https://doi.org/ 10.33472/AFJBS.6.5.2024. 7995-8005

African Journal of Biological Sciences

A Novel Stacking-Based Hybrid Ensemble Learning Model for Pneumonia Detection

Author 1: Deepak R Jain [1]

Affiliation: School of Computer Engineering and Technology, MIT Academy of Engineering, Alandi
Email id:-djain3745@gmail.com

Author 2: Shitalkumar A Jain [2*]

Affiliation: School of Computer Engineering and Technology, MIT Academy of Engineering, Alandi
Email id:-sajain@comp.maepune.ac.in

Article History
Volume 6, Issue 5, 2024
Received: 22 May 2024
Accepted: 29 May 2024
doi: *10.33472/AFJBS.6.5.2024. 7995-8005*

Abstract: This study introduces a novel stacking-based hybrid deep learning model for the detection of pneumonia, a critical task in the medical field that demands high accuracy and efficiency. The proposed model amalgamates the strengths of multiple deep learning models, including VGG16 and VGG19, to create an optimal prediction model. The stacking technique employed allows for the integration of the predictions of the different models, thereby enhancing the final prediction accuracy. The models were evaluated on a comprehensive dataset of chest X-ray images. The experimental results demonstrate that the proposed model achieves a remarkable accuracy of 96.875% in pneumonia detection. This research contributes significantly to the existing body of knowledge by presenting a unique approach to pneumonia detection and providing insights into the effectiveness of the proposed model. The potential for further improvements and integration into existing healthcare systems is also discussed.

Keywords: Stacking Classifier, Deep Learning, Machine Learning, Ensemble, Hybrid.

Introduction

Pneumonia, a common respiratory infection, is characterized by inflammation in the lungs, primarily caused by bacterial, viral, or fungal infections. It affects millions of individuals across the globe, resulting in high morbidity and mortality rates. Early and accurate detection of pneumonia plays a crucial role in effective treatment and patient outcomes. Diagnosis typically involves a physical exam, medical history review, chest X-ray, and laboratory tests such as blood tests or a sputum culture. These medical images can then be further studied and the exact nature of the disease can be identified.

In recent years, there has been a significant rise in the utilization of Deep Learning (DL) models for disease detection [1–3]. These models have shown immense potential in accurately identifying various medical conditions, including pneumonia. However, there are still challenges in achieving high accuracy due to the complexity and variability of medical images.

Pneumonia, a common respiratory infection, is characterized by inflammation in the lungs, primarily caused by bacterial, viral, or fungal infections. It affects millions of individuals across the globe, resulting in high morbidity and mortality rates. Early and accurate detection of pneumonia plays a crucial role in effective treatment and patient outcomes [4, 5]. Diagnosis typically involves a physical exam, medical history review, chest X-ray, and laboratory tests such as blood tests or a sputum culture [6]. These medical images can then be further studied and the exact nature of the disease can be identified. In this study, Pneumonia detection is the prime focus and high accuracy detection is made possible by the use of deep learning models, learning the chest X-ray images of several patients. Furthermore, machine learning paradigms can also be used to clean the input data before processing in the DL models for analysis [7–9].

Numerous studies have explored the use of Artificial Intelligence (AI) techniques and other deep learning algorithms, for the detection of pneumonia. These techniques have shown promising results in improving the accuracy and efficiency of pneumonia detection. It was reported that deep learning algorithms were capable of accurately detecting pneumonia in chest X-rays with a performance level comparable to that of experienced radiologists [10]. Furthermore, another study found that deep learning algorithms achieved high accuracy in detecting pneumonia in chest X-rays, even in cases with low-quality images [11]. Yoo et al. (2019) [12] examined chest X-ray-based deep learning pneumonia diagnosis. Pneumonia and non-pneumonia chest X-rays were examined using a pre-processing technique to enhance image quality and prepare a dataset to be deep learning compatible, later utilized to train the custom-CNN to spot pneumonia traits. Chen et al. (2021) [13] proposed a deep learning approach for pneumonia identification in radio-graphs. Shin et al. (2016) [14] presented a recurrent neural cascade model for automated image annotation of chest X-rays and focused on developing an automated system capable of reading chest X-rays and providing relevant annotations. [14] demonstrates the potential of recurrent neural networks in automated chest X-ray analysis and contributes to the field of computer-aided diagnosis. More recent works like, Minaee et al. (2020) [15] introduced Deep-COVID, a deep transfer learning approach for predicting COVID-19 from chest X-ray images. These and several other studies demonstrate the potential of AI techniques for improving pneumonia detection and diagnosis. However, further research and clinical trials are needed to validate and refine these approaches for use in clinical settings.

This study aims to address these challenges by proposing a novel stacking-based hybrid deep learning model for pneumonia detection. The model combines the strengths of multiple deep learning models, to obtain an optimal prediction models. The model uses a stacking technique that allows for the integration of the predictions of the different models to improve the final prediction accuracy. The main contributions of this paper are:

1. Proposing a novel stacking-based hybrid deep learning model for pneumonia detection.
2. Evaluating the proposed model on a large dataset of chest X-ray images.
3. Achieving a best accuracy of 96.875% in pneumonia detection.

The rest of the paper is divided into following sections: Section 2. presents the methodology and the architecture of the model. The outcomes and analysis are described in Section 3.Finally, concluding thoughts are presented in Section 4.

1. Literature Survey and Methodology

This section presents the methodology adapted by the study to perform the analysis and derive a hybrid stacking-based deep learning model for pneumonia detection. To derive an optimal stacking-based DL model, the individual DL models need to be considered and analyzed thoroughly. In this particular study, widely used DL models, namely VGG16, VGG19, and ResNet were tested for the purpose of pneumonia detection. Once each of the models has been tested, a series of stacking tests can be conducted to identify the best performing DL model.

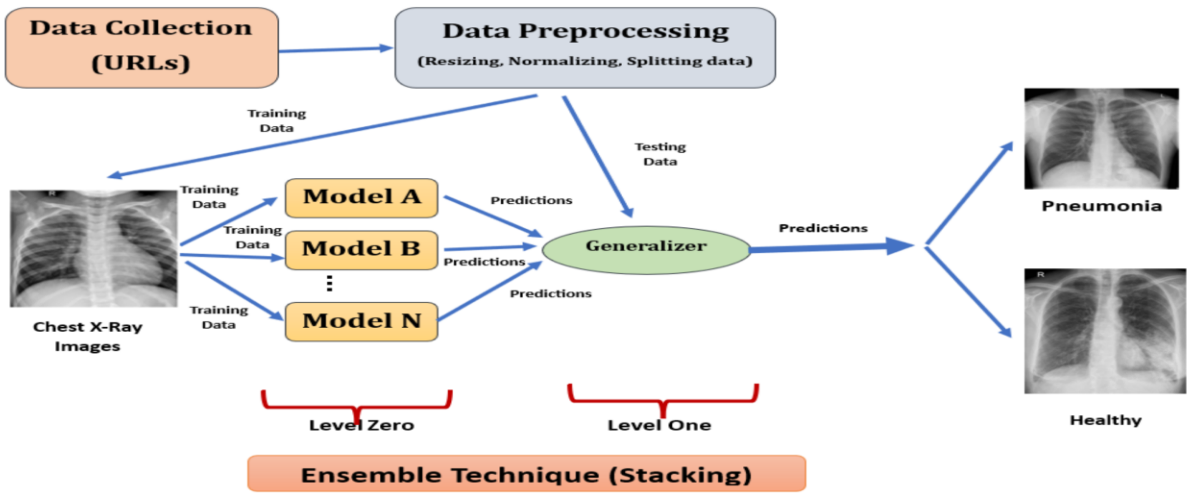

Figure 1: Stacking-based DL Model Architecture

It is evident from Figure 1 that the stacking-based approach can incorporate several DL models before processing the data under the Generalizer to produce the required output. The primary advantage of following this process (of stacking multiple DL models) is to allow the DL model as a whole to inherit the advantages and strong-points of each of the underlying DL model. From various previous works found in the literature, it is evident that each DL model shows different perception to the same input and is sensitive to different aspects of the input data. Thus, utilizing this key-feature of each model and stacking them could possibly help in improving the accuracy of the predictions. However, it should also be noted that with the advantages, the drawbacks of each DL model are also liable to be stacked and amplified in the end result. Thus, it is crucial to study the technique used in stacking model itself.

The input data processing is an equally important step. This study uses the Chest X-Ray14, Chest X-Ray8 datasets, and the COVID-19 Radiography Database, to obtain high-quality input images for the stacking model. The raw data obtained from these data sources still need to be pre-processed before they can be input into any of the models mentioned above. Each of the images was resized, normalized, and converted to grayscale format in due requirements of the DL models. The input images then need to be further classified and reshaped into a Supervised Input dataset to enable the DL models to learn and predict properly. With the input converted to a supervised dataset, it can be split into training, testing, and validation datasets for the model. It should be noted that analysis was performed with Accuracy and Loss for both Training and Testing data sets.

When comparing the performance of the various DL model for a specific task, it's common to benchmark it against a variety of other deep learning architectures and algorithms. The choice of algorithms to compare with, depends on the nature of the task and dataset. Here's a list of some commonly used deep learning architectures and algorithms that can be used for comparison with the VGG19 model:

1. VGG16: As a sibling architecture, comparing VGG19 with VGG16 helps understand the impact of model depth. VGG16 is a convolutional neural network (CNN) architecture with 16 weight layers, including 13 convolutional layers and 3 fully connected layers. VGG19, on the other hand, is a deeper CNN architecture with 19 weight layers, including 16 convolutional layers and 3 fully connected layers

2. ResNet (Residual Networks): ResNet is known for its deep architectures with skip connections. It's effective in addressing the vanishing gradient problem and can be an excellent benchmark for VGG19.

3. InceptionNet (GoogLeNet): InceptionNet uses multiple filters of different sizes within a single layer. It is known for its efficiency and is suitable for tasks requiring high accuracy.

4. DenseNet: It connects each layer to every other layer in a feed-forward fashion. It's known for its compactness and strong performance.

5. MobileNet: It is designed for mobile and embedded vision applications. It's lightweight and computationally efficient, making it a good choice for resource-constrained environments.

6. ResNeXt: It is an extension of ResNet that focuses on cardinality (the number of groups or ”paths” in a network). It's known for its scalability and efficiency.

7. Xception: It is an extension of the Inception architecture and is known for its depth wise separable convolutions. It's efficient and has shown strong performance in image classification tasks.

8. EfficientNet: It is designed using a compound scaling method that balances model depth, width, and resolution. It provides a good trade-off between accuracy and computational resources.

9. Capsule Networks (CapsNets): Capsule Networks are a different approach to deep learning that focuses on modelling hierarchical relationships in data. They can be compared to traditional CNNs like VGG19. When comparing VGG19 with other algorithms, it was essential to use appropriate evaluation metrics, such as accuracy, precision, recall, F1-score, and AUC-ROC, to assess their performance on the specific task you're tackling. Additionally, cross-validation and hyperparameter tuning is also considered to ensure a fair comparison.

2. Results and Discussion

This section presents the results obtained from the models and the system described above. The findings are broken into two parts, the first dealing with the models individually and the second one deals with the stacking model and analyses different combinations possible.

3.1 Individual DL Models

Before considering the collective effort of the models in the stacking approach, the individual performance must be considered. For this study, VGG16, VGG19, and ResNet models were used to study Pneumonia Detection. The accuracy metric for each of the three models is provided in Figures 2, 3 & 4. It can be observed from the graphs that the accuracy of VGG16 is better than the other two models owing to its capability of understanding the depth of the input better. ResNet performs with a saturation in the training scenario but fails to produce similar results for the testing phase. VGG19, on the other hand, gives a steadily rising accuracy with an increase in epochs but largely unstable accuracy for the testing phase.

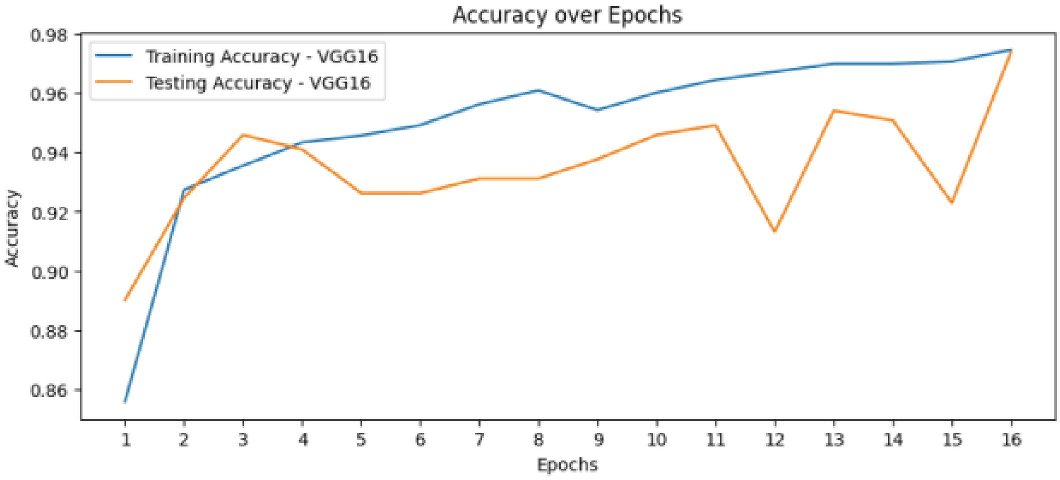

Figure 2: Accuracy Metric for Training and Testing process over multiple epochs – VGG16

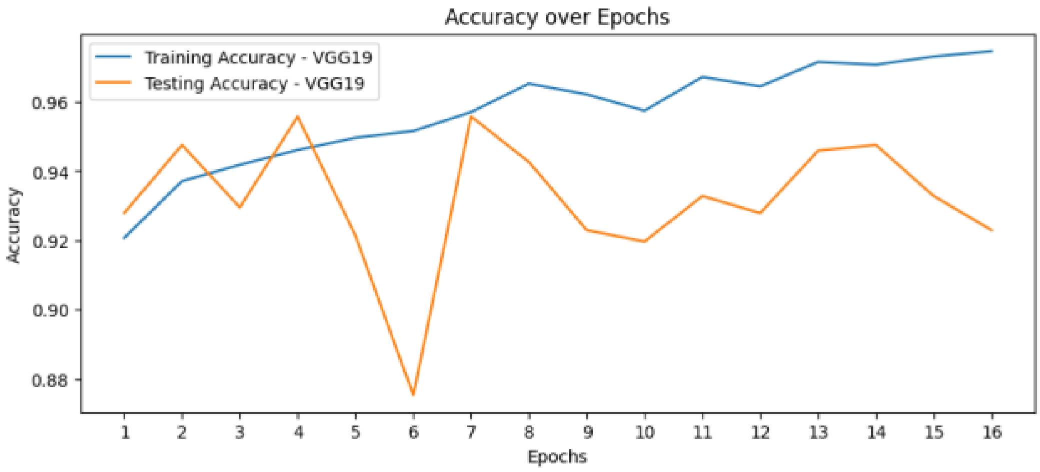

Figure 3: Accuracy Metric for Training and Testing process over multiple epochs - VGG19

Each of the models can be further studies using Figures 5, 6 & 7, representing the loss in both the phases for each model. A similar analysis can be extended from the previous paragraph. ResNet shows a saturated performance in the Training Phase for the Loss metric while a relatively poor performance in the testing phase. VGG19 also shows a steady decline in the loss for the training phase but fails to present similar performance in the testing phase. VGG16 outperforms both the models in this case as well.

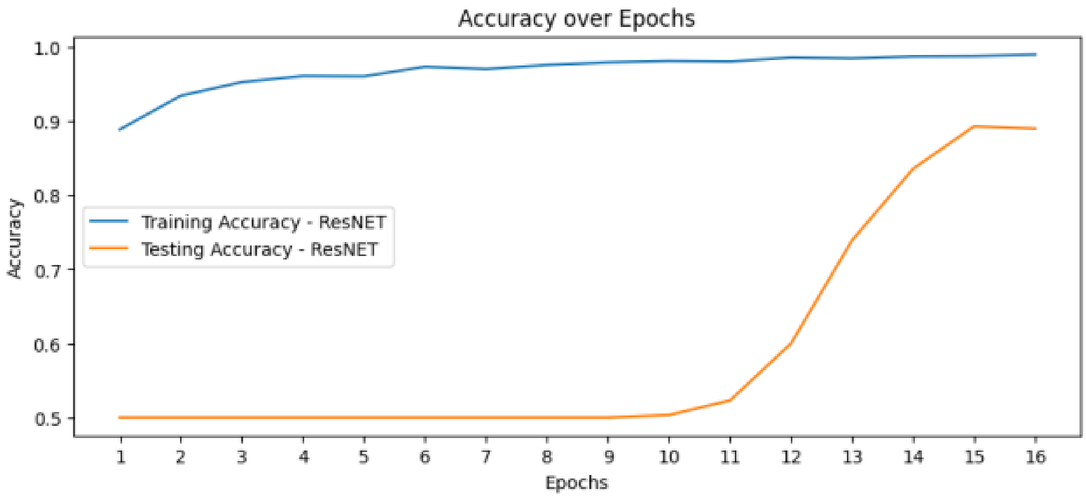

Figure 4: Accuracy Metric for Training and Testing process over multiple epochs – ResNet

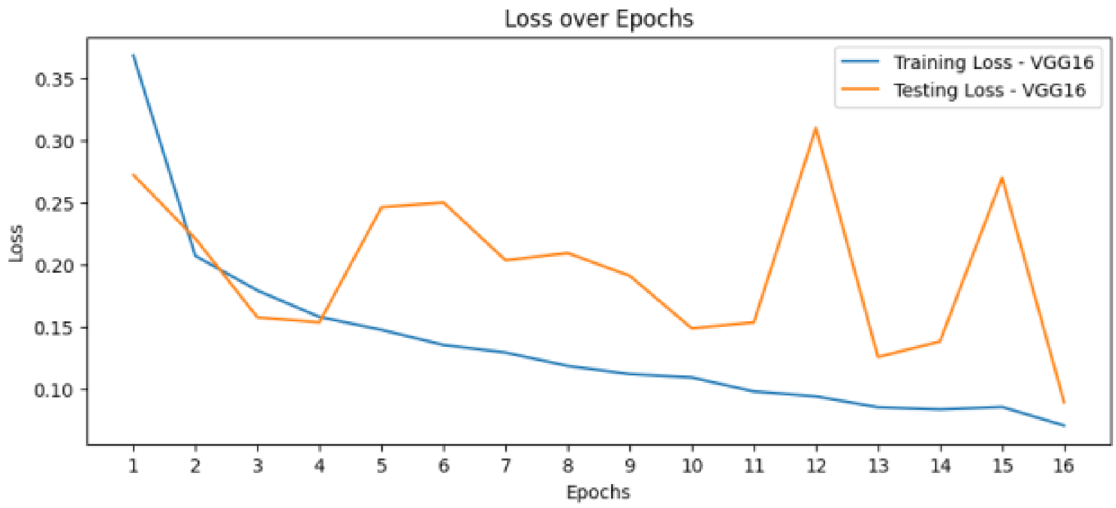

Figure 5: Loss Metric for Training and Testing process over multiple epochs - VGG16

3.2 Stacking-Based DL Model

With the individual characteristics of each model studied in the previous sub-section, this section presents the combined results. Table 1 shows the accuracy metric for various combinations obtained from the stacking-based DL model. From the table it is evident that VGG16-VGG19 combine together very well and output the best accuracy among the other combinations.

One important aspect that must be highlight from the above table is the improvement in performance of ResNet based models. Individually it was found that the ResNet model performance relatively poorly for the pneumonia detection datasets in the previous section. But when combined in the stacking model, both VGG19 and ResNet show significant performance jumps, strengthening the usage of stacking-based DL models.

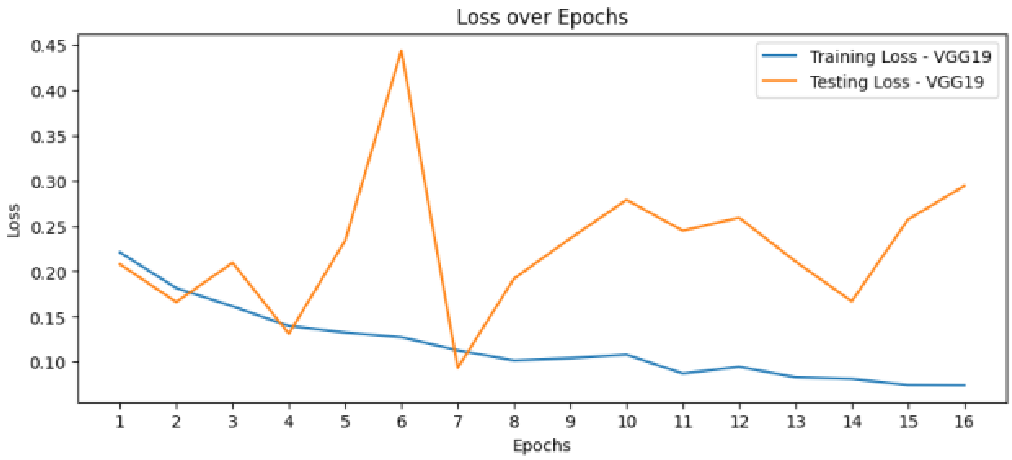

Figure 6: Loss Metric for Training and Testing process over multiple epochs - VGG19

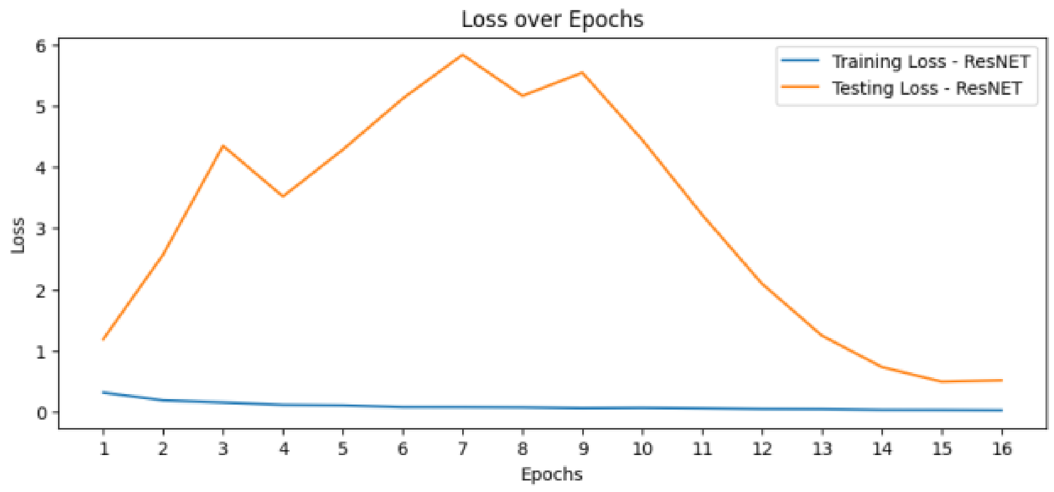

Figure 7: Loss Metric for Training and Testing process over multiple epochs – ResNet

Following table shows the accuracy obtained by stacking the combinations of all the three approaches to find out the model having higher accuracy in order to have less errors in the diagnosis of pneumonia. In further study we can improve the other parameters (TN,FN,TP,FP and F1 score).

S. No.	Stacking-Based Model	Best Accuracy
1	ResNet – VGG16	90.625%
2	ResNet – VGG19	93.750%
3	VGG16 – VGG19	96.875%
4	ResNet – VGG16 – VGG19	78.125%

Table 1 Stacking-based DL Models best accuracy

3. Conclusion

The proposed novel stacking-based hybrid deep learning model for pneumonia detection is a promising approach to ameliorate the accuracy of pneumonia detection based on chest X-ray scans. The model combines the strengths of multiple deep learning algorithms, including VGG16, and VGG19, through a stacking approach to enhance the performance of the model. VGG19 is more complex than VGG16 due to its deeper architecture with additional convolutional layers. The added layers allow VGG19 to capture more intricate features in images. The increased complexity of VGG19 can be beneficial when dealing with complex patterns or fine-grained details in medical images, such as those found in pneumonia cases. The depth of VGG19 gives it an advantage in feature extraction. It can learn hierarchical features at various levels of abstraction, which is useful in medical image analysis. VGG16, while less deep, is still capable of capturing important features, but it may not be as effective in capturing very fine details. Further research can be conducted to evaluate the generalizability of the model to diverse patient populations and to explore the possibility of incorporating other medical imaging modalities for pneumonia detection.

References

References
[1] A. A. Abdullah, A. F. D. Giong, and N. A. H. Zahri, “Cervical cancer detection method using an improved cellular neural network (cnn) algorithm,” Indones. J. Electr. Eng. Comput. Sci, vol. 14, no. 1, pp. 210–218, 2019.

[2] U. Haruna, R. Ali, and M. Man, “A new modification cnn using vgg19 and resnet50v2 for classification of covid-19 from x-ray radiograph images,” Indonesian Journal of Electrical Engineering and Computer Science, vol. 31, no. 1, pp. 369–377, 2023.

[3] M. Mohebbanaaz, Y. Sai, and L. Kumari, “Detection of cardiac arrhythmia using deep cnn and optimized svm,” Indonesian Journal of Electrical Engineering and computer science, vol. 24, no. 2, pp. 217–225,2021.

[4] “American lung association. (n.d.). pneumonia,” https://www.lung.org/lung-health-diseases/lung-disease-lookup/pneumonia, retrieved on January 5, 2024.

[5]“Mayo clinic, pneumonia,” https://www.mayoclinic.org/diseases-conditions/pneumonia/symptoms-causes/syc-20354204, retrieved on 15 October, 2023.

[6] C. Wang, C. Xu, Y. Zhang, and P. Lu, “Diagnosis of chest pneumonia with x-ray images based on graph reasoning,” Diagnostics, vol. 13, no. 12, p. 2125, 2023.

[7] M. Jesmeen, J. Hossen, S. Sayeed, C. Ho, K. Tawsif, A. Rahman, and E. Arif, “A survey on cleaning dirty data using machine learning paradigm for big data analytics,” Indonesian Journal of Electrical Engineering and Computer Science, vol. 10, no. 3, pp. 1234–1243, 2018.

[8] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, “Multi-scale convolutional neural networks for lung nodule classification,” in Information Processing in Medical Imaging: 24th International

Conference, IPMI 2015, Sabhal Mor Ostaig, Isle of Skye, UK, June 28-July 3, 2015, Proceedings 24. Springer, 2015, pp. 588–599.

[9] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, “Deep ehr: a survey of recent advances in deep learning techniques for electronic health record (ehr) analysis,” IEEE journal of biomedical and health informatics, vol. 22, no. 5, pp. 1589–1604, 2017.

[10] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya et al., “Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning,” arXiv preprint arXiv:1711.05225, 2017.

[11] S. Bhardwaj, N. M. Garg, T. Singh, and A. Gupta, “Ai based detection of covid-19 pneumonia in chest x-ray images using resnet50,” in AIP Conference Proceedings, vol. 2916, no. 1. AIP Publishing, 2023.

[12] S. H. Yoo, H. Geng, T. L. Chiu, S. K. Yu, D. C. Cho, J. Heo, M. S. Choi, I. H. Choi, C. Cung Van, N. V. Nhung et al., “Deep learning-based decision-tree classifier for covid-19 diagnosis from chest x-ray imaging,” Frontiers in medicine, vol. 7, p. 427, 2020.

[13] H. J. Chen, L. Mao, Y. Chen, L. Yuan, F. Wang, X. Li, Q. Cai, J. Qiu, and F. Chen, “Machine learningbased ct radiomics model distinguishes covid-19 from non-covid-19 pneumonia,” BMC Infectious Diseases, vol. 21, no. 1, pp. 1–13, 2021.

[14] H.-C. Shin, K. Roberts, L. Lu, D. Demner-Fushman, J. Yao, and R. M. Summers, “Learning to read chest x-rays: Recurrent neural cascade model for automated image annotation,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2497–2506.

[15] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. J. Soufi, “Deep-covid: Predicting covid-19 fromchest x-ray images using deep transfer learning,” Medical image analysis, vol. 65, p. 101794, 2020.