Engineering and Technology Journal e-ISSN: 2456-3358

Volume 11 Issue 02 February-2026, Page No.-8935-8942

DOI: 10.47191/etj/v11i02.18, I.F. – 8.482

Explainable Deep Learning for Lung Disease Detection on Chest X-ray Images Using Local Interpretable Model-Agnostic Explanations (LIME)

Muhammad Irsyad[1], Benny Sukma Negara[2]*, Sherly Ananda[3]

[1,2,3]Informatics Engineering Study Program, Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim, Pekanbaru-Riau, Indonesia.

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.

KEYWORDS: Explainable AI, LIME, ResNet18, COVID-19, Pneumonia

I. INTRODUCTION

Artificial intelligence has experienced significant growth across various fields in recent years. Along with this development, numerous methods have been reported that leverage machine learning (ML) and deep learning (DL) models. However, the majority of these models are complex in nature and do not provide clear insights into their decision-making processes, and are therefore often referred to as black-box models. This condition poses particular challenges, especially in critical domains such as healthcare, where the interpretability and transparency of system decisions are essential to ensure trust and safety for end users [1].

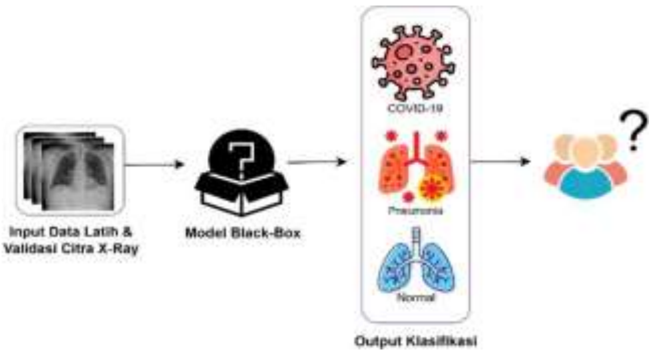


Figure 1. Black-box models are not transparent in their decision-making processes.

In the healthcare domain, lung diseases exhibit various variations with clinical symptoms that are often similar, making it difficult for the general public to accurately identify the type of disease being suffered. One of the most commonly encountered lung diseases is pneumonia, which is an acute infection of the lower respiratory tract caused by inflammation of lung tissue and alveoli (air sacs) [2]. COVID-19, caused by the SARS-CoV-2 virus, is a lung disease that has gained global attention since it was first identified in Wuhan, China, in December 2019 [3].

Artificial Neural Networks (ANNs), as one of the major breakthroughs in machine learning (ML), are models inspired by the functioning of biological nervous systems and have been shown to outperform conventional AI approaches in various machine learning tasks. One of the most prominent ANN architectures is the Convolutional Neural Network (CNN), which is widely used for visual pattern recognition, particularly in image analysis. CNNs are known for their efficient yet effective structure, making them an ideal starting point for neural network modeling [4][5]. One of the most widely used variants of CNN architectures is the Residual Neural Network (ResNet). Although based on the same fundamental principles, CNN architectures such as ResNet vary in terms of the number of layers and parameters used [6]. Research conducted by Salih Sarp et al. [7] demonstrated that the ResNet model is capable of achieving very high classification performance, with an F1-score of 98%. These findings indicate that ResNet is one of the most effective and accurate methods for detecting COVID-19 from chest X-ray images.

As the complexity of modern ML models that operate as black-box systems continues to increase, the need for transparent and interpretable methods has become increasingly urgent. This has driven the growth of the field of Explainable Artificial Intelligence (XAI), which aims to provide better insights into how ML models operate for users and decision-makers [8]. XAI is particularly relevant for models with complex architectures, such as deep neural networks, where transparency in decision-making is a critical aspect [9]. One of the most well-known XAI approaches is Local Interpretable Model-agnostic Explanations (LIME), which is designed to enhance the interpretability of black-box models. LIME works by constructing interpretable local models around individual predictions through the generation of synthetic data with random perturbations to identify the most influential features contributing to the prediction outcomes [10].

CNN-based approaches integrated with LIME have been successfully applied in various studies. Research by Mesut Togacar et al. [11] showed that the combination of CNN and LIME was able to accurately distinguish COVID-19 from pneumonia caused by bacterial and viral infections through medical image analysis. Furthermore, a study by Natasha Nigar et al. [12] developed a ResNet-18–based model trained using transfer learning techniques on the ISIC 2019 dataset. The model was able to accurately classify eight types of skin lesions, achieving accuracy, precision, recall, and F1-score values of 94.47%, 93.57%, 94.01%, and 94.45%, respectively. LIME was employed in the study to provide explanatory visualizations that support rational decision-making, as well as to reveal the model's generalization capability and potential biases toward the image data used.

This research provides several important contributions to the development of interpretable deep learning–based medical image classification systems. The identified contributions of this study are as follows: this research employs the ResNet-18 architecture to classify chest X-ray (CXR) images into three classes, namely normal, COVID-19, and pneumonia. In addition, this study integrates the LIME method as an XAI approach to provide interpretations of the deep learning model's classification results.

II. METHOD

The conduct of a scientific study requires a systematic and well-structured methodological design. Research methodology not only serves as a conceptual framework but also functions as a strategic foundation that guides each stage of the research in a sequential manner. In this study, the implementation was carried out using the Python programming language within the Windows 11 operating system environment.

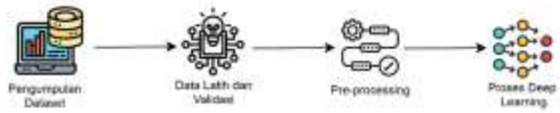


Figure 2. Stages of classification

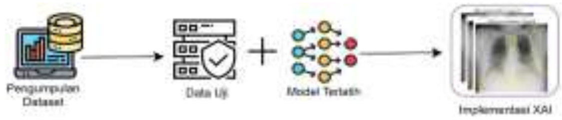


Figure 3. Stages of XAI implementation

Figures 2 and 3 show that this study consists of two main stages. The first stage is the classification process, which begins with the collection of training and validation data, followed by CXR image preprocessing and model training using deep learning techniques. The second stage is the testing phase, in which the test data are processed using the trained model to generate predictions, which are then visually explained using Explainable Artificial Intelligence (XAI) methods.

A. Data Collection

Data collection in this study was carried out by searching for public data obtained from Sachin Kumar et al. [13]. The data collected consisted of 3009 images of normal chest CXR, pneumonia, and COVID-19, of which 3000 were used for training and validation, divided equally for each class. Each class consists of the following: 1,000 normal images, 1,000 pneumonia images, and 1,000 COVID-19 images. Then, 9 images from 3 different classes were used for the test data for XAI implementation. The following are examples of images from the 3 classes of CXR images used:

(a) (b) (c)

Figure 4. CXR images of the lungs: (a) Normal, (b) COVID-19, and (c) Pneumonia

B. Pre-Processing

Resize, at this stage, the input image dimensions are adjusted to 224 × 224 pixels. This adjustment aims to ensure image compatibility with the deep learning architecture used, while facilitating computational efficiency during the training and validation processes. In addition, this size change is designed to adjust the image format to the input standards commonly used in various deep learning models, particularly CNN architectures, which generally accept 224 × 224pixel images as default input

Data splitting, data splitting is a fundamental step in the machine learning model training process. Datasets are usually divided into two main subsets: training data for building models and validation data for evaluating their performance. The splitting ratio can vary, such as 90:10, 80:20, or 70:30, depending on the needs and size of the dataset. The main purpose of this division is to avoid overfitting and to measure the model's ability to generalize to data it has never seen before [14].

C. Deep Learning Models

Model development, the model developed in this study uses one of the CNN architectures, namely ResNet18, which is initialized without pre-trained weights, in an effort to ensure that the learning process starts from a neutral condition without the influence of external parameters. Architectural adaptation is carried out strategically by replacing the fully connected layer at the end of the network with a linear layer that is adjusted to the number of target classes [15]. In addition, a dropout mechanism was added as a form of regularization to minimize the risk of overfitting and improve the model's generalization ability to previously unseen data [16]. The model development process was carried out using the Adam optimization algorithm combined with weight regularization, aiming to maintain stability and accuracy during the parameter update process [17]. As a further strategy to control learning dynamics, learning rate scheduling was applied, where the initial learning rate was set at a high value to accelerate initial convergence, then gradually decreased to refine the model parameters. This approach was designed to produce a medical image classification model that is not only efficient, but also adaptive and reliable in dealing with the complexity of clinical visual data [18].

Model training, DL model training in CXR image classification is carried out through a gradual training cycle that is repeated over a number of epochs. At the beginning of each cycle, the model is activated in training mode to process the entire image batch, where each image is analyzed to produce an initial prediction [19]. Using too many epochs can cause overfitting, while using too few epochs can cause underfitting. Therefore, it is very important to determine the appropriate number of epochs [20]. During the training process, the model parameters are iteratively optimized to minimize the loss function. The loss values for the training and validation data are evaluated at each epoch to monitor the model's performance. The model is automatically saved whenever there is a decrease in the loss value in the validation data, as an indication of improved generalization to unseen data [21].

Model evaluation, the confusion matrix is a classic evaluation tool that has long been used in testing the performance of classification models, both in scientific and technical applications. This matrix is an essential component in various disciplines such as computer vision, natural language processing, and speech recognition. In its most basic form, a confusion matrix is used to describe binary classification results through a two-row, two-column table, which represents the distribution of four possible model outputs:

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

Figure 5. Confusion matrix table

True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). Each of these components provides important information about the model's success and failure rates in distinguishing between correct and incorrect classes [22].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 - score = 2\,\frac{Precision \; x \; Recall}{Precision + Recall}$$

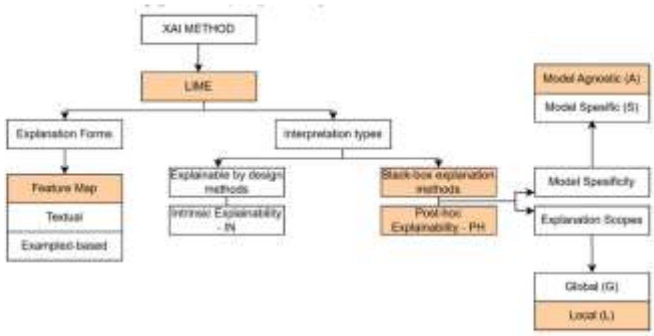
D. XAI Implementation

XAI is a branch of research that focuses on uncovering and explaining the decision-making processes of artificial intelligence systems [23]. As model architectures become increasingly complex, especially deep neural networks, XAI has emerged as a crucial element in the modern machine learning ecosystem, addressing the need for transparency and accountability in the prediction process [9]. Various XAI approaches have been developed to explore the internal mechanisms of models, with the aim of interpreting how neural networks process and respond to input data, thereby providing insights that can be understood by users and decision makers [24].

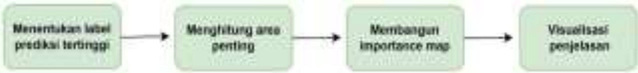
Local Interpretable Model-Agnostic Explanations (LIME) is an approach that explains complex model predictions by building simple models around the analyzed data. This method is model-agnostic, so it can be used on various types of models, including those whose structure is inaccessible or black-box. LIME focuses on explaining a single prediction locally, resulting in easy-to-understand visual interpretations [25]. To generate an explanation for a prediction, LIME generates a number of synthetic data that resemble the original examples and evaluates them using a black-box model. The prediction results from the model are then used as training data for a simple, interpretable linear model. This linear model acts as a local explainer that approximates the behavior of the complex model around that data point. From this model, LIME forms a saliency map that highlights the areas in the image that most influence the prediction [26].

The explanation of various approaches to black box model interpretability can be summarized as follows. Feature maps enable the system to identify and highlight specific areas of the input that most influence the final decision [25]. The visualization is usually displayed in the form of an original image overlaid with a feature map, which represents the intensity of each region's contribution to the prediction result. Meanwhile, the post-hoc approach aims to interpret complex models, such as neural networks, using external techniques to gain interpretive insights into trained models [8][25]. Finally, the model-agnostic approach serves to explain various types of black box models without requiring information about the model's internal structure. This approach also falls under the post-hoc category because understanding is gained through analysis of the output generated from controlled perturbations of the input data [8][25][27].

Next, the local approach is used to explain the reasons behind the model's decision for a particular instance. This approach considers the model as a non-transparent entity and focuses on the local variables that most influence the prediction results [8][25]. The saliency map method aims to interpret the model's decision by assigning a relevance score to each input component, which is visualized in the form of a heatmap or probabilistic map, highlighting the most informative input areas [28]. Finally, the feature importance approach is one of the most common forms of explanation in local interpretation, by assigning importance values to each feature that indicate how much influence that feature has on the prediction being analyzed [29].



F.6. LIME XAI Method



F.7. LIME stages

The LIME procedure identifies labels with the highest probability based on the model's prediction results for the input image. LIME estimates the parts of the image that are most relevant to the model's decision by comparing the influence of image segments on the output. The estimation results are presented in the form of a visual map showing the contribution of each image part to the classification. The importance map is visualized by displaying it as an overlay on the original image, so that the parts that influence the model's decision can be intuitively recognized. The following is the calculation for the LIME operation:

$$argmin!\epsilon G \ \{L\ (f, g, \pi) +\ \Omega\ (g)\}$$

Explanation:

x: An example of the data space we want to explain for the predicted target value. $L\,(f,\ g,\ \pi x)$: A fidelity function that measures how unfaithful g is when approaching f in the locality defined byπx. This is a loss that recognizes locality.

$\Omega(g)$: Measures the complexity of the model from the explainer (g)

f: The black-box model to be explained

g: The explainer

G: Total set of interpretable models

πx: Proximity measure [7]

III.RESULTS AND DISCUSSION

A. Hyperparameter Testing Results

Hyperparameters are important components that determine the architectural configuration and training behavior of convolutional neural networks (CNNs). The main challenge in optimizing them lies in finding the optimal combination of parameters, which not only maximizes model accuracy but also considers training time efficiency [30]. This study applies data splitting to the dataset. Data splitting is applied to the training and validation processes.

The training and validation data will be divided in the deep learning stage with ratios of 90%:10%, 80%:20%, and 70%:30%, which are divided manually.

Table 1. Best data splitting

Dataset	Class	Training Data	Validation Data
Chest *X-ray*	Normal	900	100
	COVID-19	900	100
	Pneumonia	900	100
Total		**2700**	**300**

Based on the parameter testing results in Table 1, the optimal configuration was obtained in a data split scenario of 90% for training and 10% for validation.

Table 2. Hyperparameter testing results

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

Based on the results in Table 2, the optimal configuration was obtained in the scenario with 50 epochs and a learning rate of 0.00001. This combination of parameters produced the highest validation accuracy, which was 98.33%, reflecting the model's excellent predictive performance in recognizing patterns from the input data.

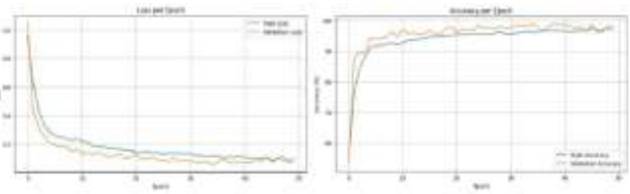


Figure 8. Plot of loss and accuracy of the best hyperparameters

B. Evaluation Metrics

From the *hyperparameter testing* results, it was found that each class has different *precision, recall, and f1-score* values for each dataset.

Table 3. Evaluation Results

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

Based on Table 3, we can see that the COVID-19 class has a higher value than the other classes. This indicates that the model classifies the COVID-19 class well. However, overall, the model has classified well, with the evaluation values for the normal and pneumonia classes being close to those of the COVID-19 class.

C. LIME Visualization

In this study, to see the LIME visualization results, 3 data per class were taken without prior training. The aim was to see the performance of the model that had been built and trained previously. Then, to see whether LIME highlighted the test data images well and correctly. This paper will display LIME visualizations of 3 data per class.

COVID-19, based on the visualization results, it can be seen that LIME highlights the inferior lobe area in both lungs, particularly in the peripheral zone or outer edge of the lungs, as the main contributor in the model's decision to classify this image as COVID-19.

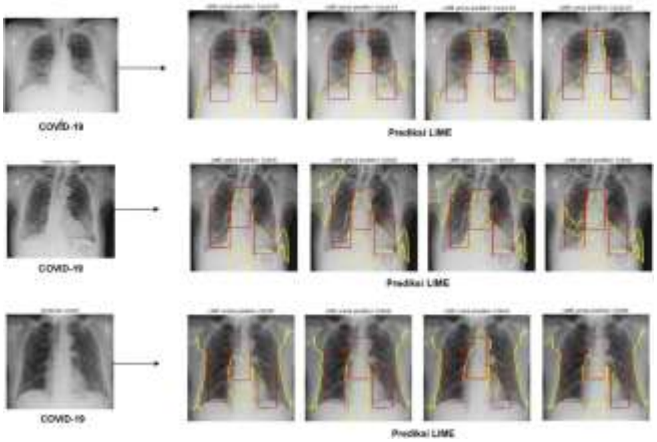


Figure 9. LIME visualization of COVID-19

Normal, based on the LIME interpretation results for normal predictions, it can be seen that the highlighted area is in the middle to lower part of both lungs. This area appears symmetrical and shows no signs of abnormality. LIME shows that the model considers this clean and normal-looking part of the lung to be the main reason for determining that this image belongs to the normal category.

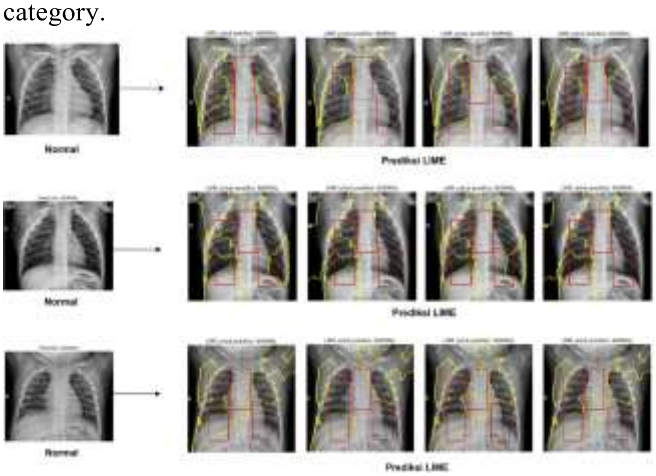


Figure 10. LIME Visualization for Normal

Pneumonia, based on the LIME results for pneumonia prediction, it can be seen that the highlighted areas are located in several areas of the upper to middle lungs. LIME indicates that the model relies on these areas to determine that this image shows pneumonia.

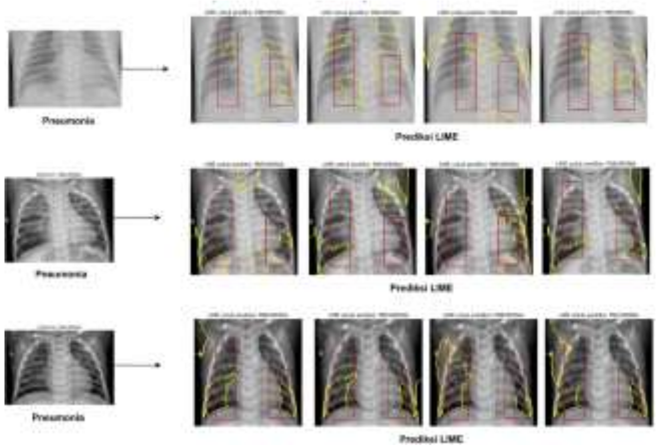


Figure 11. LIME Visualization of Pneumonia

Figures 9, 10, and 11 present the model's prediction results on the test data, accompanied by explanations using the LIME method. Each figure shows yellow superpixel regions that represent significant contributions to the classification decision, as well as red bounding boxes that indicate the ground truth areas of COVID-19, normal, and pneumonia in the lungs. This LIME explanation provides visual insight into the parts of the image that the model considers important in generating predictions. All trained models successfully classified the CXR images in the three figures accurately into classes. This reflects the model's ability to detect the presence of disease and extract relevant features from the lung structure. In particular, the ResNet18 architecture shows consistency between the LIME superpixel regions and the ground truth areas, indicating a valid interpretation. However, there are still explanatory superpixels that appear in areas outside the lungs, which are not directly related to pathology, suggesting that although the model's interpretation is largely correct, there is room for improvement in increasing the model's focus on clinically relevant areas.

CONCLUSIONS

Based on the results of interpretation using LIME, there are differences in the attention patterns of the model in classifying chest CSR images as normal, COVID-19, or pneumonia. In images with a normal prediction, the areas highlighted by LIME are in the middle to lower parts of both lungs and appear symmetrical, indicating that the model relies on the appearance of clean lung structures as the basis for classification. Conversely, in COVID-19 cases, the highlighted areas tend to be in the peripheral and lower parts of the lungs. Meanwhile, in pneumonia predictions, LIME highlights several areas in the upper to middle parts of the lungs.

ACKNOWLEDGMENT
The author would like to express their deepest gratitude to their supervisors and examiners who have provided guidance and support in the research process so that it could be completed successfully.

REFERENCES

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, "Classification of Pneumonia in Lung X-ray Images with Convolutional Neural Network," National Journal of Electrical Engineering and Information Technology, 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, "Detection of Covid-19 Disease in X-ray Images-Ray Images Using a Convolutional Neural Network (CNN) Approach," Jurnal RESTI (Engineering Systems and Information Technology), 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 Mining and Knowledge Discovery, 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. 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.