Journal of Computers, Mechanical and Management
Journal homepage: www.jcmm.co.in
e-ISSN: 3009-075X
AAN PUBLISHING

Volume 3 Issue 3

Article Number: 240126

Deep Learning-Based Diagnosis of Pneumonia Using Convolutional Neural Networks

Ayesha Karunaratna Mudiyanselage*

University of Oulu, Oulu, North Ostrobothnia, Finland, 8000

Abstract

Pneumonia is a respiratory illness characterized by lung inflammation, often caused by pathogens such as viruses, bacteria, or fungi. Timely detection of pneumonia is crucial for effective treatment. While chest X-rays are commonly used for diagnosis, manual interpretation can be time-consuming, particularly in areas with limited access to trained radiologists. Currently, deep learning models have emerged as an efficient method for pneumonia diagnosis. Numerous researchers are dedicated to enhancing pneumonia diagnostic capabilities through artificial intelligence methods. This study employs a convolutional neural network (CNN) for pneumonia diagnosis. The dataset used in this study consists of chest X-ray images of healthy individuals as well as those affected by bacterial and viral pneumonia. In this study, a CNN model is implemented using an imbalanced chest X-ray dataset with a weighted cross-entropy cost function. The outcome of the developed CNN model shows an accuracy of 75.84%, a precision of 83.16%, a recall of 68.37%, and an F1 score of 68.97% on the test dataset. Further tuning of the model's hyperparameters is necessary to improve performance metrics.

Keywords: Pneumonia Diagnosis;Deep Learning;Chest X-Rays;Convolutional Neural Networks;Medical Imaging

1 Introduction

According to the World Health Organization (WHO), pneumonia is the single largest infectious cause of death in children worldwide. Therefore, an effective and immediate pneumonia detection mechanism is essential. Given that X-ray images serve as the primary diagnostic tool for pneumonia, the development of an automatic detection mechanism using X-ray images is important. Such an advancement would significantly enhance treatment efficacy and ultimately save countless lives that might otherwise be lost due to delays or errors in disease detection. Transmission of pneumonia typically occurs through inhalation of contaminated air, particularly in unhygienic or polluted environments. High-risk groups include young children and older adults above age 60, especially in cases of community-acquired pneumonia. The incidence of pneumonia varies by country, age, and gender, and is higher in male individuals aged above 65 years [1]. With the development of imaging modalities like X-ray and CT, examinations related to diagnosing different respiratory diseases can be performed at a much faster rate. Despite the increased utilization of CT, which has enhanced the early detection of severe clinical manifestations in patients, X-ray remains the primary choice for clinical screening and imaging follow-up due to its simplicity, speed, and cost-effectiveness. Moreover, it is the prevailing examination method for chest diseases currently. The manifestations of pneumonia on chest radiographs typically involve thickening and blurring of early lung markings [2], along with decreased lung transparency.

*Corresponding author: ayesha.karunaratnamudiyanselage@student.oulu.fi
Received: 20 May 2024; **Revised:** 16 June 2024; **Accepted**: 21 July 2024; **Published:** 04 August 2024

DOI: 10.57159/gadl.jcmm.3.3.240126.

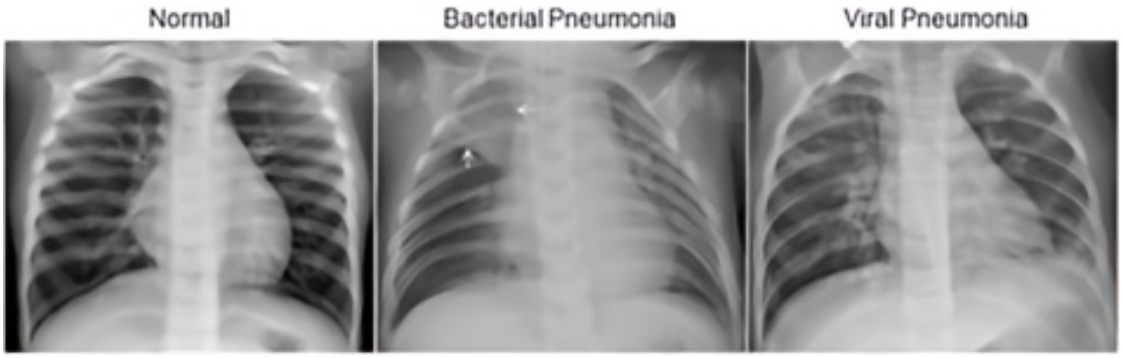

Figure 1: Examples of Chest X-Rays in Patients with Pneumonia extracted from the work of [3], which is the dataset used for this study.

As information and computer technology continue to advance, the use of image digitization and automatic recognition has become widespread across various sectors. This progress facilitates the extraction of information from images, enabling radiologists to make rapid and efficient assessments. The X-ray images from the original study of the dataset used in this paper [3], illustrated in Figure 1, depict three different conditions: a normal chest X-ray (left panel) showing clear lungs without any abnormal opacification, bacterial pneumonia (middle panel) with focal lobar consolidation in the right upper lobe (white arrows), and viral pneumonia (right panel) with a more diffuse interstitial pattern in both lungs. The remainder of this paper is organized as follows: Section 2 provides a review of related works. Section 3 outlines the proposed methodology, including model architecture and details. In Section 4, the experiments are conducted, and the dataset and software are discussed in detail. In Section 5, the results are presented, and further, the limitations are discussed. Finally, the conclusion and future recommendations are presented in Section 6.

2 Related Work

Neural networks have been applied in various domains ranging from transport demand prediction [4] to medical imaging for disease diagnosis. These models, including Convolutional Neural Networks (CNNs), have proven effective in tasks like classification, segmentation, and feature extraction. Throughout history, machine learning has been utilized in various applications within the medical field, particularly in disease diagnosis and prediction tasks. In recent times, a large variety of algorithms suitable for disease diagnosis and prediction have emerged, covering a wide range of disease conditions, including neurological disorders such as Alzheimer's disease, Parkinson's disease, dementia, and epilepsy, as well as cardiovascular diseases, prostate cancer predictions (classifying malignant and benign tumors) [5], and many other common health conditions. The scope of deep learning is broad, and in recent years, there has been a notable transformation in medical image analysis due to the emergence of deep learning-based image analysis methods [6, 7]. Deep convolutional neural networks (CNNs), renowned for their effectiveness in image classification, segmentation, and feature extraction, have played an important role in advancing automated medical imaging systems. Utilizing deep CNNs for pneumonia detection has exhibited promising results. Recent research related to the use of deep learning in pneumonia diagnosis, especially since 2020, has focused mainly on detecting COVID-19-related pneumonia [8]. Many studies, such as [9, 10], have employed transfer learning and deep convolutional neural networks (CNNs) for this purpose. Some image classifications, like [11], were performed using the Support Vector Machine (SVM) algorithm and transfer learning models such as ResNet50, VGG16, and VGG19. Transfer learning with pre-trained models has increased the accuracy of recent X-ray image classification models. The work of [12] provides a comprehensive summary of recent models used in pneumonia detection. Several related studies have been conducted using the chest X-ray dataset employed in this paper. For example, the work of [13] proposes a transfer learning-based convolutional neural network using the same dataset. This proposed model by [13] achieved a test accuracy of 98.43% and an AUC score of 0.9976. The work of [14] utilizes two pre-trained models, EfficientNetB0 and DenseNet121, achieving an accuracy of 95.19%, a precision of 98.38%, a recall of 93.84%, and an F1 score of 96.06% on the test dataset for detecting pneumonia. This model is trained using a similar X-ray dataset as in this study.

3 Proposed Method

The schematic of the Convolutional Neural Network (CNN) model used in this study for the binary classification of X-ray images into Normal and Pneumonia classes is shown in Figure 2. As shown in Figure 2, the model contains three convolutional (Conv) layers used for feature extraction and two fully connected (FC) layers used for classification. In the first Conv layer, there are 64 output nodes, and the activation function used is ReLU. ReLU is applied per pixel and replaces all negative pixel values in the feature map with zero. The second and third Conv layers produce outputs of 128 and 256 nodes, respectively. The activation function used throughout the second and third Conv layers is also ReLU. The mathematical formula for the ReLU activation function is represented in Equation 1:

$$\text{ReLU}(x) = \max(0, x) \tag{1}$$

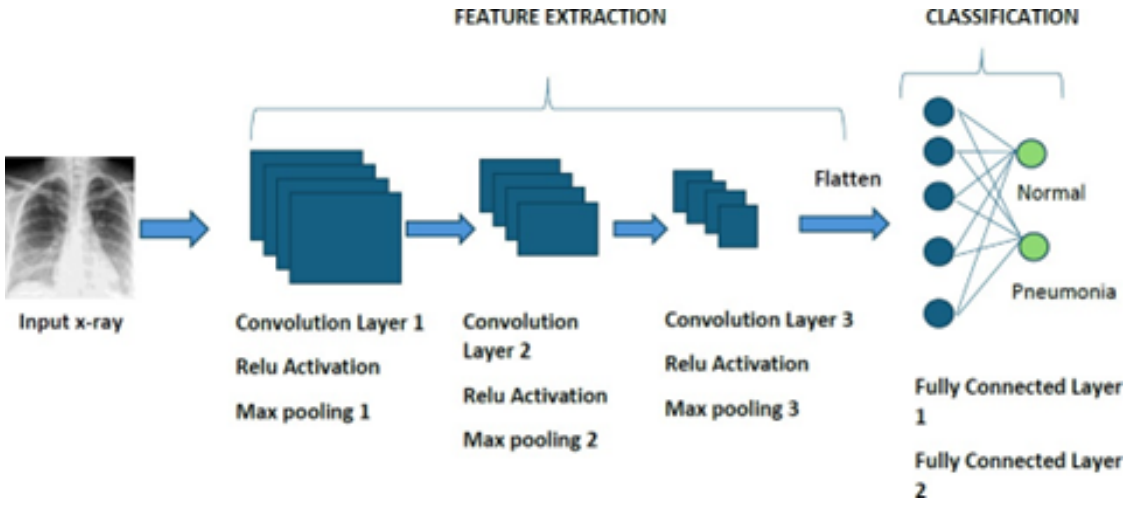

Figure 2: The schematic of the CNN architecture used in this study.

Dropout layers are not used in this model, although they can be included as part of future model development. Adding dropout layers could improve the model by randomly dropping out a percentage of nodes during operation, increasing randomness and potentially reducing overfitting. However, to keep the model computationally simpler, dropout layers were omitted, and the number of convolutional layers was limited to three.

4 Experiments

The experimentation process is carried out using PyTorch's `torch` library, and the adjusted dataset shown in Table 1 is used for training, validation, and testing of the model implemented in this study.

Table 1: The original dataset used for this study.

Dataset	**Number of Images**
Train (Normal class)	1341
Train (Pneumonia class)	3875
Validation (Normal class)	8
Validation (Pneumonia class)	8
Test (Normal class)	234
Test (Pneumonia class)	390
Total	**5863**

Additionally, for performance metric calculations, the `scikit-learn` library is utilized. The chest X-ray images are resized to dimensions of $224 \times 224 \times 3$, normalized to the range $[-1, 1]$, and transformed into tensors. Additional data augmentations such as cropping and rotations are not applied to maintain computational simplicity, given the limitations in GPU availability. The X-ray images are processed in batches. The dataset is visualized after the initial pre-processing, and a sample data visualization of a mini-batch of size 16 is displayed in Figure 3. The CNN architecture described in Figure 2 is utilized for feature extraction and classification. The loss function employed is weighted cross-entropy, and the optimizers used are Stochastic Gradient Descent (SGD) and Adam. The weight tensor of the cross-entropy loss used during the training of the model is calculated as shown in Equation 2:

$$\text{Weight of a class} = \frac{\text{Number of data instances of the class in the training dataset}}{\text{Total number of data instances in the training dataset}} \times \text{Number of classes} \tag{2}$$

Accordingly, weights of 1.92 and 0.67 are utilized for the 'Normal' and 'Pneumonia' classes, respectively, during the training of the model. Due to limited GPU availability, extensive hyperparameter tuning could not be performed during this study. However, some parameter tuning was conducted, and better performance was observed with the hyperparameter combinations shown in Table 3. The best-performing model was obtained with a learning rate of 0.1, a batch size of 32, and a convolutional kernel size of 3x3 using the SGD optimizer.

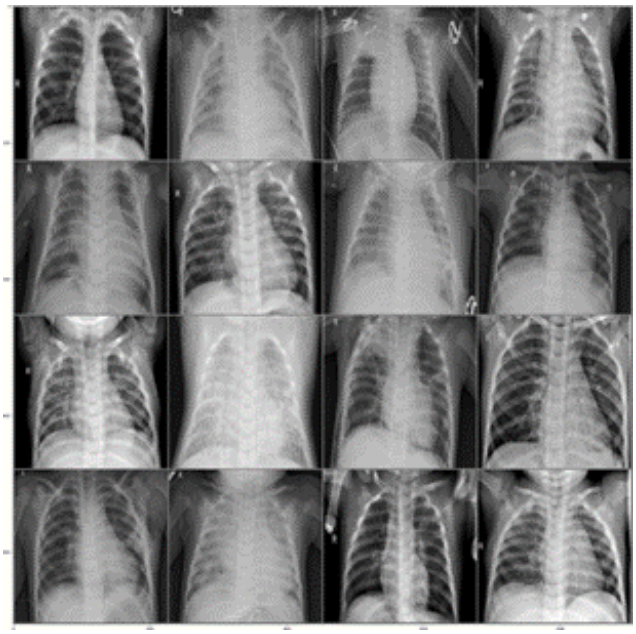

Figure 3: Sample chest X-ray images in a mini-batch of size 16 from the training dataset.

4.1 Dataset

The original dataset was downloaded from Kaggle[1]. The chest X-ray pneumonia dataset contains a total of 5863 X-ray images, arranged in three folders (train, test, validation), each containing subfolders for two categories: 'Pneumonia' and 'Normal.' The X-ray images were selected from retrospective cohorts of pediatric patients aged one to five years from Guangzhou Women and Children's Medical Center, Guangzhou. As per the dataset provider, all chest X-ray imaging was performed as part of patients' routine clinical care. For quality control, all chest radiographs were initially screened to remove low-quality or unreadable scans. The diagnoses for the images were graded by two expert physicians before being cleared for training the AI system, with a third expert reviewing the evaluation set to account for any grading errors. The original dataset is summarized in Table 1, mentioned earlier. The validation folder contained only a few data instances, so the training and validation datasets were rearranged into an 80/20 ratio. The adjusted dataset, used for this study, is presented in Table 2.

Table 2: The adjusted dataset used for this study.

Dataset	**Number of Images**
Train (Normal class)	1080
Train (Pneumonia class)	3104
Validation (Normal class)	271
Validation (Pneumonia class)	780
Test (Normal class)	234
Test (Pneumonia class)	390
Total	**5859**

4.2 Software and Hardware

The Google Colab platform and Jupyter Notebooks serve as the primary tools for editing and executing the Python code related to this study. The GPU provided by Google Colab is utilized for model execution, and as a result, the model's hyperparameters are kept computationally simple to accommodate the available resources. Despite these constraints, deep learning models require significant computational resources, particularly for operations like matrix multiplication, which is a core component of CNNs. As shown in studies on compiler optimization [15], improvements in matrix multiplication efficiency can significantly enhance neural network performance. However, due to the limited GPU availability in this study, the current model implementation remains constrained in terms of computational efficiency and the number of convolutional layers. Additionally, the Miniconda environment installed on a local device is employed for specific editing and visualization tasks, such as data augmentation and result interpretation. Future improvements to the computational setup could further enhance model training, such as integrating matrix multiplication optimizations or using more advanced hardware configurations.

[1] https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia

Table 3: The combinations of hyperparameters that displayed better performance on the testing dataset.

Hyperparameter Combinations	**1**	**2**	**3**
Number of Training Iterations	20	20	20
Learning Rate	0.1	0.01	0.1
Weight Decay of Optimizer	0	0	0.1
Optimizer Type	SGD	SGD	Adam
Kernel Size of Conv Layer	3x3	3x3	5x5

5 Results and Discussion

The best performance is observed on the testing dataset at a learning rate of 0.1. The best-performing model is trained using hyperparameter combination 1, as displayed in Table 3. The relevant code and the saved model can be accessed via footnote[2]

All the results presented in this section are obtained from the best-performing model. Figure 4 shows the variation in training and validation accuracy during the training iterations. After 15 epochs, the model starts to slightly overfit the training data. Hence, the optimal number of training epochs for the model developed in this study lies between 15 and 20 epochs.

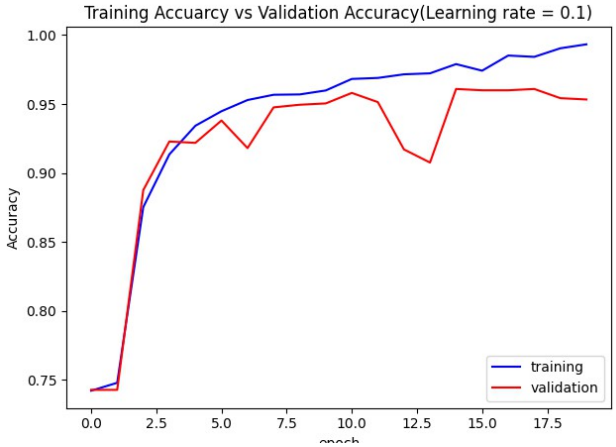

Figure 4: Variation of training and validation accuracy for the model at learning rate = 0.1, number of iterations = 20, weight decay = 0, and optimizer type = SGD.

The performance metrics of the model on the validation and testing datasets are shown in Tables 4 and 5, respectively. The training and validation accuracies of the model are 97.89% and 96.09%, respectively. The testing accuracy is comparatively lower at 75.84%. This indicates that the model has not fully captured the characteristics of the dataset yet.

Table 4: Performance of the model on the validation dataset.

Performance Metric	**Output**
Accuracy	96.09%
Precision	95.34%
Recall	93.34%
F1 Score	94.83%

Table 5: Performance of the model on the testing dataset.

Performance Metric	**Output**
Accuracy	75.84%
Precision	83.16%
Recall	68.37%
F1 Score	68.97%

The confusion matrices for the validation and testing datasets are shown in Figures 5 and 6, respectively. The confusion matrix for the testing dataset shows that the model is biased toward the 'PNEUMONIA' class, which is the majority class in the dataset. While the weighted cross-entropy function has managed to address class imbalance during training, the model

[2]The code and saved model can be requested from the corresponding author.

still exhibits an effect from this imbalance. This suggests that the weight tensor used could be erroneous and may require adjustments in future implementations. Additionally, other methods of addressing class imbalance, such as SMOTE or class balancing, could also be considered in future work.

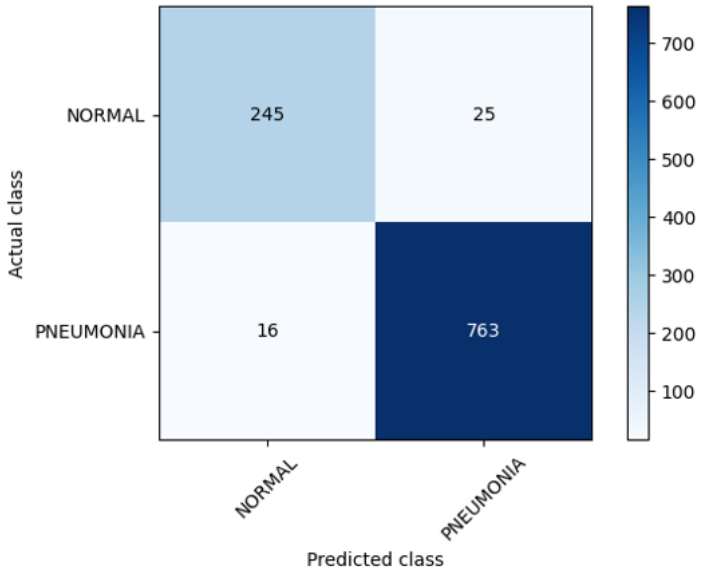

Figure 5: Confusion matrix for the validation dataset.

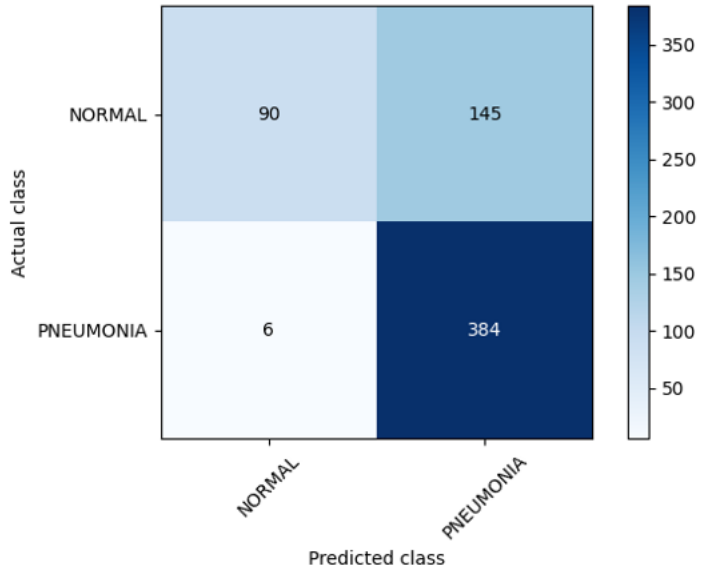

Figure 6: Confusion matrix for the testing dataset.

6 Conclusions and Recommendations

The performance metrics indicate that the CNN model developed in this study demonstrates better classification results on the validation data but fails to exhibit similarly strong performance on the test data. Furthermore, it is evident that the imbalanced nature of the dataset has impacted the final results, despite the use of a weighted cost function during model training. Due to limited GPU availability, the CNN model is constrained by the restricted number of convolutional layers, and no additional data augmentations were applied to the dataset. These factors have also contributed to the relatively low performance of the CNN model on the test dataset. Therefore, the current model is not suitable for clinical applications at this stage of training. Further optimization of hyperparameters is expected to improve performance metrics in future iterations. Additionally, enhancing the model architecture by increasing the number of convolutional layers and incorporating dropout layers could potentially yield better results. Exploring appropriate data augmentations is another possibility for future improvements. A significant enhancement to the model's performance can be achieved through transfer learning, using pre-trained models such as VGG19, ResNet50, or Inceptionv3. By fine-tuning these models for the specific task of pneumonia detection, the network can leverage features learned from vast datasets and adapt them to the current dataset. This approach not only enhances classification accuracy but also reduces the computational burden and training time required. Hence, the use of transfer learning presents a promising avenue for future improvements to the CNN models in this study.

7 Acknowledgment

The author expresses sincere gratitude to the University of Oulu for providing the necessary resources and support for this research. Special thanks are also extended to the reviewers for their insightful comments and suggestions that significantly enhanced the quality of this work.

Declaration of Competing Interests

The author declares no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding Declaration

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author Contribution

Ayesha Karunaratna Mudiyanselage: Conceptualization, Methodology, Data Analysis, Writing - original draft, review, and editing.

References

[1] T. Welte, A. Torres, and D. Nathwani, "Clinical and economic burden of community-acquired pneumonia among adults in europe," *Thorax*, vol. 67, no. 1, pp. 71–79, 2012.

[2] A. Depeursinge, A. S. Chin, A. N. Leung, D. Terrone, M. Bristow, G. Rosen, and D. L. Rubin, "Automated classification of usual interstitial pneumonia using regional volumetric texture analysis in high-resolution computed tomography," *Investigative Radiology*, vol. 50, pp. 261–267, 2015.

[3] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, Feb 2018.

[4] P. Kunekar, K. Jadhav, A. Bhagwat, A. Kirar, A. Singh, S. Devesh, and R. Bhat, "Transport vehicle demand prediction using context-aware neural networks," in *Engineering Proceedings*, vol. 59, p. 232, 2024.

[5] S. Bhattacharjee, H.-G. Park, C.-H. Kim, D. Prakash, N. Madusanka, J.-H. So, N.-H. Cho, and H.-K. Choi, "Quantitative analysis of benign and malignant tumors in histopathology: Predicting prostate cancer grading using svm," *Applied Sciences*, vol. 9, 2019.

[6] J. Wang, H. Zhu, S.-H. Wang, and Y.-D. Zhang, "A review of deep learning on medical image analysis," *Mobile Networks and Applications*, vol. 26, no. 1, pp. 351–380, 2021.

[7] M. Puttagunta and S. Ravi, "Medical image analysis based on deep learning approach," *Multimedia Tools and Applications*, vol. 80, no. 16, pp. 24365–24398, 2021.

[8] "Covid-rdnet: A novel coronavirus pneumonia classification model using the mixed dataset by ct and x-rays images," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 3, pp. 977–994, 2022.

[9] R. B. Nair and C. Kurian, "Pneumonia detection using convolutional neural network," in *Proceedings of National Seminar on Artificial Intelligence & Machine Learning*, p. 15, 2020.

[10] R. Kundu, R. Das, Z. W. Geem, G.-T. Han, and R. Sarkar, "Pneumonia detection in chest x-ray images using an ensemble of deep learning models," *PLoS ONE*, vol. 16, no. 9, p. e0256630, 2021.

[11] G. L. E. Maquen-Niño, J. G. Nuñez-Fernandez, F. Y. Taquila-Calderon, I. Adrianzén-Olano, P. De-La-Cruz-VdV, and G. Carrión-Barco, "Classification model using transfer learning for the detection of pneumonia in chest x-ray images," *International Journal of Online & Biomedical Engineering*, vol. 20, no. 5, 2024.

[12] M. Lakshmi, R. Das, and B. Manohar, "A new covid-19 classification approach based on bayesian optimization svm kernel using chest x-ray datasets," *Evolving Systems*, pp. 1–20, 2024.

[13] M. F. Hashmi, S. Katiyar, A. G. Keskar, N. D. Bokde, and Z. W. Geem, "Efficient pneumonia detection in chest x-ray images using deep transfer learning," *Diagnostics*, vol. 10, no. 6, 2020.

[14] Q. An, W. Chen, and W. Shao, “A deep convolutional neural network for pneumonia detection in x-ray images with attention ensemble,” *Diagnostics*, vol. 14, no. 4, 2024.

[15] R. Kumar, K. C. Negi, N. K. Sharma, and P. Gupta, “Deep learning-driven compiler enhancements for efficient matrix multiplication,” *Journal of Computers, Mechanical and Management*, vol. 3, no. 2, pp. 08–18, 2024.