Detecting Indications of Pneumonia from X-Ray Images Using VGG-19

Jesslyn Hok[1,a)], Messya Carment[1,b)], Michelle Imanuela Winata[1,c)], Muhammad Fikri Hasani[1,d)], Pandu Wicaksono[1,e)]

[1]*School of Computer Science, Bina Nusantara University, Jakarta, Indonesia*

[a)] jesslyn.hok@binus.ac.id
[b)] messya.carment@binus.ac.id
[c)] michelle.winata001@binus.ac.id
[d)] muhammad.fikri003@binus.ac.id
[e)] Corresponding author: pandu.wicaksono005@binus.ac.id

Abstract. Pneumonia is a condition characterized by inflammation of the lungs, resulting in reduced lung function. It is a major cause of mortality globally and imposes a substantial health burden, particularly in Indonesia, where the number of cases remains high. This study proposes a deep learning-based approach using the Visual Geometry Group (VGG-19) to detect Pneumonia from chest X-ray images which highlights its potential for improving diagnostic accuracy. Our model achieved an accuracy of 91.34% in detecting Pneumonia. Even though this study encountered some limitations such as hardware CPU constraints and runtime issues, our findings indicate that VGG-19 can be a reliable method for Pneumonia detection. Future research can build on these results to further enhance the diagnostic performance.

Keywords: X-Ray Images, Image Classification, VGG-19.

INTRODUCTION

Pneumonia is an acute infection caused by microorganisms which include viruses, bacteria, and fungi. Bacteria such as Streptococcus and Mycoplasma are among those that could cause Pneumonia [1]. In addition to that, viruses such as Influenza viruses and Adenoviruses could also contribute to being the cause. One of the ways that these microorganisms could be transmitted is by air pollution, which is common in Indonesia. The infection that these microorganisms create would then cause the inflammation of the lungs, specifically the alveoli, where the alveolar sacs are filled with fluid. This could result in breathing difficulty and extreme chest pain. Pneumonia itself could be detected through x-ray images of the lungs. Pneumonia has been one of the leading causes of death in the world, affecting those living in the West to those living in the East, from third world countries to first, paying no exception, seeing that it affects the youngest to the oldest of individuals within the society. Apart from the transmission of infection across the body, people with Pneumonia can face mortality risks. Indonesia is no exception to that. Pneumonia cases in East Java remained notably elevated until August 2022 [2]. Such high mortality rates of pneumonia deaths among the population have ranked Indonesia one of the highest in the world, outranking other Southeast Asian countries, which causes a substantial health burden in Indonesia [3].

A neural network model could then be created and trained to identify whether a particular image of a lung X-ray shows indications of Pneumonia [4]. CNN's limitations in Pneumonia classification are primarily attributed to its inability to effectively detect edges, a critical step in accurately identifying the presence of Pneumonia. This limitation is substantiated by the findings of Amrulloh, Y.A., et al., who demonstrated that edge detection techniques in Pneumonia classification are inherently flawed, hindering the model's ability to accurately diagnose the disease [5]. Another example, a study by Wang et al. also highlighted the challenges of using CNN in Pneumonia classification, especially in terms of prioritizing hyperparameters to achieve optimal performance. This

suggests that the choice of hyperparameters can be critical in determining the accuracy of the model [6]. In addition to that, CNN models show struggle when being trained on limited datasets, making them very inaccurate. It is also found that CNN models work better with specific CNN architecture and preprocessing techniques. Thus, this study specifically emphasizes the utilization of deep Convolutional Neural Network (CNN) architectures and efficient preprocessing strategies to enhance the accuracy of the model.

This work aims to study the difference between Pneumonia and non-Pneumonia lungs using deep learning. Specifically, we employ the VGG-19 deep CNN architecture to perform image processing and identify the presence of Pneumonia in X-ray images. The dataset, sourced from Kaggle [7], includes two sets of images: normal lung X-rays and Pneumonia-infected lung X-rays. Through Image preprocessing and combining these datasets, we compute the number of white pixels present in the X-ray image. Afterwards, the processed images are then classified to determine whether they indicate signs of Pneumonia.

LITERATURE REVIEW

a. Disease Detection

There are multiple methods in detecting a disease using machine learning algorithms. Usually, disease detection is obtained through a patient's medical background and physical assessments such as X-Ray examinations. By using machine learning, it is often easier to identify malignant cells in a microscopic image rather than visually inspect the images. Early disease detection can significantly improve public health outcomes by identifying and controlling the spread of infectious diseases, thereby aiding government vaccination campaigns, and benefiting the public. A study conducted by Ashan et al., shows that several diseases like heart disease, kidney disease, breast cancer, diabetes, Parkinson's, Alzheimer's, and COVID-19 can be detected by using machine learning [8]. Most of these diseases can be identified using tabular data types, whereas COVID-19 and breast cancer are notable exceptions, as they require image classification for diagnosis in addition to Pneumonia [9].

b. Image Classification using X-Ray Image for Pneumonia

Image classification is a process in machine learning, a part of computer vision, that consists of many kinds of performance and application level, which includes object detection, image feature extraction, video classification, object tracking, super-resolution technology, and many more [10]. There exist many methods that can be used to perform image classification, one of them being CNN (Convolutional Neural Network). CNN is a computer neural network that mimics the anatomical vision system, constructed by researchers through their inspiration of Huber and Weisel's study on a cat's visual cortex.

Besides CNN, there exists SVM (Support Vector Machine) as an alternative method in image classification. SVM is a kernel-based learning algorithm that is one of the most widely used in machine learning. They function as linear binary classifiers, which is a type of classification that decides based on the linear combination of the image's features. SVM works through a learning process, where an iterative process of creating a constructive classifier with an optimal decision boundary is done [8]. A study by Rela et al, has shown an image classification of detecting liver tumor using SVM method with an accuracy of 84.6% [11].

Another popular method that is commonly used in image classification is Random Forest. Random Forest is a classifier and an algorithm that is most used to solve classification and regression problems [8]. Unlike SVM, Random Forest uses ensemble classification, which is a classification method that involves learning a set of classifiers and combining their predictions.

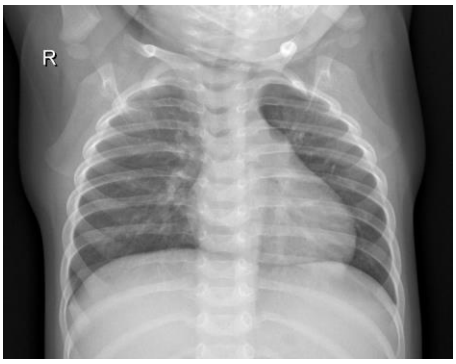

FIGURE 1. X-Ray Image indicating the presence of Pneumonia.

Image classification has a significant application in medical diagnostics, such as detecting and segmenting diseases like kidney disease detection and segmentation, breast cancer, COVID-19, brain tumor, melanoma skin cancer, and Pneumonia [8]. Pneumonia, for instance, is a lung disease which affects the alveolar sacs in either one or both lungs [12]. Studies have shown that X-ray imaging emerges as a pivotal technology due to its rapid acquisition and straightforward accessibility for Pneumonia detection [13]. X-ray images of patients diagnosed with Pneumonia and healthy controls can also be used to train and validate the deep learning model for Pneumonia detection [14].

c. Visual Geometry Group (VGG-19)

Visual Geometry Group network is a deep neural network with multilayered operation. It is based on the CNN model and is applied on the ImageNet dataset. It has 3x3 convolutional layers on top to increase depth and uses max pooling layers to reduce the size of the data. According to research by Mateen, et al., when compared to other methods like AlexNet or SIFT, VGG-19 exhibits superior performance, as evidenced by its higher classification accuracy [15]. Another research by Ikechukwu, et al. demonstrated that VGG-19 achieved an accuracy rate of 97.3% outperforming ResNet-50 and IykeNet [16]. Thus, we chose VGG-19 as a method for our Pneumonia detection owing to its ability to accurately classify medical images which is crucial for precise detection for Pneumonia cases.

METHOD

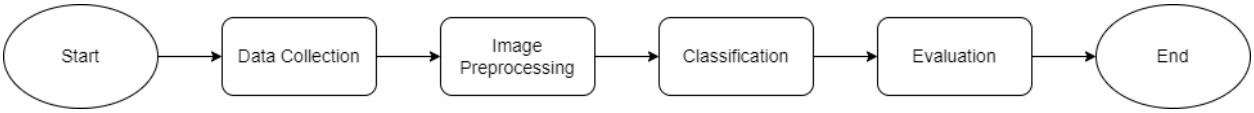

FIGURE 2. Flow diagram of the method.

The method's flow diagram is shown in Figure 2. encompasses sequential stages starting from data collection, proceeding through image preprocessing, classification, and evaluation.

a. Data Collection

This research utilizes publicly available Chest X-Ray images from Kaggle [7]. First, there are two sets of datasets, the first one being the dataset for training, and the second one being the dataset for testing. The dataset consists of two categories of x-ray images, which are x-ray images of normal lungs, and x-ray images of lungs with Pneumonia. Table 1. states in detail the amount of Chest X-Ray images for the training and testing of the model.

TABLE 1. Normal and pneumonia training and testing data

	Normal	**Pneumonia**
Training	1341	3875
Testing	234	390
Total	1575	4265

Figure. 3. (A) and (B) show two examples of Chest X-Ray images that are part of the dataset. (A) belongs to that of a normal lung, and (B) right one belongs to a lung with Pneumonia.

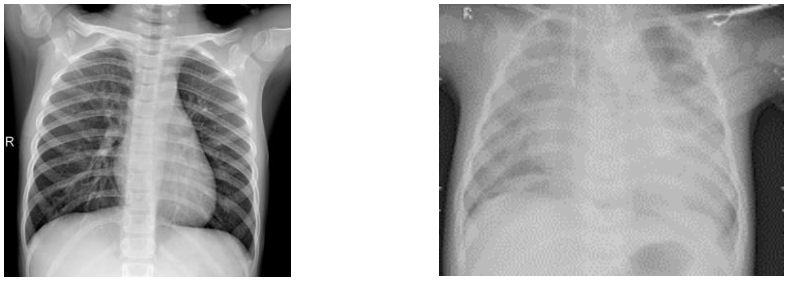

FIGURE 3. (A) CXR example of normal lungs. (B) CXR example of lungs with Pneumonia.

b. Image Preprocessing

Before the model training was executed, we did some data preprocessing. First, to ensure the balance of uniformity for the dataset, whitening transforms [17], rotating, shifting, rescaling, and adjusting the zoom levels of the image datasets were done to reduce correlations as well as create different imaging conditions. This helps to create a wider variety and adds a new perspective for the model to learn from.

c. Classification

In this research, we adapted the pretrained VGG-19 model to suit the input data. The specific architecture employed is described in Table 2. In order to mitigate overfitting, this work implements early stopping with a patience value of 3. Additionally, the model is constructed using Adam's optimizer with an initial learning rate of 0.0001 for a total of 20 epochs.

TABLE 2. VGG-19 Architecture

No	Layer Type	Size	Activation
1	Input	224 x 224 x 3	-
2	Convolution	224 x 224 x 64	ReLU
3	Convolution	224 x 224 x 64	ReLU
4	Max Pooling	112 x 112 x 64	-
5	Convolution	112 x 112 x 128	ReLU
6	Convolution	112 x 112 x 128	ReLU
7	Max Pooling	56 x 56 x 128	-
8	Convolution	56 x 56 x 256	ReLU
9	Convolution	56 x 56 x 256	ReLU
10	Convolution	56 x 56 x 256	ReLU
11	Max Pooling	28 x 28 x 256	-
12	Convolution	28 x 28 x 512	ReLU
13	Convolution	28 x 28 x 512	ReLU
14	Convolution	28 x 28 x 512	ReLU
15	Max Pooling	14 x 14 x 512	-
16	Convolution	14 x 14 x 512	ReLU
17	Convolution	14 x 14 x 512	ReLU
18	Convolution	14 x 14 x 512	ReLU
19	Max Pooling	7 x 7 x 512	-
20	Flatten	25088	-
21	Dense	512	ReLU
22	Dense	1	Sigmoid

d. Evaluation

The results obtained from the proposed model are estimated using various metrics including accuracy, precision, and F1 score. Additionally, this metric also requires brief definitions of four terms, namely “false positive (FP)”, “true positive (TP)”, “false negative (FN)”, and “true negative (TN)”. FP is a term that refers to samples that are part of the negative class but are predicted to belong to the positive class. TP is a term that refers to samples that are classified as positive and are also part of the positive class. FN is a sample that is part of the positive class but is anticipated or prevented from becoming part of the negative class. TN is a sample that is classified as negative and is also included in the negative class. Below are the metrics used for predictions of the proposed model:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (1)$$

$$Recall = TP/(TP + FN) \quad (2)$$

$$Precision = TP/(TP + FP) \quad (3)$$

$$F1 - score = 2/((1/Precision) + (1/Recall) \quad (4)$$

In addition to these ethical considerations, we ensured fairness by eliminating any possible biases that could potentially be present in the trained model. Much attention was paid to any possible source of bias, such as towards demographic imbalances or unequal representation in the dataset. Since the dataset had more data representing CXR images of lungs with pneumonia, we decided to create duplicates of data representing CXR images of normal lungs. This was done to reduce the gap in amount between CXR images of normal lungs and lungs with pneumonia.

All in all, efforts were taken to foster accountability and integrity within every research process. Through these ethical considerations, this research aims to uphold the highest standards of accountability and integrity.

RESULTS AND DISCUSSION

In this section, the results of this research are according to the steps in the method's sections.

a. Evaluation

The dataset used in this experiment was taken from the publicly available Chest X-ray images dataset from Kaggle [7]. This dataset contains a total of 5.840 frontal view Chest X-ray images, all sourced from unique patients. These CXR images consist of those of normal lungs and those of lungs with Pneumonia. There are also three separate files, each containing CXR images, used for training, testing, and evaluating.

b. Image Preprocessing

To prepare the CXR images for training the VGG-19 model, some data preprocessing was done. To enhance the model's ability to generalize, and reduce augmentation bias, data augmentation techniques were done. These techniques included shifting the width, height, rotation range, as well as zooming the CXR images in or out [18].

c. Analysis & Classification

The VGG-19 architecture was employed as the primary deep learning model for detecting indications of pneumonia from CXR images. The model was trained using the preprocessed CXR dataset, while keeping a focus on optimizing the performance metrics, such as accuracy, precision, recall, support, and the F1-score.

During training, the model underwent 6 epochs with an early stop at 3 epochs (to take the best epoch), having a batch size of 32 and learning rate of $1x10^{-4}$. The performance of the VGG-19 models in detecting indications of Pneumonia from CXR images is good, with an overall accuracy of 0.9134 or 91.34% on the test.

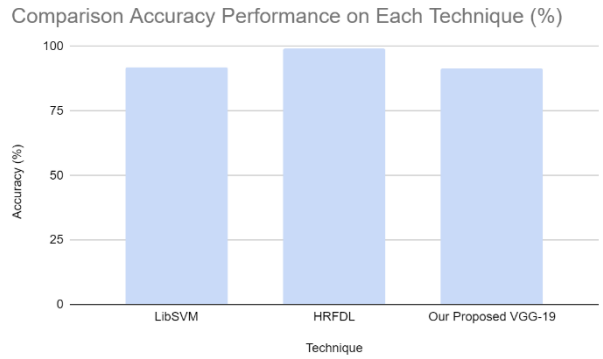

FIGURE 4. Comparison Accuracy Performance on Each Technique (%)

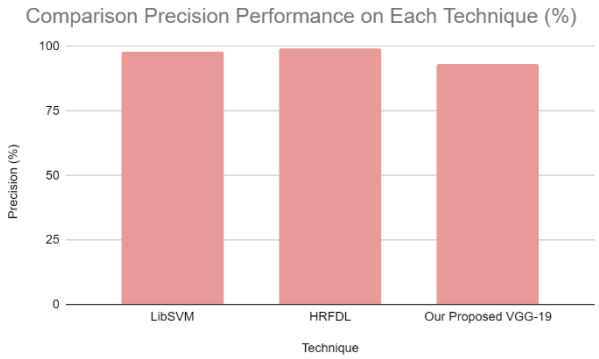

FIGURE 5. Comparison Precision Performance on Each Technique (%)

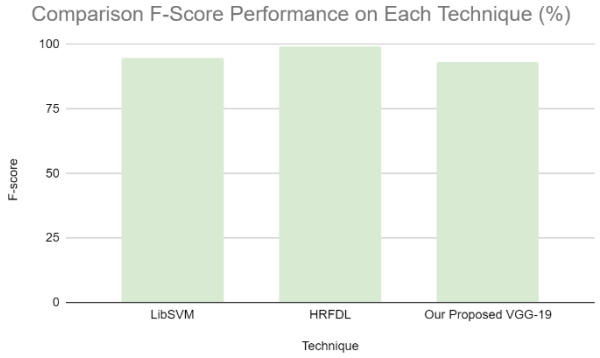

FIGURE 6. Comparison F-Score Performance on Each Technique (%)

Our proposed model is compared with other methods such as SVM and Random Forest Classification. Using LibSVM, the SVM method to detect Pneumonia obtained an accuracy of 91.9% [19], while MOMHTS algorithm optimized HRFDL classifier using Random Forest obtained an accuracy of 99% [20]. LibSVM achieved 97.9% precision, while HRFDL achieved 99% precision. In comparison, our proposed VGG-19 model obtained 93% precision. The comparison of precision for each method is shown in Figure 5. Furthermore, LibSVM achieved an F-

score of 94.6%, HRFDL achieved 99%, and our proposed VGG-19 model achieved 93%. The comparison of F-score for each method is shown in Figure 6.

d. Evaluation Metrics

With an overall accuracy of 91.34%, our model demonstrates a high level of correctness in classifying both Pneumonia and non-Pneumonia from CXR images. Written below are the precision, recall, F1-score, and support score:

TABLE 3. Normal and pneumonia evaluation metrics

	Normal	**Pneumonia**
Training	1341	3875
Testing	234	390
Total	1575	4265

Precision quantifies the proportion of positive predictions that are correct. This means that for our model, 89% of the cases predicted as non-Pneumonia were indeed non-Pneumonia. Recall measures the ability to correctly identify all positive instances out of the actual positives. F1-score is the harmonic mean of precision and recall, providing a balanced measure of the model's overall accuracy.

CONCLUSIONS

In conclusion, this study presents a deep learning-based approach using the VGG-19 to detect Pneumonia from Chest X-Ray images. The proposed model demonstrates a high level of accuracy in classifying both Pneumonia and non-Pneumonia cases from CXR images, with an overall accuracy of 91.34%. The results show that the model is effective in identifying Pneumonia with a precision and recall at 0.93 for positive while 0.89 for negative cases. Although it isn't as high as some other methods compared, this suggests that the VGG-19 model has the potential to improve diagnostic accuracy in Pneumonia detection.

Despite our efforts, we couldn't reach a higher accuracy due to hardware CPU limitations and runtime constraints. Nevertheless, this study highlights that VGG-19 can be a reliable method for detecting diseases like Pneumonia. We hope that future studies can build on this work and achieve even greater results. Future works can build upon our work by addressing the limitations we encountered. Specifically, enhancing the model, improving the computational infrastructure, and optimizing preprocessing techniques that can contribute to even better accuracy in detecting Pneumonia. Additionally, during the preprocessing process, future studies can focus on creating a more balanced dataset, which could improve the model's performance to greater accuracy in Pneumonia detection.

This study emphasizes data preprocessing and augmentation techniques which highlights the importance of these steps in enhancing the model's ability to generalize and reduce biases. Overall, the research contributes to the development of more accurate and accessible diagnostic tools for Pneumonia which could help the public health sectors especially in areas where the disease is prevalent like Indonesia.

REFERENCES

[1] W. S. Lim, "Pneumonia—Overview," in *Encyclopedia of Respiratory Medicine*, Elsevier, 2022, pp. 185–197. doi: 10.1016/B978-0-12-801238-3.11636-8.

[2] F. A. Farisa, S. N. H. Salby, A. A. Rahman, and P. Purhadi, "Modeling the Number of Pneumonia in Toddlers in East Java Province in 2021 with Generalized Poisson Regression," *Inferensi*, vol. 6, no. 2, p. 91, Sep. 2023, doi: 10.12962/j27213862.v6i2.15339.

[3] T. Sudrajat, W. Hakim, T. Rahman, and R. Shah, ""STOP Pneumonia" – A successful advocacy campaign for preventing and protecting Indonesia's children from pneumonia," *Public Health of Indonesia*, vol. 7, no. 1, pp. 1–4, Mar. 2021, doi: 10.36685/phi.v7i1.400.

[4] D. Alapat, M. Meron, and S. Ashok, "A Review on Detection of Pneumonia in Chest X-ray Images Using Neural Networks," *J Biomed Phys Eng*, vol. 12, no. 6, Dec. 2022, doi: 10.31661/jbpe.v0i0.2202-1461.

[5] Y. A. Amrulloh *et al.*, "Computer Aided Classification of X-ray Images from Pediatric Pneumonia Subjects Collected in Developing Countries," *ELKHA*, vol. 15, no. 2, p. 139, Oct. 2023, doi: 10.26418/elkha.v15i2.69981.

[6] T. Gao and G. Wang, "Chest X-ray image analysis and classification for COVID-19 pneumonia detection using Deep CNN," *medRxiv*, 2020.

[7] M. Madhav, "Pneumonia Detection using CNN(92.6% Accuracy)," Aug. 2020. [Online]. Available: https://www.kaggle.com/code/madz2000/pneumonia-detection-using-cnn-92-6-accuracy

[8] M. M. Ahsan, S. A. Luna, and Z. Siddique, "Machine-Learning-Based Disease Diagnosis: A Comprehensive Review," *Healthcare*, vol. 10, no. 3, p. 541, Mar. 2022, doi: 10.3390/healthcare10030541.

[9] L. Cai, J. Gao, and D. Zhao, "A review of the application of deep learning in medical image classification and segmentation," *Ann Transl Med*, vol. 8, no. 11, pp. 713–713, Jun. 2020, doi: 10.21037/atm.2020.02.44.

[10] L. Chen, S. Li, Q. Bai, J. Yang, S. Jiang, and Y. Miao, "Review of Image Classification Algorithms Based on Convolutional Neural Networks," *Remote Sens (Basel)*, vol. 13, no. 22, p. 4712, Nov. 2021, doi: 10.3390/rs13224712.

[11] "Performance analysis of liver tumor classification using machine learning algorithms," *International Journal of Advanced Technology and Engineering Exploration*, vol. 9, no. 86, Jan. 2022, doi: 10.19101/IJATEE.2021.87465.

[12] G. Waterer, "What is pneumonia?," *Breathe*, vol. 17, no. 3, p. 210087, Sep. 2021, doi: 10.1183/20734735.0087-2021.

[13] X. Ou *et al.*, "Recent Development in X-Ray Imaging Technology: Future and Challenges," *Research*, vol. 2021, Jan. 2021, doi: 10.34133/2021/9892152.

[14] M. F. Hashmi, S. Katiyar, A. W. Hashmi, and A. G. Keskar, "Pneumonia detection in chest X-ray images using compound scaled deep learning model," *Automatika*, vol. 62, no. 3–4, pp. 397–406, Oct. 2021, doi: 10.1080/00051144.2021.1973297.

[15] M. Mateen, J. Wen, Nasrullah, S. Song, and Z. Huang, "Fundus Image Classification Using VGG-19 Architecture with PCA and SVD," *Symmetry (Basel)*, vol. 11, no. 1, p. 1, Dec. 2018, doi: 10.3390/sym11010001.

[16] A. Victor Ikechukwu, S. Murali, R. Deepu, and R. C. Shivamurthy, "ResNet-50 vs VGG-19 vs training from scratch: A comparative analysis of the segmentation and classification of Pneumonia from chest X-ray images," *Global Transitions Proceedings*, vol. 2, no. 2, pp. 375–381, Nov. 2021, doi: 10.1016/j.gltp.2021.08.027.

[17] V. Hespanhol and C. Bárbara, "Pneumonia mortality, comorbidities matter?," *Pulmonology*, vol. 26, no. 3, pp. 123–129, May 2020, doi: 10.1016/j.pulmoe.2019.10.003.

[18] L. Chen, S. Li, Q. Bai, J. Yang, S. Jiang, and Y. Miao, "Review of Image Classification Algorithms Based on Convolutional Neural Networks," *Remote Sens (Basel)*, vol. 13, no. 22, p. 4712, Nov. 2021, doi: 10.3390/rs13224712.

[19] S. S. Guddanti, A. Padhye, A. Prabhakar, and S. Tayur, "Pneumonia detection by binary classification: classical, quantum, and hybrid approaches for support vector machine (SVM)," *Front Comput Sci*, vol. 5, Jan. 2024, doi: 10.3389/fcomp.2023.1286657.

[20] M. Hemalatha, "A hybrid random forest deep learning classifier empowered edge cloud architecture for COVID-19 and pneumonia detection," *Expert Syst Appl*, vol. 210, p. 118227, Dec. 2022, doi: 10.1016/j.eswa.2022.118227.