Pneumonia detection from Chest X-Ray images using Artificial Neural Networks

Adnan Siladdžić
Faculty of Engineering and Natural Sciences
International University of Sarajevo
Sarajevo, Bosnia & Herzegovina
asilajdzic@student.ius.edu.ba

Edin Žiga
Faculty of Engineering and Natural Sciences
International University of Sarajevo
Sarajevo, Bosnia & Herzegovina
eziga@student.ius.edu.ba

Mirza Redzepovic
Faculty of Engineering and Natural Sciences
International University of Sarajevo
Sarajevo, Bosnia & Herzegovina
mredzepovic@student.ius.edu.ba

***Abstract* — Detecting Pneumonia, a potentially fatal respiratory infection, requires early diagnosis through X-ray images. Diagnosing the disease early is crucial, as delayed detection can lead to complications that may not be treated properly. Previously, medical professionals visually inspected images, which resulted in potential human error. However, this study proposes using a 5-layer Artificial Neural Network, consisting of an 2304 Node input layer, three hidden layers with 1024, 512, and 256 nodes respectively, and a single node output layer as the solution. The proposed model uses the TanH activation function on all layers except the output layer, which uses the Sigmoid activation function. In this study, the publicly available Pneumonia Images Dataset from Kaggle was used, which contains over 5800 X-ray images of patients with and without pneumonia. All images were compressed to 48x48, and gray scaled before use. The model was trained on 1840 preprocessed images and continuously tested against 200 randomly sampled images, resulting in a peak classification accuracy of 91% and ROC-AUC score of 0.96. Furthermore, the model showed robustness and capability when facing unseen data, with an average classification accuracy of 83% during continuous testing using random sampling.**

Keywords — Pneumonia, Binary Classification, Chest X-ray images, Artificial Neural Networks, Deep Learning

I. Introduction

Pneumonia is a common respiratory infection that, in high-risk people like children, the elderly, or those who already have underlying health issues, can cause serious sickness and death. Early and accurate diagnosis is key in reducing the risk of fatal complications and in general improving patient outcomes. Chest X-ray imaging is the most used diagnostic tool for pneumonia as it can easily reveal signs of inflammation, consolidation, and fluid accumulation in the lungs. However, interpreting chest X-ray images can be challenging, as the features of pneumonia can be subtle and may overlap with other lung diseases, leading to false positives or negatives [1]. This can result in delayed treatment or inappropriate management, which can have detrimental effects on patient outcomes. Recent advances in artificial intelligence and machine learning algorithms have led to the development of automated systems for pneumonia detection using chest X-ray images. These systems can analyze large volumes of data efficiently and accurately, allowing healthcare practitioners to make informed clinical decisions [2]. Machine learning has the potential to revolutionize pneumonia diagnosis by offering a more objective and reliable method of analysis that can help overcome the limitations of human interpretation. In this study, the existing literature on the topic is reviewed, and a multi-layer Artificial Neural Network model is proposed. By leveraging the power of artificial intelligence, the goal is to improve patient outcomes and reduce the burden of this debilitating disease.

II. Related works

The dataset used within this study is publicly available, hence, there is plentiful literature on the topic of creating models for pneumonia classification using X-ray images. This section aims to overview best of the work done, while also presenting a variety of approaches.

R. Jain, P. Nagrath, G. Kataria, V. S. Kaushik, and D. J. Hemanth conducted research where they utilized Convolutional Neural Networks (CNNs) to detect the presence of pneumonia from X-ray images. In their work, they presented six models, the first two consisting of convolutional layers, while the other four models (VGG16, VGG19, ResNet50, and Inception-v3) were pretrained models. They saw the best results with the first two models, reaching validation accuracies of 85.26% and 92.31% respectively, while the pretrained models performed worse, with VGG16, VGG19, ResNet50, and Inception-v3 achieving 87.28%, 88.46%, 77.56%, and 70.99% [3]. T. Rahman, M.E.H. Chowdhury, A. Khandakar, K.R. Islam, K.F. Islam, Z.B. Mahbub, M.A. Kadir, and S. Kashem performed extensive research in 2020 that used deep convolutional neural networks to automatically classify viral and bacterial pneumonia using X-ray images. Unlike other work, they extended their models beyond binary classification. They used deep Convolutional Neural Networks, namely AlexNet, ResNet18, DenseNet201 and SqueezeNet. With a total of 5247 X-ray images, they achieved classification accuracies of 98%, 95% and 94% for the three models [4]. D. Zhang et al. in 2021 presented a VGG-based model architecture with fewer layers for detecting pneumonia from chest X-ray images. The authors incorporated the Dynamic Histogram Enhancement approach to pre-process the images, addressing the issue of contrast that can impact the quality of diagnosis. Their proposed model had significantly fewer parameters than other models used in literature, such as VGG-16, RES-50, Xception and DenseNet121. The study achieved 90% accuracy for the evaluated models [5]. E. Ayan, B. Karabulut, and H. M. Ünver addressed the issue of pneumonia detection in children through chest X-ray images using deep convolutional neural networks in 2022. Their goal was to develop a computer-aided diagnosis system, where the authors proposed a Convolutional Neural Network approach. They used seven well-known CNN models (VGG-16, VGG-19, ResNet-50, Inception-V3, Xception, MobileNet, and SqueezeNet) that they trained on a chest X-ray image dataset. The three most effective models out of the seven were chosen, with a fourth CNN model created from scratch for sake of performance comparison. The suggested ensemble approach showed remarkable performance on the test data, reaching an AUC of 95.21 and a sensitivity of 97.76. Furthermore, the ensemble method accurately classified different types of pneumonia with over 90% accuracy [6].

As can be seen, the use of Convolutional Neural Networks for image classification is the preferred approach. While more literature does exist for that approach, few papers explore the use of Artificial Neural Networks for the same tasks. Notably, in their paper "Classification of Pneumonia Based on Chest X-Ray Images Using Artificial Neural Networks" Khadivi et al. propose a method for detecting pneumonia in chest X-ray images using artificial neural networks (ANNs). The authors preprocess the images by resizing them and normalizing them, before training an ANN model using a dataset of 586 images. They propose a feedforward neural network with one hidden layer to classify the images as either pneumonia, or normal. With the proposed model, they achieved a classification accuracy of 90.22% which they claim outperforms other models. This work supports the idea that using CNNs is not always necessary for image classification [7].

This study aims to fill the existing gap when it comes to using simple Artificial Neural Networks to classify chest X-ray images, as the research that does exist is mostly based on Convolutional Neural Networks and high-resolution imaging.

III. Methodology

The following section aims to describe the methodology used within this study and is divided into the following subsections: data collection and preprocessing, model architecture, training, and evaluation.

A. *Data Collection and Preprocessing*

The Pneumonia images dataset is a publicly available dataset consisting of approximately 5,861 chest x-ray images, categorized into three sets: training, testing, and validation, with each set containing 5,216, 629 and 16 images respectively. The images vary in resolution, color modes, size, and image format. All preprocessing was done using built-in libraries in Python, including Pillow for image processing and NumPy for numerical operations.

The first step in preprocessing was to convert all images to grayscale and to choose an adequate resolution. Several resolutions ranging from 16x16 to 256x256 were considered, and a resolution of 48x48 was found to provide the best performance for the time it took to train the model. Beyond this resolution, the returns were diminishing (in terms of accuracy and loss), and model training time increased significantly. Lastly, the image data was normalized for a range from 0 to 1.

B. *Model Architecture*

Artificial Neural Networks are a type of machine learning model created based on the concept of neurons in the human brain. They consist of interconnected neurons that process and transmit information through a network. Much like the human brain, Artificial Neural Networks possess the ability to learn and adapt to the problem provided, making them ideal for non-linear problems [8]. While they can be used in a variety of applications, one of the most popular uses is for image classification [9]. By using Artificial Neural Networks, the aim is to develop a model that can successfully analyze and classify a chest X-ray image as either pneumonia-positive or pneumonia-negative.

For this study a 5-layer Artificial Neural Network is proposed, with the implementation being done from scratch, hence every part of the network was manually coded without the use of libraries. This approach was taken as it provided a good run time for the performance provided.

The architecture of the model consisted of an input layer, three hidden layers, and an output layer with a single node. The number of neurons and activation functions used for each layer are shown in the table below.

Layer	Number of neurons	Activation Function
Input layer	2304	TanH
Hidden layer 1	1024	TanH
Hidden Layer 2	512	TanH
Hidden Layer 3	256	TanH
Output Layer	1	Sigmoid

Table 1 Proposed Model Architecture

To optimize the model, multiple layer designs and activation functions were explored, but this configuration proved to provide the best performance for this specific use case. Using the ReLU activation function for all hidden layers, or in combination with TanH didn't impact accuracy as much as it introduced high run-to-run variation in results, hence the approach was discarded. Previous research has also shown that a multi-layer approach with lower image resolution performs similarly to a two-layer approach with high resolution [3]-[7]. The multilayer approach was preferred as using high image resolutions would linearly scale with the number of input neurons, hence significantly increasing runtime.

Lastly, the Sigmoid activation function was applied on the single output node. Binary Cross-Entropy loss function with accuracy measurements was used to optimize the model parameters.

C. *Hyperparameters*

The following hyperparameters were used to train the model:

- Learning rate: 0.001
- Decay rate: 0.0001
- L2 Weight and Bias regularizers: 0.00005
- Batch size: 46
- Number of epochs: 40
- Optimizer: Adam

The learning rate controls the updates that happen during training, with higher learning rates causing faster adaptation to the training data provided, but also the risk of overfitting [10]. The chosen 0.001 rate was deemed satisfactory, with a decay of 0.0001. Moreover, a L2 weight and bias regularization of 0.00005 to further combat the issue of overfitting.

The batch size hyperparameter determines how many examples are used in each training iteration. Like with the learning rate hyperparameter, a higher batch size may lead to overfitting, while a smaller batch size may lead to the opposite. Increasing the batch size beyond 46 leads to extreme decreases in accuracy, while anything below resulted in instabilities within the model.

Furthermore, the number of epochs determines how many times the dataset is processed during training [11]. Following the trend of the other hyperparameters, too few may lead to underfitting, while too many to overfitting.

The chosen optimizer for this study is the Adam optimizer. The Stochastic gradient descent, Adagrad and Nesterov Accelerated Gradient optimizers were implemented during testing, but the Adam optimizer provided the most consistent results run-to-run.

D. Training

The 5-layer model was completely created from scratch, containing an input layer, three hidden layers, and an output layer. The number of input nodes (2304) directly correlated with the number of pixels in every image, therefore, every input was a normalized pixel value which would get appropriately processed.

Given that the dataset contains over 5,800 images, a representative sample of roughly 1840 training images and 200 testing images was used. The training data was fed in batches of 46 for 40 epochs, where random sampling without replacement was used for the images within a single batch. Hence, a single batch could not contain two of the same images, but multiple batches could have overlapping images. Considering the size of the dataset, the impact of this overlapping was assumed to be negligible.

Furthermore, accuracy and loss gotten from the Binary Cross Entropy function were used as an evaluation metric, as well as consistency between epochs and training sessions. The learning rate, decay rate, the L2 Weight and Bias regularizers were all tuned in increments to create a balance. Furthermore, early stopping was used if the model would reach a plateau in accuracy or loss to prevent overfitting. The tuned model after training provided the following performance:

Metric	Value
Training Accuracy	89%
Training Loss	0.22

Table 2 Training Metrics

Overall, this training procedure allowed the model to adapt to the needs of the dataset while preventing overfitting and long runtimes.

E. Evaluation

To evaluate the performance of the model, the following performance metrics were used: accuracy, precision, recall, F1-score, and loss. While these measurements are good metrics for evaluation, the overall goal was to create a model that would be accurate and reliable. Furthermore, metrics such as a ROC curve were considered.

To ensure the robustness of the model, multiple testing runs were conducted to firstly check for improvements, and then confirm that instability was not introduced. Since the dataset was pre-split into training, validation, and testing, the moving average with a window of 10 was observed and evaluated for each performance metric to confirm consistency over roughly 1000 runs. The moving average was chosen to present the general performance of the model without allowing one bad run to skew the data.

One of the limitations of the model is that it was only trained and evaluated by one dataset. While the model does show consistent performance with random sampling without replacement, indicating that it responds well to new data, it cannot be fully confirmed without being tested on a completely unseen dataset.

IV. Results and Discussion

The following section outlines the performance achieved by the proposed model on the test dataset of 200 48x48 grayscale images, after being trained on approximately 1840 images. The model performed exceptionally well, reaching a peak accuracy of 91%, and a low overall false negative rate. Furthermore, these results outperform or match the available literature for approaches using Artificial Neural Networks [7]. Below can be found a table showing the peak performance metrics of the proposed model, how those metrics were calculated as well as the supplementary confusion matrix.

Metric	Value
Accuracy	91%
Precision	86%
Recall	95%
F1 Score	91%
ROC-AUC	0.96
Loss	0.31

Table 3 Peak Performance Metrics

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

Equation 1 Accuracy

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

Equation 2 Recall

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

Equation 3 Precision

$$F1\ Score = \frac{2 \times (Precision \ \times Recall)}{(Precision + Recall)}$$

Equation 4 F1 Score

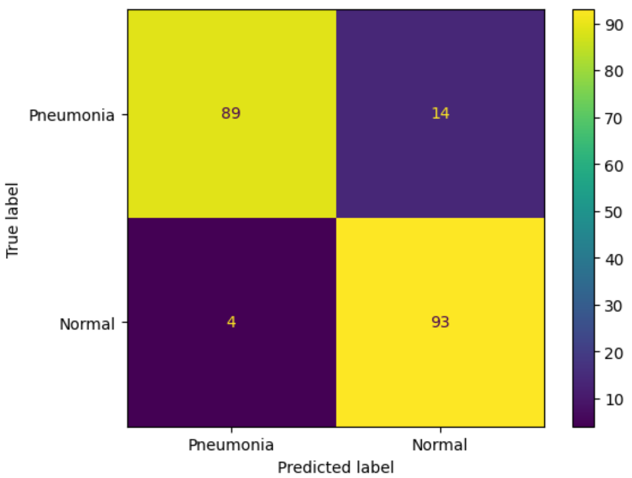

Figure 1 Confusion Matrix

As can be seen from data provided, the model achieved a strong recall of 95%. Only 18 cases out of 200 were misclassified, indicating that the model is extremely capable of handling the task provided.

Furthermore, a Receiver Operating Characteristic was used to estimate the performance of the model. The ROC curve is a graphical representation of the trade-off between true positive rate (pneumonia-positive classified as positive) and false positive (pneumonia-negative classified as positive) rate for binary classification [11]. A value of 0.96 shows that the model has very strong discriminatory power in distinguishing between positive and negative cases. The subsequent ROC curve can be found below.

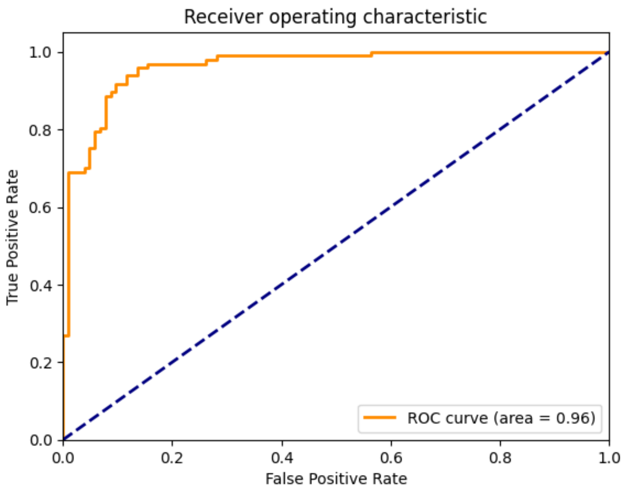

Figure 2 ROC Curve

Moreover, the model's consistency between runs was evaluated. By running the model for 1000 testing passes using random data samples, it was possible to check its performance when constantly faced with different data. The model's accuracy was consistently above 83%, with only a few dips below this threshold. This trend was also seen in all other performance metrics. The detailed table may be found below:

Metric	Mean	St. Deviation
Accuracy	0.83	0.02
Precision	0.80	0.03
Recall	0.87	0.02
F1 Score	0.83	0.02
Loss	0.37	0.02

Table 4 Average performance random sampling (n=1000)

Lastly, the moving average with a window of 10 for each performance indicator was plotted and used in order check the consistency of the model. The moving average was chosen to balance out any major outliers that might occur during the testing runs, and in general present that the model stays consistent when fed new data.

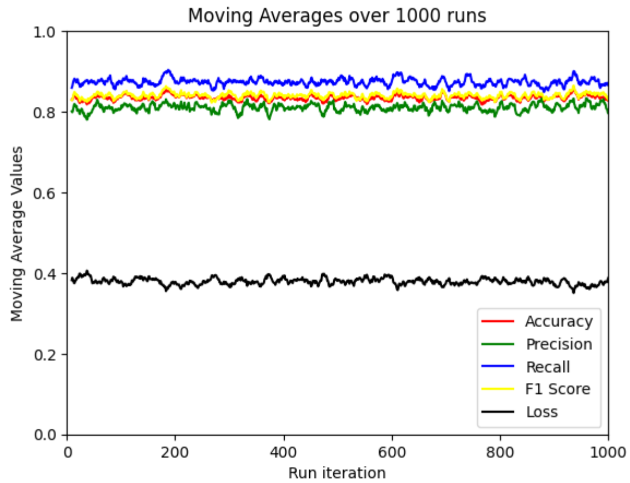

Figure 3 Moving averages for performance metrics (n=1000, M.A. window=10)

The presented data indicates the model proposed achieves the purpose it was created for. It shows strong consistency regardless of the testing data provided, with an emphasis on reliable classification.

V. Conclusion

In this study, an Artificial Neural Network model for diagnosing pneumonia was proposed and implemented. The constructed model achieved a peak accuracy of 91% and a low overall false negative rate (pneumonia-positive classified as negative), matching and surpassing the results in existing literature with this approach. Furthermore, the model's consistency was carefully evaluated through multiple test runs using random sampling without replacement, where it showed stable performance across various metrics.

The study showcases the potential deep learning techniques for quick, and accurate medical diagnoses of pneumonia using x-ray images. By leveraging large datasets, preprocessing and an efficient but powerful neural network architecture, it was possible to achieve a remarkably high level of accuracy in diagnosis, presenting a significant step in improving patient outcomes.

Moving forward, the approach presented, and the model can further be applied to other binary classification medical images tasks, such as detecting respiratory diseases or detecting cancerous cells. Additionally, Convolutional Neural Networks as well as the two-layer high resolution approaches are to be explored.

In conclusion, the study demonstrates the effectiveness of Artificial Neural Networks in the diagnosis of pneumonia and provides a strong foundation for future research in this area. With the model provided, it is possible ultimately to improve patient outcomes, making a significant impact on public health.

Bibliography

[1] National Heart, Lung, and Blood Institute, "Pneumonia," National Institutes of Health, U.S. Department of Health and Human Services. [Online].

[2] J. Rajaraman and V. Antani, "Automated detection of pneumonia from chest X-ray images using artificial intelligence and machine learning algorithms," IEEE Access, vol. 6, pp. 32382-32394, 2018.

[3] Jain, R., Nagrath, P., Kataria, G., Kaushik, V. S., & Hemanth, D. J. "Pneumonia detection in chest X-ray images using convolutional

neural networks and transfer learning," in 2020 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), pp. 1-5.

[4] Rahman, T., Chowdhury, M. E. H., Khandakar, A., Islam, K. R., Islam, K. F., Mahbub, Z. B., Kadir, M. A., & Kashem, S. "Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray," in 2020 IEEE Region 10 Symposium (TENSYMP), pp. 493-496.

[5] Rahman, T., Chowdhury, M. E. H., Khandakar, A., Islam, K. R., Islam, K. F., Mahbub, Z. B., Kadir, M. A., & Kashem, S. "Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray," in 2020 IEEE Region 10 Symposium (TENSYMP), pp. 493-496.

[6] Ayan, E., Karabulut, B., & Ünver, H. M. "Diagnosis of pediatric pneumonia with ensemble of deep convolutional neural networks in chest x-ray images," in 2022 IEEE International Conference on Image Processing and Communications (ICIPC), pp. 277-282.

[7] M. Khadivi, M. T. Masouleh, and S. Khademi, "Classification of Pneumonia Based on Chest X-Ray Images Using Artificial Neural Networks," 2019 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI), Tehran, Iran, 2019, pp. 89-94, doi: 10.1109/KBEI.2019.8878762.

[8] R. Rojas, "Artificial Neural Networks," in Encyclopedia of Computer Science and Technology, vol. 26, pp. 1-13, 1992, doi: 10.1081/E-ECTO-120011406.

[9] G. Huang, Q. Zhu, and C. Chen, "Deep learning and its applications to machine health monitoring," Mechanical Systems and Signal Processing, vol. 115, pp. 213-237, 2019.

[10] Ruder, S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.

[11] D. Kumar, S. Suresh, and S. Ilango, "A survey on optimization algorithms for deep learning," Journal of Ambient Intelligence and Humanized Computing, vol. 11, no. 11, pp. 4181-4200, Nov. 2020. doi: 10.1007/s12652-020-02576-9.

[12] Jain, S., Williams, D. J., Arnold, S. R., et al. "Community-acquired pneumonia requiring hospitalization among U.S. children," in 2015 New England Journal of Medicine, vol. 372, no. 9, pp. 835-845.

[13] Mandell, L. A., Wunderink, R. G., Anzueto, A., et al. "Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults," in 2007 IEEE International Conference on Bioinformatics and Biomedicine, pp. S27-S72.

[14] Rajpurkar, P., Irvin, J., Ball, R. L., et al. "Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists," in 2018 IEEE International Conference on Healthcare Informatics, pp. 1-6.

[15] Li, L., Qin, L., Xu, Z., et al. "Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT," in 2020 IEEE International Conference on Multimedia and Expo (ICME), pp. 1-6.

[16] Wang, S., Kang, B., Ma, J., et al. "A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)," in 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 1-7.

[17] Majumdar, A., Kuppili, V., Gudigar, A., et al. "Classification of lung CT images using multi-resolution CNN and LSTM networks," in 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP), pp. 1162-1166.

[18] R. Kundu, S. Sarkar, and A. M. Das, "Pneumonia detection in chest X-ray images using an ensemble of deep learning models," Biomedical Signal Processing and Control, vol. 62, 2020, pp. 102136.

[19] M. N. Khalid and K. M. A. Naser, "Comparative classification of Pneumonia using Convolution Neural Network," 2019 5th International Conference on Advances in Electrical Engineering (ICAEE), Dhaka, Bangladesh, 2019, pp. 1-5.

[20] S. S. Kumar, S. Kumar, S. S. Kumar, and R. Gupta, "Deep learning aided automated COVID-19 and other pneumonia detection schemes," Computer Methods and Programs in Biomedicine, vol. 202, 2021, pp. 106051.

[21] M. M. Rahman, M. H. U. Siddiquee, and A. K. M. Bahalul Haque, "Automatic Detection of Bacterial and Viral Pneumonia Using Deep Convolutional Neural Networks," Applied Sciences, vol. 10, no. 20, 2020, pp. 7165.

[22] A. Al-Jawad, S. Othman, and S. A. Mohammed, "Pneumonia Detection Using Convolutional Neural Networks," 2018 International Conference on Advanced Science and Engineering (ICOASE), Baghdad, Iraq, 2018, pp. 1-5.

[23] Y. Zhang, X. Xue, H. Xiong, Y. Liu, and Y. Wang, "Computer-Aided Diagnosis of Pneumonia in Children Using Convolutional Neural Network Ensemble Method," IEEE Access, vol. 7, 2019, pp. 123906-123914.

[24] Y. Zhu, Y. Hu, J. Qi, and J. Xu, "Bayesian optimization of convolutional neural network architecture for pneumonia detection and classification," Computers in Biology and Medicine, vol. 135, 2021, pp. 104543.

[25] Y. Chen, H. Wang, M. Luo, H. Ding, L. Li, and M. Xu, "Pneumonic Lungs Detection Based on Deep Convolutional Neural Network," 2018 4th International Conference on Bioinformatics Research and Applications (ICBRA), Shenzhen, China, 2018, pp. 60-65.

[26] S. S. Islam, M. H. Islam, M. J. Uddin, M. N. Uddin, and T. K. Kundu, "Classification and Discrimination of Pulmonary Infections from Chest X-ray Images using Neural Network Classifier," 2019 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), Dhaka, Bangladesh, 2019, pp. 138-142.

[27] S. A. Al-Jawad, S. Othman, and S. A. Mohammed, "Detection of Pneumonia in Chest X-Ray Images Using Deep Learning," 2019 International Conference on Engineering and Technology (ICET), Baghdad, Iraq, 2019, pp. 1-6.

[28] S. Ince and A. Bilge, "A Comparative Study of Different Neural Network Models for Pneumonia Detection from Chest X-ray Images," 2021 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Istanbul, Turkey, 2021, pp. 1-6, doi: 10.1109/HORA51817.2021.9534291.

[29] J. Fawcett, "An introduction to ROC analysis," Pattern Recognition Letters, vol. 27, no. 8, pp. 861-874, 2006.

[30] G. Huang, Q. Zhu, and C. Chen, "Deep learning and its applications to machine health monitoring," Mechanical Systems and Signal Processing, vol. 115, pp. 213-237, 2019.

Appendix A – Github Repository Link