \pi) + \Omega(g)\} \quad (5)$$

Keterangan;

x: Sebuah contoh dari ruang data yang kita inginkan penjelasannya untuk nilai target yang diprediksi.
$L(f, g, \pi x)$: Fungsi fidelitas yang mengukur seberapa tidak setianya g saat mendekati f di lokalitas yang ditentukan oleh πx. Ini adalah kerugian yang menyadari lokalitas.
$\Omega(g)$: Mengukur kompleksitas model dari penjelas (g)
f: Model kotak hitam yang akan dijelaskan
g: Penjelas
G: Total set model yang dapat ditafsirkan
πx: Ukuran kedekatan [7]

3. HASIL DAN PEMBAHASAN

3.1 Hasil *Hyperarameter Testing*

Hyperparameter merupakan komponen penting yang menentukan konfigurasi arsitektural serta perilaku pelatihan jaringan saraf konvolusional (CNN). Tantangan utama dalam pengoptimalannya terletak pada pencarian kombinasi parameter yang optimal, yang tidak hanya memaksimalkan akurasi model, tetapi juga mempertimbangkan efisiensi waktu pelatihan [30]. Penelitian ini menerapkan *splitting* data pada dataset. *Splitting* data ini diterapkan pada proses *training dan validation*. Data *training* dan *validation* akan dibagi pada tahapan *deep learning* dengan rasio 90%: 10%, 80%: 20%, dan 70%:30% yang dibagi secara manual.

Tabel 1. *Splitting* data terbaik

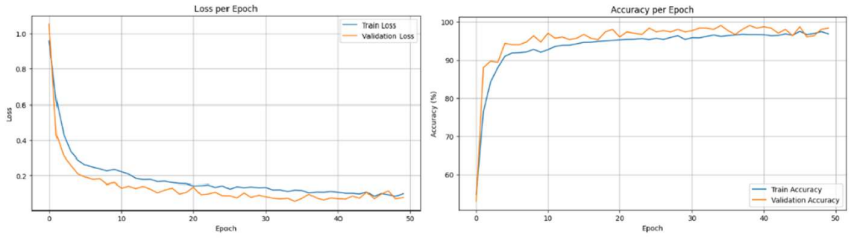
Dataset	Kelas	Data Latih	Data Validasi
X-ray Dada	Normal	900	100
	COVID-19	900	100
	Pneumonia	900	100
Total		**2700**	**300**

Berdasarkan hasil pengujian parameter pada tabel 1, konfigurasi optimal diperoleh pada skenario pembagian data sebesar 90% untuk pelatihan dan 10% untuk validasi.

Tabel 2. Hasil hyperparameter testing

No.	Splitting Data		Epoch	Learning Rate	Loss %		Accuracy%	
	Latih	Validasi			Latih	Validasi	Latih	Validasi
Percobaan 1	70%	30%	25	0,001	0.1216	0.0880	96.25%	96.83%
Percobaan 2	70%	30%	25	0,0001	0.0812	0.0962	97.25%	96.33%
Percobaan 3	70%	30%	25	0,00001	0.1482	0.1096	95.25%	96.50%
Percobaan 4	70%	30%	50	0.001	0.4507	0.3568	83.83%	90.00%
Percobaan 5	70%	30%	50	0.0001	0.0679	0.1069	97.33%	96.17%
Percobaan 6	70%	30%	50	0,00001	0.0324	0.0671	98.88%	97.83%
Percobaan 7	80%	20%	25	0,001	0.0842	0.0899	97.04%	96,83%
Percobaan 8	80%	20%	25	0,0001	0.2782	0.2272	91.21%	92.83%
Percobaan 9	80%	20%	25	0.00001	0.0625	0.0971	97.67%	96.78%
Percobaan 10	80%	20%	25	0.000001	0.1033	0.1008	96.62%	96.78%
Percobaan 11	80%	20%	50	0.001	0.1424	0.1083	95.29%	96.78%
Percobaan 12	80%	20%	50	0.0001	0.0710	0.3250	97.67%	86.89%
Percobaan 13	80%	20%	50	0.00001	0.0388	0.0921	98.52%	96.89%
Percobaan 14	80%	20%	50	0.000001	0.0987	0.0971	97.10%	96.56%
Percobaan 15	90%	10%	25	0.001	0.1245	0.0882	95.74%	97.67%
Percobaan 16	90%	10%	25	0.0001	0.0842	0.0638	97.15%	98.00%
Percobaan 17	90%	10%	25	0.00001	0.1442	0.0881	95.33%	98.33%
Percobaan 18	90%	10%	25	0.000001	0.4016	0.3378	85.67%	90.00%
Percobaan 19	90%	10%	50	0.001	0.0679	0.0861	97.37%	97.00%
Percobaan 20	90%	10%	50	0.0001	0.0249	0.0711	99.19%	98.00%
Percobaan 21	90%	10%	50	0.00001	0.0998	0.0777	96.81%	98.33%
Percobaan 22	90%	10%	50	0.000001	0.2711	0.2052	91.11%	92.67%

Berdasarkan hasil pada tabel 2, konfigurasi optimal diperoleh pada skenario dengan jumlah *epoch* sebanyak 50 dan *learning rate* sebesar 0.00001. Kombinasi parameter tersebut menghasilkan tingkat akurasi validasi tertinggi, yaitu sebesar 98,33%, yang mencerminkan kinerja prediktif model yang sangat baik dalam mengenali pola dari data yang dimasukkan.



Gambar 8. *Plot loss* dan *accuracy hyperparameter* terbaik

Gambar 8 mengilustrasikan performa model yang stabil selama proses pelatihan, ditunjukkan oleh *plot loss* dan *accuracy* yang seimbang antara data pelatihan dan validasi, tanpa indikasi *overfitting* maupun *underfitting*. Pola ini mencerminkan kemampuan generalisasi model yang baik terhadap data baru. Dengan demikian, grafik ini dapat dijadikan sebagai validasi bahwa pemilihan dan penerapan hyperparameter telah dilakukan secara tepat dan mendukung pembelajaran model secara optimal.

3.2 Metrik Evaluasi

Dari hasil *hyperparameter testing* yang dilakukan, didapatkan bahwasannya setiap kelas memiliki nilai *precision, recall, dan f1-score* yang bebeda untuk setiap dataset.

Tabel 3. Hasil evaluasi

Kelas	Precision	Recall	F1-Score	Support
COVID-19	100%	100%	100%	100
Normal	99%	93%	96%	100
Pneumonia	93%	99%	96%	100

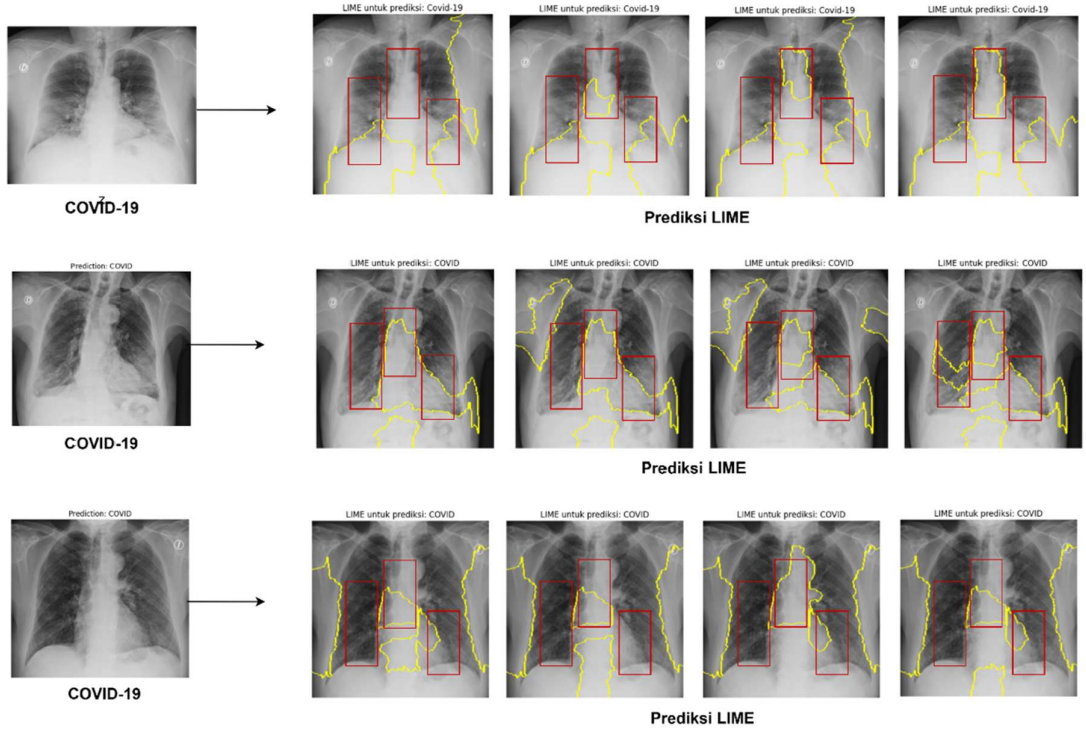
Berdasarkan tabel 3, bisa kita lihat bahwa kelas COVID-19 memiliki nilai lebih tinggi dari pada kelas lainnya. Ini menunjukkan bahwa model mengklasifikasikan dengan baik pada kelas COVID-19. Tetapi secara keseluruhan model telah mengklasifikasikan dengan baik dengan hasil nilai evaluasi pada kelas normal dan pneumonia hampir mendekati nilai kelas COVID-19.

3.3 Visualisasi LIME

Pada penelitian ini, untuk melihat hasil visualisasi LIME maka diambil 3 data per kelas tanpa di latih sebelumnya. Tujuannya adalah melihat performa dari model yang telah dibangun dan dilatih sebelumnya. Kemudian melihat apakah LIME menyoroti citra data uji dengan baik dan benar. Pada makalah ini akan ditampilkan visualisasi LIME sebanyak 3 data per kelas.

a. COVID-19

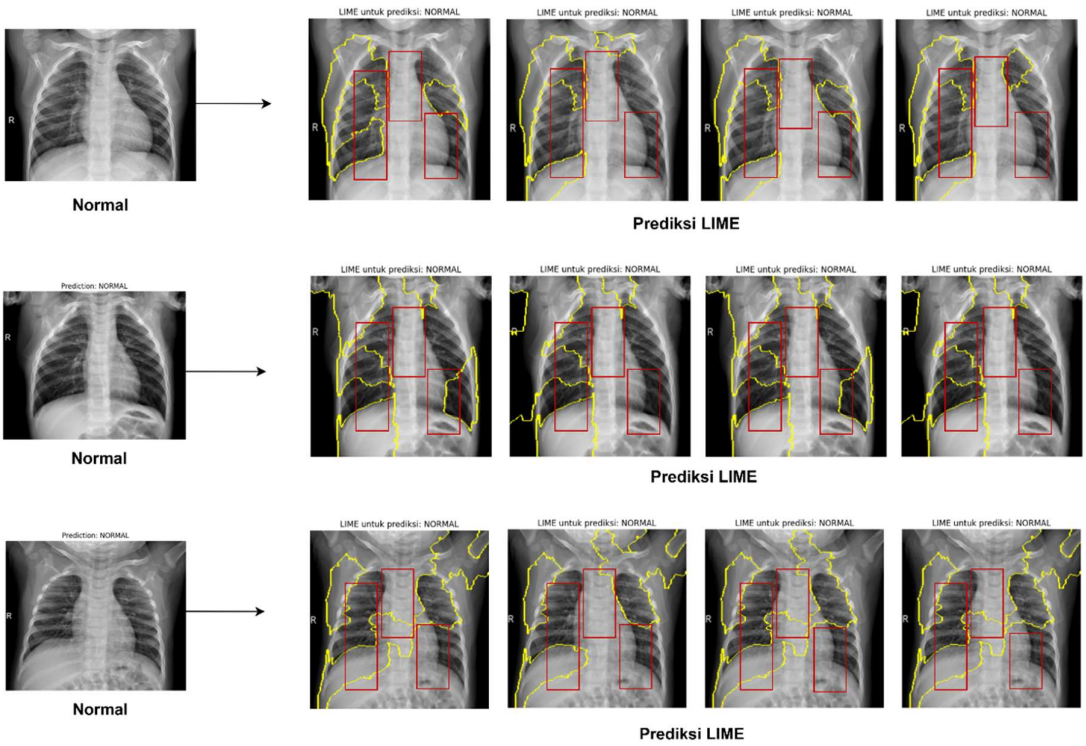
Berdasarkan hasil visualisasi, terlihat bahwa LIME menyoroti area lobus inferior pada kedua paru-paru, khususnya di zona perifer atau tepi luar paru-paru, sebagai kontributor utama dalam keputusan model untuk mengklasifikasikan citra ini sebagai COVID-19.



Gambar 9. Visualisasi LIME COVID-19

b. Normal

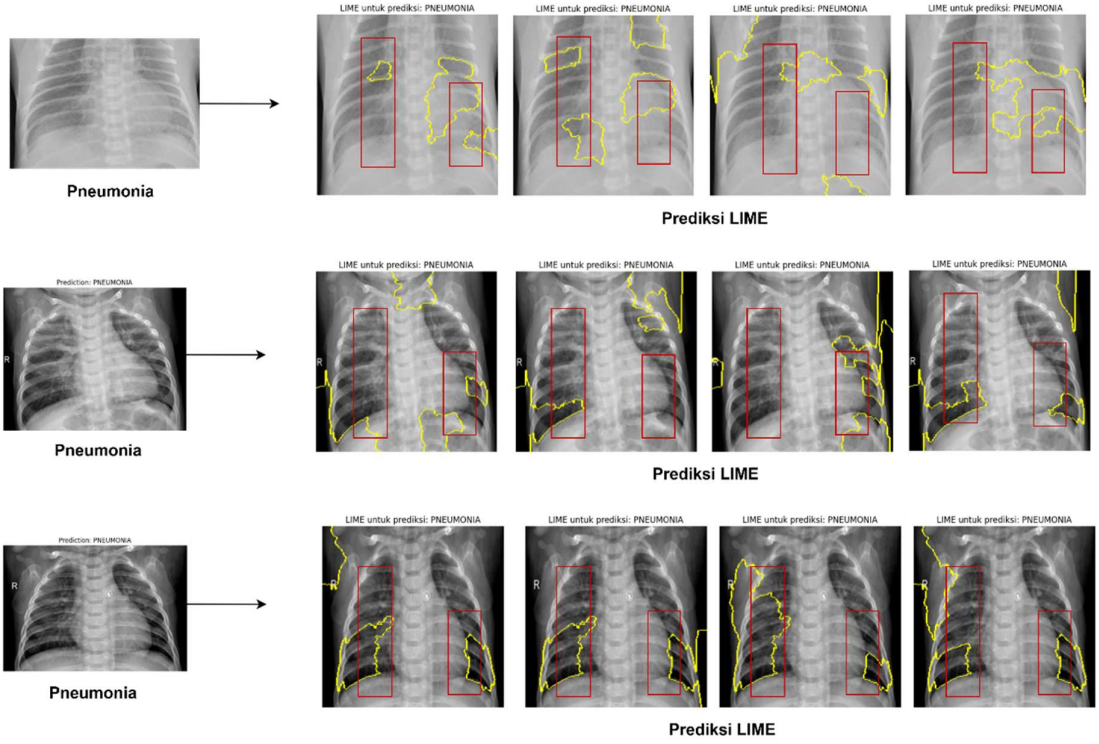
Berdasarkan hasil interpretasi LIME untuk prediksi normal, terlihat bahwa area yang disorot berada di bagian tengah sampai bawah dari kedua paru-paru. Area ini tampak simetris dan tidak menunjukkan tanda-tanda kelainan. LIME menunjukkan bahwa model menganggap bagian paru-paru yang terlihat bersih dan normal ini sebagai alasan utama dalam menentukan bahwa citra ini termasuk kategori normal.



Gambar 10. Visualisasi LIME Normal

c. Pneumonia

Berdasarkan hasil dari LIME untuk prediksi pneumonia, terlihat bahwa bagian yang disorot berada di beberapa area paru-paru bagian atas hingga tengah. LIME menunjukkan bahwa model mengandalkan area-area ini untuk memutuskan bahwa gambar ini menunjukkan pneumonia.



Gambar 11. Visualisasi LIME Pneumonia

Gambar 9, 10 dan 11 menyajikan hasil prediksi model pada data uji, yang dilengkapi dengan penjelasan menggunakan metode LIME. Pada masing-masing gambar, ditampilkan wilayah superpiksel berwarna kuning yang merepresentasikan kontribusi signifikan terhadap keputusan klasifikasi, serta kotak pembatas berwarna merah yang menunjukkan area *ground truth* lokasi COVID-19, normal, dan pneumonia pada paru-paru. Penjelasan LIME ini memberikan wawasan visual mengenai bagian citra yang dianggap penting oleh model dalam menghasilkan prediksi. Seluruh model yang dilatih berhasil mengklasifikasikan citra CXR pada ketiga gambar tersebut secara akurat ke dalam kelas. Hal ini mencerminkan kemampuan model dalam mendeteksi keberadaan penyakit serta mengekstraksi fitur relevan dari struktur paru-paru. Secara khusus, arsitektur ResNet18 menunjukkan konsistensi antara wilayah superpiksel LIME dan area *ground truth*, yang menandakan interpretasi yang valid. Namun demikian, masih terdapat superpiksel penjelas yang muncul di area luar paru-paru, yang tidak secara langsung berkaitan dengan patologi, sehingga menunjukkan bahwa meskipun interpretasi model sebagian besar tepat, terdapat ruang untuk perbaikan dalam meningkatkan fokus atensi model terhadap area klinis yang relevan.

4. KESIMPULAN

Berdasarkan hasil interpretasi menggunakan LIME, terdapat perbedaan pola perhatian model dalam mengklasifikasikan citra *CSR* dada sebagai normal, COVID-19, atau pneumonia. Pada citra dengan prediksi normal, area yang disorot oleh LIME berada di bagian tengah hingga bawah kedua paru-paru dan tampak simetris, menunjukkan bahwa model mengandalkan tampilan struktur paru-paru yang bersih sebagai dasar klasifikasi. Sebaliknya, pada kasus COVID-19, area yang disorot cenderung berada di bagian perifer dan bawah paru-paru. Sementara itu, pada prediksi pneumonia, LIME menyoroti beberapa area di paru-paru bagian atas hingga tengah,

Penelitian ini menunjukkan bahwa integrasi metode XAI, khususnya LIME, dengan model DL berbasis ResNet18, mampu meningkatkan interpretabilitas dalam tugas klasifikasi citra CXR. Model yang dibangun berhasil mencapai performa yang tinggi. Melalui visualisasi LIME, area-area penting pada citra yang menjadi dasar keputusan model dapat diidentifikasi secara jelas, sehingga memberikan gambaran yang lebih transparan mengenai bagaimana model membedakan antara kondisi normal, COVID-19, dan pneumonia. Temuan ini menegaskan bahwa pendekatan XAI seperti LIME berperan penting dalam menjembatani kesenjangan antara performa dan interpretasi model, serta berpotensi mendukung kepercayaan dan adopsi sistem berbasis AI dalam konteks diagnosis medis. Namun demikian, sebagai pihak non-medis, peneliti tidak memiliki otoritas untuk memastikan sejauh mana hasil interpretasi dari metode LIME selaras dengan keputusan klinis yang diambil oleh tenaga medis profesional. Dengan demikian, disarankan agar penelitian di masa mendatang melibatkan proses validasi langsung oleh ahli radiologi guna memastikan akurasi, relevansi, dan kredibilitas interpretasi yang dihasilkan oleh metode tersebut. Penelitian selanjutnya juga disarankan

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada dosen pembimbing dan penguji yang telah memberi arahan dan dukungan dalam proses pelaksanaan penelitian sehingga dapat terselesaikan dengan baik.

REFERENSI

[1] V. Hassija *et al.*, “Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence,” Jan. 01, 2024, *Springer*. doi: 10.1007/s12559-023-10179-8.

[2] I. Md. D. Maysanjaya, “Klasifikasi Pneumonia pada Citra X-rays Paru-paru dengan Convolutional Neural Network,” *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 9, no. 2, pp. 190–195, 2020, doi: 10.22146/jnteti.v9i2.66.

[3] M. Harahap, Em Manuel Laia, Lilis Suryani Sitanggang, Melda Sinaga, Daniel Franci Sihombing, and Amir Mahmud Husein, “Deteksi Penyakit Covid-19 Pada Citra X-Ray Dengan Pendekatan Convolutional Neural Network (CNN),” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 1, pp. 70–77, Feb. 2022, doi: 10.29207/resti.v6i1.3373.

[4] A. Saxena, “An Introduction to Convolutional Neural Networks,” *Int J Res Appl Sci Eng Technol*, vol. 10, no. 12, pp. 943–947, 2022, doi: 10.22214/ijraset.2022.47789.

[5] F. Xoliyarov, S. Gulomov, and S. Bozorov, “The Impact of Artificial Neural Network Architecture on Network Attack Detection,” in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Dec. 2023, pp. 532–539. doi: 10.1145/3644713.3644792.

[6] S. A. Hasanah, A. A. Pravitasari, A. S. Abdullah, and I. N. Yulita, “applied sciences A Deep Learning Review of ResNet Architecture for Lung Disease Identification in CXR Image,” 2023.

[7] S. Sarp *et al.*, “An XAI approach for COVID-19 detection using transfer learning with X-ray images,” *Heliyon*, vol. 9, no. 4, p. e15137, 2023, doi: 10.1016/j.heliyon.2023.e15137.

[8] F. Bodria, F. Giannotti, R. Guidotti, F. Naretto, D. Pedreschi, and S. Rinzivillo, “Benchmarking and survey of explanation methods for black box models,” *Data Min Knowl Discov*, vol. 37, no. 5, pp. 1719–1778, Sep. 2023, doi: 10.1007/s10618-023-00933-9.

[9] S. Sharma, K. Kaushik, R. Sharma, and N. Chaturvedi, “IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES Explainable Artificial Intelligence (XAI),” 2012.

[10] M. Rehman Zafar and N. Khan, "machine learning & knowledge extraction Deterministic Local Interpretable Model-Agnostic Explanations for Stable Explainability," 2021, doi: 10.3390/make.

[11] M. Toğaçar, N. Muzoğlu, B. Ergen, B. S. B. Yarman, and A. M. Halefoğlu, "Detection of COVID-19 findings by the local interpretable model-agnostic explanations method of types-based activations extracted from CNNs," *Biomed Signal Process Control*, vol. 71, no. May 2021, pp. 0–3, 2022, doi: 10.1016/j.bspc.2021.103128.

[12] N. Nigar, M. Umar, M. K. Shahzad, S. Islam, and D. Abalo, "A Deep Learning Approach Based on Explainable Artificial Intelligence for Skin Lesion Classification," *IEEE Access*, vol. 10, no. October, pp. 113715–113725, 2022, doi: 10.1109/ACCESS.2022.3217217.

[13] S. Kumar *et al.*, "LiteCovidNet: A lightweight deep neural network model for detection of COVID -19 using X-ray images," *Int J Imaging Syst Technol*, vol. 32, no. 5, pp. 1464–1480, Sep. 2022, doi: 10.1002/ima.22770.

[14] I. O. Muraina, "IDEAL DATASET SPLITTING RATIOS IN MACHINE LEARNING ALGORITHMS: GENERAL CONCERNS FOR DATA SCIENTISTS AND DATA ANALYSTS." [Online]. Available: https://www.researchgate.net/publication/358284895

[15] P. Mohammadinasab, "Pneumonia Detection Using Deep Convolutional Neural Networks," no. September, 2023, doi: 10.13140/RG.2.2.25567.02720.

[16] I. Salehin and D. K. Kang, "A Review on Dropout Regularization Approaches for Deep Neural Networks within the Scholarly Domain," *Electronics (Switzerland)*, vol. 12, no. 14, 2023, doi: 10.3390/electronics12143106.

[17] M. Reyad, A. M. Sarhan, and M. Arafa, "A modified Adam algorithm for deep neural network optimization," *Neural Comput Appl*, vol. 35, no. 23, pp. 17095–17112, 2023, doi: 10.1007/s00521-023-08568-z.

[18] H. Iiduka, "Appropriate Learning Rates of Adaptive Learning Rate Optimization Algorithms for Training Deep Neural Networks," *IEEE Trans Cybern*, vol. 52, no. 12, pp. 13250–13261, 2022, doi: 10.1109/TCYB.2021.3107415.

[19] N. Das and S. Das, "Epoch and accuracy based empirical study for cardiac MRI segmentation using deep learning technique," *PeerJ*, vol. 11, 2023, doi: 10.7717/peerj.14939.

[20] W. M. Oboya, A. W. Gichuhi, and A. Wanjoya, "A Hybrid DNN-RBFNN Model for Intrusion Detection System," *Journal of Data Analysis and Information Processing*, vol. 11, no. 04, pp. 371–387, 2023, doi: 10.4236/jdaip.2023.114019.

[21] V. R. Mishra, "Image classification of Cow Teat by implementing Convolution Neural Network using PyTorch and Residual Block Image classification of Cow Teat by implementing Convolution Neural Network using PyTorch and Residual Block," no. November, pp. 0–4, 2023.

[22] D. Krstinić, M. Braović, L. Šerić, and D. Božić-Štulić, "Multi-label Classifier Performance Evaluation with Confusion Matrix," Academy and Industry Research Collaboration Center (AIRCC), Jun. 2020, pp. 01–14. doi: 10.5121/csit.2020.100801.

[23] G. Schwalbe and B. Finzel, "A comprehensive taxonomy for explainable artificial intelligence: a systematic survey of surveys on methods and concepts," *Data Min Knowl Discov*, vol. 38, no. 5, pp. 3043–3101, 2024, doi: 10.1007/s10618-022-00867-8.

[24] Y.-S. Lin, W.-C. Lee, and Z. B. Celik, "What Do You See? Evaluation of Explainable Artificial Intelligence (XAI) Interpretability through Neural Backdoors," Sep. 2020, [Online]. Available: http://arxiv.org/abs/2009.10639

[25] A. Chaddad, J. Peng, J. Xu, and A. Bouridane, "Survey of Explainable AI Techniques in Healthcare," Jan. 01, 2023, *MDPI*. doi: 10.3390/s23020634.

[26] E. G. Cervantes and W. Y. Chan, "LIME-Enabled Investigation of Convolutional Neural Network Performances in COVID-19 Chest X-Ray Detection," *Canadian Conference on Electrical and Computer Engineering*, vol. 2021-Septe, pp. 1–6, 2021, doi: 10.1109/CCECE53047.2021.9569029.

[27] P. P. Angelov, E. A. Soares, R. Jiang, N. I. Arnold, and P. M. Atkinson, "Explainable artificial intelligence: an analytical review," *Wiley Interdiscip Rev Data Min Knowl Discov*, vol. 11, no. 5, Sep. 2021, doi: 10.1002/widm.1424.

[28] E. Tjoa and C. Guan, "A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI," *IEEE Trans Neural Netw Learn Syst*, vol. 32, no. 11, pp. 4793–4813, Nov. 2021, doi: 10.1109/TNNLS.2020.3027314.

[29] F. Bodria, F. Giannotti, R. Guidotti, F. Naretto, D. Pedreschi, and S. Rinzivillo, *Benchmarking and survey of explanation methods for black box models*, vol. 37, no. 5. Springer US, 2023. doi: 10.1007/s10618-023-00933-9.

[30] F. M. Talaat and S. A. Gamel, "RL based hyper-parameters optimization algorithm (ROA) for convolutional neural network," *J Ambient Intell Humaniz Comput*, vol. 14, no. 10, pp. 13349–13359, Oct. 2023, doi: 10.1007/s12652-022-03788-y.