Enhancing Pneumonia Diagnosis: Deep Learning Approaches with Chest X-Ray Images

CHAI WEN ZUN, MOHAMMAD SYAHIR BIN MOHD ASRI, MUHAMMAD ASYRAAF BIN MOKHTAR SANI, MUHAMAD KHAIRUL AKMAL BIN KHAIRUL ANUAR

Faculty of Computer Science and Information Technology,
Universiti Tun Hussein Onn Malaysia,
Parit Raja 86400, Batu Pahat, Johor, Malaysia
CI230043@student.uthm.edu.my, CI230019@student.uthm.edu.my, CI230050@student.uthm.edu.my, AI220382@student.uthm.edu.my

Abstract. This research delves into the exploration of deep learning methodologies to revolutionize pneumonia diagnosis using chest X-ray images. Leveraging sophisticated algorithms such as Convolutional Neural Network (CNN), Visual Geometry Group (VGG), Residual Networks (ResNet), and LeNet-5 (LN-5), our research aims to significantly enhance the precision and efficiency of pneumonia detection through robust classification models.Through meticulous analysis and interpretation of the intricate features within chest X-ray images, this research pioneers innovative models, poised to elevate diagnostic accuracy. The investigation accentuates the transformative capacity of deep learning, illuminating its pivotal role in advancing medical imaging techniques for refined and timely interventions in pneumonia cases.Notably, our analysis highlights LN-5 as the most proficient algorithm, demonstrating an impressive accuracy of 0.8365 on a rigorously tested 70% dataset, accompanied by notably low error rates (0.1635), commendable precision (0.8365), substantial recall (0.94598), an exceptional F1 score (0.96531), and an AUC value of 0.9215. These compelling findings underscore the instrumental role of deep learning in reshaping pneumonia diagnosis, heralding a future of more precise and expedited medical interventions.

1. Introduction

Pneumonia, a prevalent and potentially life-threatening respiratory condition, continues to be a significant public health concern worldwide. Timely and accurate diagnosis is paramount for effective treatment and patient care. In recent years, the integration of deep learning techniques in medical imaging has shown promising results in aiding disease diagnosis, particularly in analyzing chest X-ray images for pneumonia detection.

The Chest X-Ray Images (Pneumonia) dataset comprises chest X-ray images used to discern pneumonia types. A normal X-ray depicts clear lungs devoid of abnormal opacification, contrasting with bacterial pneumonia's focal lobar consolidation, which showcases localized, denser areas due to fluid accumulation. Conversely, viral pneumonia presents a diffuse 'interstitial' pattern, spreading haziness across both lungs instead of localized consolidation. Distinguishing these patterns aids in diagnosing pneumonia's bacterial or viral origins, guiding tailored treatment approaches. deep learning models trained on this dataset aim to recognize these telltale patterns, facilitating automated pneumonia diagnosis or screening.

Our primary objective outlined in the statement is to explore the effectiveness of various advanced deep learning algorithms—such as Convolutional Neural Network (CNN), Visual Geometry Group (VGG), Residual Networks (ResNet), and LeNet-5 (LN-5)—in improving the accuracy and efficiency of diagnosing pneumonia through the analysis of chest X-ray images. The focus is on leveraging these powerful models known for their proficiency in image classification to differentiate between pneumonia-infected lungs and healthy lung conditions. The ultimate goal is to advance diagnostic procedures within clinical settings by enhancing the precision and speed of pneumonia diagnosis using these cutting-edge algorithms.

The utilization of deep learning methodologies for medical image analysis presents a compelling opportunity to augment traditional diagnostic practices, potentially offering more reliable and swift assessments. Through rigorous evaluation and comparative analysis of these four deep learning models, this study endeavors to discern their model performance characteristics, strengths, and limitations in the context of pneumonia detection.

Moreover, this research seeks it outlines the core objectives, intentions and aim of the project. The focus on leveraging deep learning for enhancing accuracy in pneumonia diagnosis using chest X-ray images aligns with the broader goal of advancing medical imaging practices. Additionally, the discussion of ethical considerations, interpretability of results, and the practical implications underlines the comprehensive approach taken to ensure the responsible integration of artificial intelligence in healthcare.

The remaining of this paper is organized as follows. Section 2 reviews all work related to CNN, VGG, ResNet and LN-5. Section 3 presents our chosen methodology for knowledge discovery used to perform the data mining task along with the dataset and the evaluation metrics, showcase our results (Section 4), discuss the implications of our findings (Section 6), and conclude by outlining directions for future research (Section 5).

2. Related Work

Pneumonia is the infection where lung’s air sacs are inflamed by variety of organisms, including bacteria, viruses, and fungi. It may cause cough with phlegm or pus, fever, chills, and difficulty breathing due to the air sacs filled with fluid or pus. Due to the bounce back of cases Corona viruses’ infection, this surge of investigation be absorbed in analytics using deep learning model to diagnose pneumonia by classifying it.

Table 1. Result from related works

Ref.	Model	Accuracy	Precision	Recall	F1-score	AUC
[1]	NN with VGG16	92.1500	0.9428	0.9308	0.9370	0.9880
[7]	CNN	95.6700	0.9550	0.9554	0.9552	0.9700
[4]	ResNet50V2	90.0700	-	-	-	-
[11]	CNN with ResNet	78.7300	-	-	-	-
[12]	DNN	98.9000	0.9770	0.9970	0.9870	-
[13]	Xception	-	0.8400	0.9900	0.9100	0.9600

Table 3 shows the result of related works on lung x-ray using CNN.

All research in table 2.1 uses the similar datasets as used in this report which related to pneumonia (X-ray or CT scan images of patient's lungs). These images are scanned from patients suffered in COVID-19, SARS, Streptococcus, ARDS, Pneumocystis, and others.

For results of references [1], D. Varshni, K. Thakral, L. Agarwal, R. Nijhawan, and A. Mitta used Neural Network (NN) with VGG16 to make analysis on the pneumonia and make classification on it. The NN model with VGG16 used by them obtained accuracy of 92.15%, F1-score of 0.9370, and Area Under Curve (AUC) of 0.9880.

For results in references [7], Patrik Szepesi and László Szilágyi used CNN models with no dropout to classify the chest X-rays image taken from Guangzhou Women and Children's Medical Center were verified by medical experts. The datasets consist of 5868 images, and it return accuracy of 95.67% which tested with CNN model with no dropout. It also returns precision of 0.9550, 0.9554 recall, 0.9552 F1-score, and 0.9700 AUC.

For results in references [4], Rahimzadeh M, Attar A used ResNet50V2 which is a modified version of ResNet50. A modification is made in propagation formation of the connections between blocks to get better results. With covid chest x-ray dataset, taken from GitHub repository, the model gets a result of 90.07% accuracy.

For results in references [11], C.J. Saul, D.Y. Urey, and C.D. Taktakoglu used CNN with lightened image on increased contrast with RN to train the model with dataset released by Radiological Society of North America about pneumonia and gets an accuracy of 78.73%. They also proceed with the same model without interface of RN but results a lower accuracy.

For results in references [12], B. Almaslukh used fine-tuned DNN-based feature classification method to process chest X-ray (CXR) dataset. The dataset was split into 3 and used to proceed with 3 experiments. The accuracy for all 3 experiments is 94.4%, 98.9%, and 96.3%.

For results in references [13], J. E. Luján-García et al used a dataset presented by Kermany at al. in 2018 which contains 5232 chest x-ray images of children from one to five years old. When applying the dataset with Xception model, he gets a validation loss of 0.0453 with F1-score of 0.96, which is the best among VGG16 model, ResNet50V2 model, DenseNet121 model, and Xception model.

3. Methodology

The classification methodology is shown in Fig. 1.

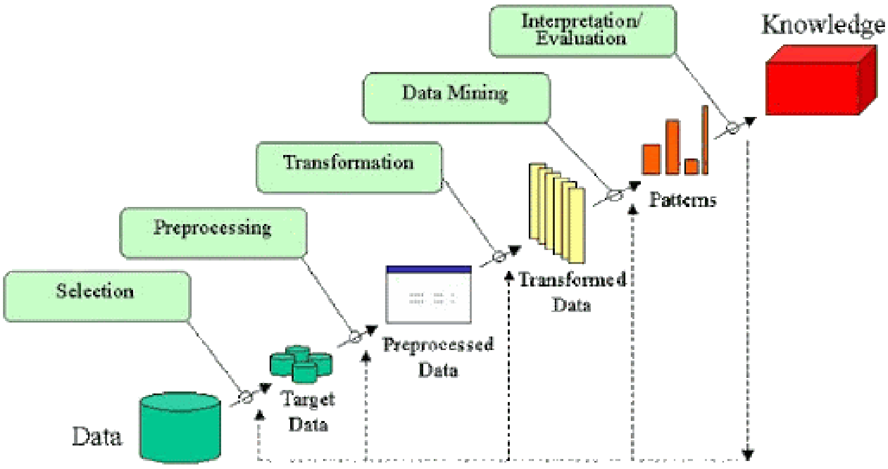

Fig. 1 Knowledge Discovery in Databases (KDD)

In this KDD Methodology for enhancing pneumonia diagnosis with chest X-ray images from Kaggle, a meticulous approach was adopted. The dataset underwent rigorous preprocessing, including augmentation and was organized into batches for efficient processing. Utilizing CNN, ResNet, VGG, and LN--5 architectures, models were fine-tuned and rigorously evaluated on performance metrics like accuracy and recall. Through comprehensive visualization techniques, insights into model behavior were gained, shedding light on their strengths and limitations. This KDD-driven analysis provides a promising direction for leveraging deep learning in improving pneumonia diagnosis classification, emphasizing efficient data handling and model evaluation within the database-driven knowledge discovery process.

3.1 Dataset

The Chest X-ray Pneumonia dataset is a comprehensive collection of medical images focusing on the classification of pneumonia in chest X-ray scans. It incorporates crucial medical and contextual information related to patients, encompassing their health status with respect to pneumonia. This dataset is essential for constructing and training deep learning models geared towards accurate pneumonia diagnosis.
The dataset includes the following critical attributes:

• Image Data: High-resolution chest X-ray images capturing diverse cases of pneumonia and normal conditions.

These attributes collectively serve as a foundation for developing deep learning models capable of distinguishing between normal and pneumonia-affected X-rays. The dataset is a valuable resource for healthcare professionals, aiding in the early identification of pneumonia cases and facilitating prompt medical intervention. Moreover, researchers can leverage this dataset to explore the intricate patterns within chest X-ray images, unraveling the complexities associated with pneumonia development. By analyzing the interplay between various medical and contextual elements, researchers gain insights into the factors influencing pneumonia susceptibility.

In the dataset, the total images of lungs showing pneumonia (4,266 bacterial) and images of normal lungs (1,576). An example of the normal lungs and pneumonia lung images is shown in figure 2.

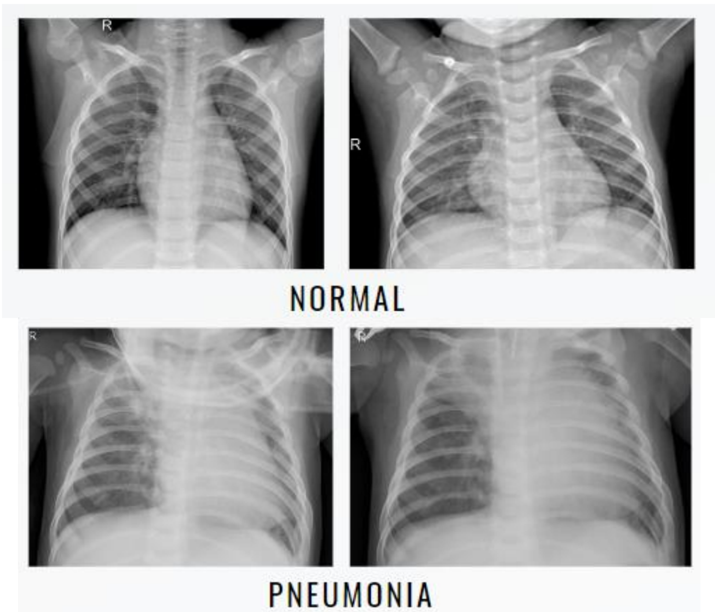

Fig. 2 X-ray images of healthy and pneumonia afflicted lungs from the dataset.

Table 2. Dataset Information

Publishing Year	2018
Owner	PAUL MOONEY

Image Category	Normal | Pneumonia
Number of Images	[5,842] chest X-ray images, including both normal and pneumonia

3.2 Algorithms

The study employs four deep learning algorithms for pneumonia classification:

- Convolutional Neural Network (CNN): The CNN is a network model proposed by Lecun et al.In 1998,.CNN is a feedforward neural network with good performance in natural language processing in image processing and image processing. The prediction of the time series can be effectively applied. Local perception and weight sharing of CNN can greatly reduce the number of networks and thus improve the efficiency of model learning. CNN is mainly composed of two parts: convolutional layer and pooling layer, each with a complex convolution kernel and its calculation formula (Wang & Zhou, 2022) The formula shown in figure 3.

- $f_t=\delta(W_f \cdot [h_{t-1},x_t]+b_f)$

Fig. 3 CNN Formula

- Visual Geometry Group (VGG): VGG stands for Visual Geometry Group; it is a standard deep Convolutional Neural Network (CNN) architecture with multiple layers. The "deep" refers to the number of layers with VGG-16 or VGG-19 consisting of 16 and 19 convolutional layers. The VGG architecture is the basis of ground-breaking object recognition models. Developed as a deep neural network, the VGGNet also surpasses baselines on many tasks and datasets beyond ImageNet. Moreover, it is now still one of the most popular image recognition architectures. The formula shown in figure 4.

$$l_{VGG/i.j} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left(\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y}\right)^2$$

Fig. 4 VGG Formula

- Residual Network (RN) : RN was specifically designed to address the challenge of training very deep neural networks by introducing skip connections or residual connections. Architecture and design principles of the RN model used in our study: The fundamental building block of RN is the residual block. A residual block consists of two or more convolutional layers with a shortcut or skip connection that bypasses one or more of the convolutional layers (D & Bhavani, 2023). The formula shown in figure 5.

$$Y = F(X, \{Wi\}) + X$$

Fig. 5 RN Formula

- • LeNet-5: LeNet-5 model which is the commonly used in the computer vision area regarding its best performances. LeNet5 is a deep learning-based CNN architecture proposed by Yann LeCun, for the purpose of recognizing handwritten digits in images (Amna et al., 2021). The formula shown in figure 6.

$$y_i = \sum_j \left(x_j - w_{ij}\right)^2$$

Fig. 6 LeNet-5 Formula

3.3 Evaluation Metrics

The evaluation metrics used in the experiments are accuracy, precision, recall and F1-score.

- Accuracy: Accuracy measures the overall correctness of predictions, indicating the proportion of correctly predicted instances**.** The formula for calculating accuracy is shown in Eq. 1.

$$Accuracy = \frac{\sum_{i=1}^{N} C_{ii}}{\sum_{i=1}^{N} \sum_{j=1}^{N} C_{ij}} \tag{1}$$

- Precision: Precision evaluates the accuracy of positive predictions, minimizing false positives. The formula for calculating precision is shown in Eq. 2.

$$Precision = \frac{C_{ii}}{\sum_{j=1}^{N} C_{ij}} \tag{2}$$

- Recall: Recall measures the ability to identify all relevant instances, minimizing false negatives. The formula for calculating recall is shown in Eq. 3.

$$Recall = \frac{C_{ii}}{\sum_{j=1}^{N} C_{ji}} \tag{3}$$

- F Score: F Score is the harmonic mean of precision and recall, providing a balanced metric for model performance. The formula for calculating F1 score is shown in Eq. 4.

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4}$$

- AUC: The AUC score refers to the Area Under the Receiver Operating Characteristic Curve. The AUC score quantifies the overall performance of the model by measuring the area under the ROC curve. The formula for calculating F1 score is shown in Eq. 5.

$$AUC = \int_{0}^{1} TPR(FPR)d \tag{5}$$

4. Results

For the CNN model among all the data split, the 60% train and 40% test set produces the highest accuracy, recall, F1-score, and lowest error rate. For our RN model among all the data split, 30% train and 70% test set produces the highest accuracy, F1-score, and lowest error rate. For our VGG model among all the data split, 70% train 30% test set has the lowest error rate, highest accuracy, precision, F1-score. For our LN-5 model among all the data split, 30% train 70% test set has the lowest error rate, highest accuracy, F1-score. The results are shown in Table 3.

Table 3. Experimental results

Data split (%)	Algorithm	Error Rate	Accuracy	Precision	Recall	F1-score	AUC
30-70	CNN	0.2476	0.7524	0.7335	0.9780	0.8383	0.9198
	RN	0.3918	0.6082	0.6261	1.0000	0.7701	0.5855
	VGG	0.3918	0.6082	0.6261	1.0000	0.7701	0.5000
	LN-5	0.1635	0.8365	0.8254	0.9524	0.8844	0.9215
40-60	CNN	0.2642	0.7358	0.7183	0.9915	0.8330	0.8882
	RN	0.5597	0.4403	0.6395	0.2350	0.3438	0.5075

	VGG	0.3977	0.6023	0.6257	1.0000	0.7697	0.4530
	LN-5	0.2614	0.7386	0.7219	0.9872	0.8339	0.8923
50-50	CNN	0.3125	0.6875	0.6895	0.9808	0.8476	0.9537
	RN	0.4063	0.5938	0.6250	1.0000	0.7692	0.6475
	VGG	0.4063	0.5938	0.6250	1.0000	0.7692	0.4385
	LN-5	0.2847	0.7153	0.7132	0.9692	0.8217	0.8715
60-40	CNN	0.2455	0.7545	0.7463	0.9808	0.8476	0.9537
	RN	0.4152	0.5848	0.6265	1.0000	0.7704	0.9062
	VGG	0.4152	0.5848	0.6265	1.0000	0.7704	0.5032
	LN-5	0.2500	0.7500	0.7358	1.0000	0.8478	0.9187
70-30	CNN	0.3438	0.6563	0.6824	0.9915	0.8084	0.8971
	RN	0.4938	0.5063	0.6735	0.2821	0.3976	0.5582
	VGG	0.2188	0.7813	0.7832	0.9573	0.8615	0.9099
	LN-5	0.4375	0.5625	0.6257	1.0000	0.7697	0.5000

Table 4. Selected Data Split

Algorithm	Data Split (%)	Error Rate	Accuracy	Precision	Recall	F1-score	AUC
CNN	60-40	0.2455	0.7545	0.7463	0.9808	0.8476	0.9537
RN	30-70	0.3918	0.6082	0.6261	1.0000	0.7701	0.5855
VGG	70-30	0.2188	0.7813	0.7832	0.9573	0.8615	0.9099
LN-5	30-70	0.1635	0.8365	0.8254	0.9524	0.8844	0.9215

Table 4 shows the data split with highest accuracy and F1-score.

Table 5. Mean of result

Algorithm	Error Rate	Accuracy	Precision	Recall	F1-score	AUC
CNN	0.2827	0.7173	0.7140	0.9845	0.8350	0.9225
RN	0.4533	0.5467	0.6381	0.7034	0.6102	0.6410
VGG	0.3659	0.6341	0.6573	0.9915	0.7882	0.5609
LN-5	0.2794	0.7206	0.7244	0.9818	0.8315	0.8208

Table 5 shows the average value of all data split.

5. Discussion

With the 60-40 data split, the CNN model performs well, as evidenced by its rather low error rate of 0.2455. Recall of 0.8476 and F1-score of 0.9537 show a strong balance between precision and recall, while accuracy of 0.7463 and precision of 0.9808 indicate a high ability to properly classify positive cases. Strong discriminative capability is also indicated by the AUC value of 0.8476. Compared to the CNN model, the RN model, which has a 30-70 data split, has a greater error rate of 0.3918. 0.6261 is a lower accuracy, while 1.0000 is a precision that indicates 100% positive instance classification. The F1-score and recall, however, are significantly lower at 0.5855 and 0.7701, respectively. With respect to the CNN model, the AUC value of 0.5855 indicates a lower degree of discriminative capacity. Compared to CNN and RN, the VGG model shows a reduced error rate of 0.2188 when using a 70-30 data split. The precision of 0.9573 and accuracy of 0.7832 both point to strong overall performance. Furthermore, a balanced performance in terms of detecting favorable cases can be seen by the recall and F1-score of 0.8615 and 0.9099, respectively. The model's ability to discriminate is further supported by the AUC value of 0.9099. Finally, the LN-5 model has a high accuracy of 0.8365 and a low error rate of 0.1635 with a 30-70 data split. While the recall and F1-score of 0.8844 and 0.9215 reflect a balanced performance, the precision of 0.9524 indicates a great ability to correctly detect positive events. The great discriminative capacity of the model is demonstrated by its AUC value of 0.9215. In summary, based on the selected data split, each method displays a different set of advantages and disadvantages. While the RN model exhibits perfect precision but worse recall and F1-score, the CNN and LN-5 models generally perform well across a range of criteria. Additionally, the VGG model performs admirably, particularly when it comes to discriminative power.

6. Conclusions and Future Work

In summary, the analysis compares the performance of four algorithms—CNN, RN, VGG, and LN-5—across different data splits. The CNN model, with a 60-40 split, stands out with low error rates, balanced precision and recall, high accuracy, and strong discriminative capability. In contrast, the RN model, with a 30-70 split, shows higher error rates, perfect precision but lower recall and F1-score. The VGG model, utilizing a 70-30 split, demonstrates reduced error rates, strong overall performance, and discriminative ability. The LN-5 model, with a 30-70 split, exhibits high accuracy, low error rates, and balanced precision and recall, indicating great ability to detect positive events. Each model has its strengths and weaknesses, emphasizing the importance of considering specific classification task objectives when selecting an algorithm and data split. This is particularly relevant in the context of diagnosing pneumonia, where distinguishing between normal and afflicted lungs is crucial.

For pneumonia diagnosis, the CNN model's ability to balance precision and recall becomes crucial in correctly identifying positive cases (pneumonia-afflicted lungs) while minimizing false positives or negatives. The model's discriminative power, reflected in the AUC value, further enhances its capability to distinguish between normal and pneumonia cases. The considerations of accuracy, precision, and recall are especially vital in a medical context, ensuring a reliable and balanced classification of lung conditions. The specific objectives and requirements of pneumonia diagnosis can be well-addressed by the strengths demonstrated by the CNN model in this analysis.

Future research endeavors should focus on refining existing models, exploring novel algorithms, optimizing hyperparameters, incorporating diverse datasets, enhancing interpretability, and collaborating with medical professionals. This holistic approach can contribute to the continued improvement and application of artificial intelligence in pneumonia diagnosis and, more broadly, in medical image classification tasks.

Acknowledgement. This research is supported by Universiti Tun Hussein Onn Malaysia.

References

[1] Shagun Sharma, Kalpna Guleria (2023).: “A Deep Learning based model for the Detection of Pneumonia from Chest X-Ray Images using VGG-16 and Neural Networks”, Elsevier B.V., Chitkara University Institute of Engineering & Technology (2023)

[2] D. Varshni, K. Thakral, L. Agarwal, R. Nijhawan, and A. Mittal, "Pneumonia Detection Using CNN based Feature Extraction," 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 2019, pp. 1-7, doi: 10.1109/ICECCT.2019.8869364.

[3] Dimpy Varshni, Rahul Nijhawan, Kartik Thakral, Ankush Mittal, Lucky Agarwal, “Pneumonia Detection Using CNN based Feature Extraction”, College of Engineering Roorkee, Haridwar, India, 2019.8869364

[4] Rahimzadeh M, Attar A. “A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2”. Inform Med Unlocked. 2020;19:100360. doi: 10.1016/j.imu.2020.100360. Epub 2020 May 26. PMID: 32501424; PMCID: PMC7255267.

[5] Tawsifur Rahman, Muhammad E. H. Chowdhury, Amith Khandakar, Khandaker R. Islam, Khandaker F. Islam, Zaid B. Mahbub, Muhammad A. Kadir, and Saad Kashem, “Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray”, Appl. Sci. 2020, 10, 3233; doi:10.3390/app10093233

[6] S. Xu, H. Wu and R. Bie, "CXNet-m1: Anomaly Detection on Chest X-Rays With Image-Based Deep Learning," in IEEE Access, vol. 7, pp. 4466-4477, 2019, doi: 10.1109/ACCESS.2018.2885997.

[7] Patrik Szepesi, László Szilágyi, “Detection of pneumonia using convolutional neural networks and deep learning, Biocybernetics and Biomedical Engineering”, Volume 42, Issue 3, 2022,Pages 1012-1022, ISSN 0208-5216, https://doi.org/10.1016/j.bbe.2022.08.001.

[8] Kaushik, V. & Nayyar, Anand & Kataria, Gaurav & Jain, Rachna. (2020). Pneumonia Detection Using Convolutional Neural Networks (CNNs). Lecture Notes in Networks and Systems. 471-483. 10.1007/978-981-15-3369-3_36.

[9] Suraj Kumar, Shiva Prakash. Identification and Classification of Pneumonia using CNN Model with Chest X- ray Image, 28 February 2023, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-2602640/v1]

[10] Ahmad T. Al-Taani, Ishraq T. Al-Dagamseh, “AUTOMATIC DETECTION OF PNEUMONIA USING CONCATENATED CONVOLUTIONAL NEURAL NETWORK”, Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 09, No. 02, June 202

[11] C. J. Saul, D. Y. Urey and C. D. Taktakoglu, "Early Diagnosis of Pneumonia with Deep Learning," ArXiv. /abs/1904.00937, DOI: 10.48550/arXiv.1904.00937, 2019.

[12] B. Almaslukh, "A Lightweight Deep Learning-Based Pneumonia Detection Approach for EnergyEfficient Medical Systems," Wireless Communications and Mobile Computing, DOI: 10.1155/2021/5556635, 2021.

[13] J. E. Luján-García et al., "A Transfer Learning Method for Pneumonia Classification and Visualization," Applied Sciences, vol. 10, no. 8, p. 2908, DOI: 10.3390/app10082908, 2020.