Attention Enhanced Explainable Ensemble Framework for Pneumonia Detection

Swathi Padarthi[1], B D Dheeraj Anand[2], B Harikiran[3],
Brahmandlapally Manaswini[4], Noone Manikeshav[5]
Dept. of CSE (CyS, DS) and AI&DS
VNR Vignana Jyothi Institute of Engineering & Technology
Hyderabad, India
[1]swathi.522@gmail.com, [2]dheerajanandbhoomi@gmail.com,
[3]harikiran7989@gmail.com, [4]manaswini.bhramandlapally2@gmail.com,
[5]keshavmani713@gmail.com

***Abstract*—The complete workflow of the proposed system is illustrated in Fig. 1. The first step involves obtaining the chest X-ray images from various datasets followed by performing several preprocessing activities like resizing, normalization, and lung region segmentation to ensure that their sizes are standardized while excluding any extra components from them. In order to achieve more robustness and generalization, a diverse set of augmentation methods like rotation, flipping, zooming, and shifting is employed to increase the dataset size. After that, the augmented images are provided to deep neural networks like ResNet50, MobileNetV2, and DenseNet121 based on the transfer learning concept. Simultaneously, these neural networks are capable of extracting hierarchal deep features from the images and classifying them into two categories, including pneumonia and normal cases. Soft voting ensemble technique is implemented on their outputs using the average of probability scores to obtain higher accuracy and stability of the final result. An attention-based technique is further used to give more emphasis to the infected lung parts by weighting them. Lastly, Grad-CAM is implemented to interpret the working principle of the models.**

***Index Terms*—Pneumonia Detection, Chest X-ray, Deep Learning, Ensemble Learning, Explainable AI, Transfer Learning, Grad-CAM**

I. Introduction

Pneumonia is an infectious illness of the lung organ that causes many deaths and disabilities around the world. It occurs more frequently among children, older people, and individuals with immune system deficiencies. The process of detecting pneumonia at its early stages can be done through chest X-ray image recognition. Nevertheless, this method poses the challenge of inconsistent diagnoses by experts since such a process demands particular competencies.

There is a rising trend of employing machine learning systems that automate the process of recognizing illnesses based on medical pictures. Alongside other models, deep learning systems are successfully applied for pneumonia identification. For example, transfer learning techniques that utilize convolutional neural networks like ResNet, VGG, MobileNet, and DenseNet121 are applied because of their high effectiveness and accuracy. Nonetheless, current algorithms still face some challenges with regard to overfitting, poor generalizability, and limited attention on particular areas.

In order to overcome the problems of the existing deep learning techniques, the idea of applying ensembles was born. Yet, the strategy for developing ensembles is quite straightforward, assuming that all characteristics are equally important in making predictions. This does not work for medical images since each part carries unique clinical importance.

However, another impediment associated with the use of this deep learning model is that it lacks interpretabillty. As a result, doctors have trouble trusting the results produced by this framework. Some possible solutions to the mentioned problem include the use of XAI approaches such as Grad-CAM since it provides visual explanation through heatmaps, allowing us to see which areas of the image influence the decision-making process of the model.

In light of the challenges discussed above, the present research proposes an approach called *Attention-Enhanced Explainable Ensemble Framework (A3EF)* that can be used to diagnose patients with pneumonia on the basis of chest X-ray images. This model is based on transfer learning-based CNNs, soft voting technique in ensembling, and Grad-CAM as the method of making the decision-making process understandable.

The main contributions of this work are summarized as follows:

- Development of an attention-enhanced ensemble framework for improved pneumonia detection accuracy.
- Implementation of a soft voting ensemble to enhance robustness and generalization.
- Incorporation of Grad-CAM for visual interpretability of model predictions.
- Evaluation on multiple datasets to ensure reliability and practical applicability.

Experimental results demonstrate that the proposed framework achieves an accuracy of 96%, outperforming individual baseline models while providing interpretable and clinically relevant predictions.

II. Literature Survey

In recent times, the application of deep learning methods has been significantly effective in the automatic detection of pneu-

monia from chest X-rays. Earlier research has focused on the design of CNNs using transfer learning techniques like VGG, MobileNet, ResNet, and DenseNet. The pre-trained weights of these neural networks are based on ImageNet database. As a result, they provide acceptable classification accuracy even for a relatively small dataset of medical images. While a number of researchers have achieved impressive results through the classification using these CNNs, one disadvantage of single model classification is the potential risks of overfitting and generalization.

Ensemble learning methods can be applied effectively in medical image analysis to resolve the shortcomings of single models. Some methods, such as bagging, boosting, and stacking, are useful in integrating the outputs from several models to increase their classification accuracy, robustness, and stability. By integrating the outputs of various networks, it becomes possible for an ensemble classifier to have improved generalization capabilities. When it comes to pneumonia detection, ensemble methods perform better than single CNNs concerning classification accuracy. Nevertheless, a critical challenge in traditional ensemble learning is that it involves a simple approach to combining outputs without considering spatial dependencies.

The second trend worth mentioning is the development of attention mechanisms that can be used in order to enhance feature representation within deep learning systems. Using attention mechanisms allowed the researchers to significantly improve the classification of medical images since it was possible to focus on the particular areas where the certain characteristics were detected, remove all irrelevant factors, and make this process more interpretable.

However, another crucial issue associated with the introduction of deep learning algorithms in clinical practice is the problem of interpretability. Although deep learning algorithms tend to work well, they often become black boxes due to the complexity of calculations that makes them difficult to implement. Therefore, approaches associated with XAI become especially popular. One of such solutions is Grad-CAM and SHAP.

Further, recent studies have emphasized the importance of using multiple source datasets in order to create models that are more generalized and robust. Models that train only from one type of dataset will not work effectively during testing due to the challenges faced such as domain shifts and biases of data. Multiple source datasets are important in solving this challenge as they provide variability for various types of images, patient conditions and diseases.

Although progress is being made regarding individual studies for transfer learning, ensembles, attention network models and AI explainability, there is limited research that combines all these aspects in one paper. Additionally, many of the studies available in literature have not done an assessment of the effectiveness of these models using multiple datasets besides considering the interpretability of these systems.

Thus, there is the need to formulate an interpretability framework that uses advanced attention-based feature extraction techniques and an ensemble classifier combined with explainable artificial intelligence. In this regard, the study seeks to fill this gap by developing the Attention Enhanced Explainable Ensemble Framework.

III. Problem Statement

The situation of pneumonia is one of the crucial conditions that have to be diagnosed timely in order not to cause deaths of children and older people. One of the most common methods of diagnoses is related to the examination of chest using X-rays due to simplicity and relatively inexpensive cost of the technique. On the other hand, implementation of this method poses several challenges since it requires significant knowledge and experience from a radiologist.

There were many studies conducted in the field of computer science concerning the issue that have been able to develop systems of deep learning that could successfully diagnose the patients with pneumonia basing on the images of chest. Nevertheless, there were some major disadvantages in such techniques that included inability to learn generalizations of samples and inability to identify relationships among the zones of images. Moreover, complicated structure of such algorithms made their interpretation rather difficult.

Thus, it makes sense to introduce an improved technique to make the processes easier.

IV. Objectives

The main objective of this research is to develop an efficient and interpretable deep learning framework for pneumonia detection using chest X-ray images. The specific objectives are as follows:

- To develop a pneumonia detection system using transfer learning-based CNN models.
- To implement an ensemble learning approach to improve model performance, robustness, and generalization.
- To address class imbalance issues using techniques such as class weighting and focal loss.
- To integrate Explainable Artificial Intelligence (XAI) techniques, such as Grad-CAM, for visual interpretation of model predictions.
- To train and evaluate the model on multiple datasets to ensure improved generalization and real-world applicability.
- To achieve high classification accuracy while maintaining interpretability and reliability for potential clinical use.

V. Methodology

A. System Overview

The suggested approach would use the Attention Enhanced Explainable Ensemble Framework (A3EF) for developing an automated detection system for pneumonia using chest X-ray images. In general, the A3EF follows four main phases. Firstly, chest X-rays are collected from diverse datasets where the images undergo some pre-processing steps such as normalization, resizing, and segmentation of lungs. Data augmentation is performed to diversify the existing datasets.

During the second phase, transfer learning-based CNN models such as ResNet50, MobileNetV2, and DenseNet121 are used to extract meaningful deep features from the input images. Specifically, each of the models is further trained for binary classification that classifies the images into pneumonia and healthy patients. To ensure accuracy and minimize the bias in model predictions, an ensembling method named soft voting is applied that merges the prediction outputs of all models. An attention module is further integrated into the framework to emphasize the relevant portions of the lungs.

The final step towards addressing the significance of interpretability in the field of medicine is to incorporate the use of Grad-CAM in the model. The justification for incorporating Grad-CAM is that it facilitates the visualization process through heat maps, which highlight the contributing regions of the model's decision-making process.

B. Dataset

Multiple datasets of chest X-rays are employed to enhance generalization and robustness. The first dataset comprises chest X-rays of children and has been categorized into two classes, namely *Normal* and *Pneumonia*. The second dataset involves lung masks corresponding to the first dataset, to promote regional-based learning.

C. Data Preprocessing

All the pictures included in the databases are normalized in terms of dimensions and resized to the dimensionality of (224 * 224) pixels in order to ensure consistency regarding the input sizes utilized in the deep learning models. Standardizing the input sizes of images helps in achieving efficiency in processing batches as well as ensuring that the convolutional neural network can be trained effectively. Following resizing, the segmentation of lungs is performed using the help of binary masks. This technique involves multiplication of the images with binary masks, thereby removing all the unnecessary background components like rib bones, markers, and other tissue elements. By following this technique, it is ensured that the machine learning model will only pay attention to the lung fields in the input images.

Following segmentation, the images are then merged to create an even more varied database of images. In order to prevent overfitting of the model and improve its generalization capability, various data augmentation techniques have been utilized. These data augmentation techniques include random rotation of images based on patient positioning, horizontal flipping of the images to allow for symmetry variation, zooming of images to cover scale variation, and spatial shifting of images.

D. Class Imbalance Handling

For balancing the class distribution, class weighting and focal loss methods are used. The weight for each class is calculated based on the automatic calculation, which will ensure balanced contribution of classes. On the other hand, focal loss helps in focusing on hard examples.

E. Model Architecture

The selected deep learning architecture will be a deep convolutional neural network, complemented by transfer learning method. In particular, modifications will be introduced into such well-known architectures as ResNet50, MobileNetV2, and DenseNet121. First, the considered architecture is trained using large ImageNet database where the architecture is able to create hierarchical features. Then, in the proposed system, the last classifier layer of each architecture is modified and replaced with custom-designed binary classifier that determines whether a patient has pneumonia or not.

Furthermore, besides replacing the classifier layer with a new one, another modification is made to the existing deep learning architecture. Specifically, attention is used in order to select those elements of the image that are of high importance for further processing by the network. With the use of attention, the model is able to pay attention only to the most critical parts of an image to detect whether there is any pathology present in it. In particular, thanks to the use of attention in the model, its attention is directed to the areas of lungs that tend to become infected.

F. Training Strategy

Attention-Enhanced Explainable Ensemble Framework's Training Approach The training approach of the Attention-Enhanced Explainable Ensemble Framework is designed to optimize learning efficiency, increase the model's generalization capability, and ensure robustness in various datasets. To begin with, all prepared and transformed chest X-ray images will be divided into training, validation, and testing sets so that the model can be assessed effectively without leaking information between the datasets. The process of transfer learning is used, wherein the ResNet50, MobileNetV2, and DenseNet121 architectures are pretrained using ImageNet weights and thus learn from their feature extraction steps.

The techniques of class balancing, such as class weighting and focal loss, are applied during the training phase to address the issue of class imbalance and ensure that the algorithm considers hard examples. The augmentation process is performed dynamically during training in order to introduce variation and prevent overfitting. The optimization process involves optimizing with the aid of an optimizer such as Adam, among others, together with optimal learning rates and learning rate schedules.

Training is conducted independently per model but with identical conditions for all to produce unique features. Following the training stage, the model with the best weights that performs optimally on the validation dataset is selected. Finally, a weighted voting scheme is applied whereby predictions made by the models trained are combined using the averaging of probabilities. This structured approach ensures the achievement of optimal accuracy and performance.

G. Ensemble Learning

Models ResNet50, MobileNetV2, and DenseNet121 learn independently on the same dataset, which is preprocessed and

augmented. Hence, the models learn different features due to differences in their architecture and the ability of models to learn. In other words, it is important to mention that this feature is essential in the construction of an effective ensemble. Moreover, using different models makes a contribution to the increase of diversity, which is also essential in building ensembles. Every model learns probability distribution for each example for two classes: pneumonia and normal.

In the proposed approach, soft voting is utilized for aggregation of results provided by the models. The difference from simple voting lies in the fact that the results obtained with different levels of confidence are considered. In order to compute the average probabilities of the classes, we have to average all values obtained for models. After that, we select the class having the highest average probability.

Such an approach to result aggregation improves stability and effectiveness of the algorithms. In case of discrepancy in results for some samples between models, it helps to reduce a negative effect on the outcome. Thus, a more general and accurate solution can be provided.

H. Explainability Module

For increasing the accuracy of results, Grad-CAM technique is used for generating the heat maps indicating the regions affecting the model's decision-making process. These interpretations help in identifying whether the model focuses on important clinical regions of lungs or not.

I. Workflow

A schematic diagram indicating the process flow for the proposed architecture is provided in Fig. 1, indicating a complete process flow from data collection to generation of interpretable outputs. The process begins with the collection of chest X-ray images from various freely available datasets, ensuring that diversity is considered in terms of the demographic characteristics of the individuals involved and the environment in which the imaging process took place. In the next step, the images undergo some preprocessing operations such as scaling the image size to a fixed size, normalizing the pixel intensities, and segmenting the lungs using a binary mask to eliminate the background image.

Upon successful completion of the pre-processing stage, the next phase that follows is called data augmentation, where the amount of data in the dataset is expanded in order to enhance the generalization capability of the model when faced with unknown data. The use of data augmentation techniques like rotation, flipping, zooming, and translation is applied in order to change the conditions of the imaging process. After training the models with the augmented data, the output is then passed through several deep learning algorithms such as ResNet50, MobileNetV2, and DenseNet121 using a technique known as transfer learning.

Following this, the predictions from each individual classifier are pooled using the soft voting method, which averages the probability predictions. The accuracy increases as a result since it becomes less sensitive to the mistakes made by each individual model. Finally, the visualization is created using the Grad-CAM method in order to identify the regions of the chest X-ray image that have impacted the predictions made by the model. This explanation will enable better comprehension and understanding of the predictions for medical practitioners.

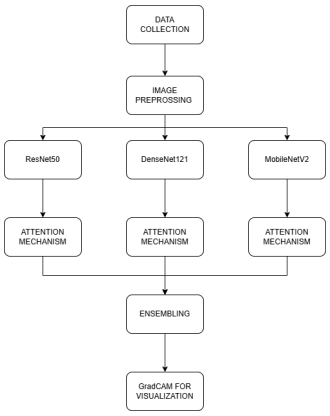

Fig. 1. Workflow of the proposed Attention-Enhanced Explainable Ensemble Framework for pneumonia detection.

VI. Results and Discussion

Evaluation of the Attention Enhanced Explainable Ensemble Framework performance is achieved by testing with different data sets with chest x-rays classified into two categories, normal and pneumonia. In this study, deep learning technique has been implemented utilizing GPUs, and the visualization method used is Grad-CAM.

A. Training and Validation Performance

Accuracy charts from training and validation are shown in Fig. 2, while loss plots are presented in Fig. 3. The performance of the model achieves a steady state of validation accuracy of about 96%.

There is a steady decline in the validation loss that signifies the success of convergence of the neural network. The incorporation of attention networks helps focus on learning infection regions.

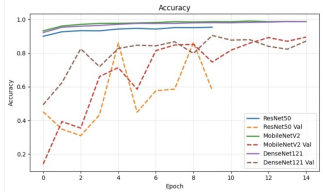

Fig. 2. Training vs. Validation Accuracy Curve

B. Comparative Analysis with Base Models

To evaluate the effectiveness of the proposed framework, comparative experiments were conducted with individual base models, namely ResNet50, MobileNetV2 and DenseNet121. The results are summarized in Table 1.

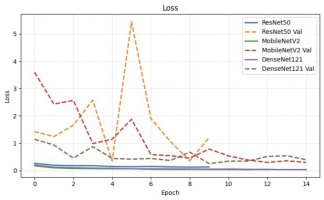

Fig. 3. Training vs. Validation Loss Curve

TABLE I
COMPARISON OF MODEL PERFORMANCE

Model	Accuracy (%)	Precision	Recall	F1-Score
ResNet50	92.68	0.93	0.92	0.92
MobileNetV2	95.61	0.96	0.95	0.95
DenseNet121	96.73	0.96	0.96	0.96
Proposed Ensemble	**96.60**	**0.96**	**0.96**	**0.96**

As can be seen from Table I, the proposed framework performs better than each individual model in accuracy and classification results.

C. Interpretation Using Grad-CAM

For the problem of explainability in deep learning systems, Grad-CAM was used for generating visual interpretation of results. Representative images generated using this technique are provided in Fig. 4 and Fig. 5.

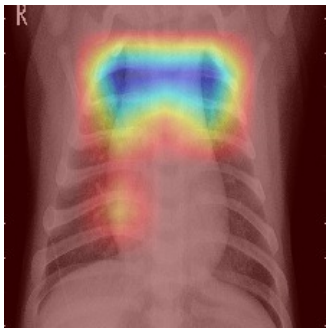

Fig. 4. Grad-CAM Visualization for Pneumonia Case

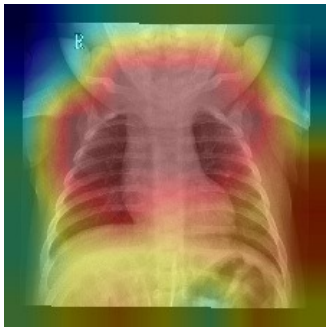

Fig. 5. Grad-CAM Visualization for Normal Case

Based on the visualizations, it can be said that the algorithm concentrates on clinically important parts of the lungs. In the case of pneumonia patients, the identified locations include those which have been affected by the infection, whereas in the normal subjects, the attention has been paid to all healthy parts of the lungs.

D. Interpretation of Confusion matrix

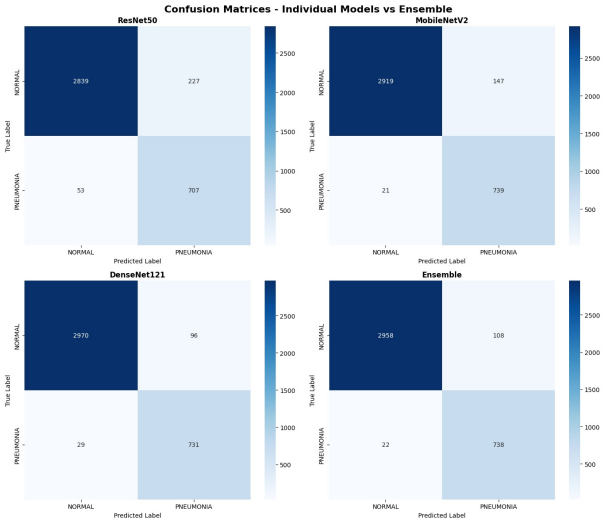

Fig. 6. Confusion Matrix of the Proposed Model

VII. CONCLUSION

The proposed research seeks to address one of the most critical issues, which is the automation of the process of diagnosing pneumonia through chest X-rays by means of the Attention-Enhanced Explainable Ensemble Framework (A3EF). By means of the proposed technique, the researcher succeeded in combining the properties of convolutional neural networks based on transfer learning, attention mechanisms to improve feature localization, soft voting ensemble technique to increase predictive accuracy, and Grad-CAM methodology to provide model interpretability.

Based on the empirical study of the results obtained via the application of the proposed methodology, the framework has managed to obtain the highest possible degree of accuracy, amounting to 96%. In comparison with the baseline individual models and other approaches described above, the proposed methodology demonstrated greater efficiency. In particular, the use of the ensemble technique helped avoid overfitting and increased the stability of the algorithm, while attention mechanisms facilitated higher accuracy of localization of the affected areas.

Moreover, the application of explainable AI techniques, in particular, Grad-CAM, makes it possible to visualize the process of decision making within the proposed model. In this case, the increase in transparency and confidence is guaranteed as far as the opinion of medical specialists is concerned. In this regard, one should say that it is extremely significant to consider the process of explainability in addition to the accuracy of results in the field of medicine.

Finally, based on what has been said above, the suggested solution is considered highly efficient as far as the enhancement in accuracy, robustness, and explainability are concerned. Speaking about future development in the sphere, there are many ways to go; some of them are multi-class classification, sophisticated attention techniques, and application in a practical setting.

In general terms, one may assume that the suggested solution is rather advantageous to diagnose various thoracic diseases using computer systems.

REFERENCES

[1] H. Bysania, S. Garg, A. Danda, T. Singh, J. C, and P. Duraisamy, "Detection of pneumonia in chest X-ray using ensemble learners and transfer learning with deep learning models," Proc. 14th IEEE Int. Conf. Computing, Communication and Networking Technologies (ICCCNT), Indian Institute of Technology Delhi, India, Jul. 6–8, 2023.

[2] A. Sharma, A. Verma, S. Verma, N. Duhan, and P. Priyanka, "Attention-enhanced CBAM-VGG-16 for optimized pneumonia detection in chest X-ray imaging," Proc. IEEE Int. Conf. Computer, Electronics, Electrical Engineering and their Applications (IC2E3), 2025.

[3] Sharma, Ayush, et al. "Attention-Enhanced CBAM-VGG-16 for Optimized Pneumonia Detection in Chest X-Ray Imaging." 2025 IEEE International Conference on Computer, Electronics, Electrical Engineering their Applications (IC2E3). IEEE, 2025.

[4] Cyriac, Sibu, Nidhin Raju, and Yong-Woon Kim. "Pneumonia Detection using Ensemble Transfer Learning." 2022 13th International Conference on Information and Communication Technology Convergence (ICTC). IEEE, 2022.

[5] Gunasundari, B., and R. Thiagarajan. "Attention-Enhanced Ensemble Transfer Learning Framework for Lung and Colon Cancer Prediction." International Conference on Communication and Computational Technologies. Singapore: Springer Nature Singapore, 2025.

[6] Jana, Yudhajit, and N. Vinutha. "Attention-Enhanced Transfer Learning for Emphysema Classification Using CBAM-Augmented ResNet-50." 2025 6th International Conference on Recent Advances in Information Technology (RAIT). IEEE, 2025.

[7] Khater, Omar H., et al. "Attcdcnet: Attention-enhanced chest disease classification using x-ray images." 2025 IEEE 22nd International Multi-Conference on Systems, Signals Devices (SSD). IEEE, 2025.

[8] Jahanian, Mojtaba, et al. "AXNet: Attention-enhanced X-ray network for pneumonia detection." Biomedical Signal Processing and Control 118 (2026): 109618.

[9] Potharaju, Saiprasad, et al. "Enhanced X-ray image Classification for Pneumonia Detection Using Deep Learning Based CBAM and SE Mechanisms." Intelligence-Based Medicine (2025): 100299.

[10] Yang, Yuting, Gang Mei, and Francesco Piccialli. "A deep learning approach considering image background for pneumonia identification using explainable AI (XAI)." IEEE/ACM Transactions on Computational Biology and Bioinformatics 21.4 (2022): 857-868.

[11] Zou, Lin, et al. "Ensemble image explainable AI (XAI) algorithm for severe community-acquired pneumonia and COVID-19 respiratory infections." IEEE Transactions on Artificial Intelligence 4.2 (2022): 242-254.

[12] Behera, Saroj Kumar, K. Murali Gopal, and Sudheer Babu Punuri. "A Deep Learning-Based Pneumonia Detection System with Explainable AI for Medical Decision Support." 2025 11th International Conference on Communication and Signal Processing (ICCSP). IEEE, 2025.

[13] Dagnaw, Getamesay Haile, and Meryam El Mouthadi. "Towards explainable artificial intelligence for pneumonia and tuberculosis classification from chest x-ray." 2023 International Conference on Information and Communication Technology for Development for Africa (ICT4DA). IEEE, 2023.

[14] Pai, Prathiksha P., and Sarika Hegde. "Early Detection of Pneumonia Using Deep Learning Approach." Artificial Intelligence: First International Symposium, ISAI 2022, Haldia, India, February 17-22, 2022, Revised Selected Papers. Cham: Springer Nature Switzerland, 2023.

[15] Wang, Tianmu, et al. "PneuNet: deep learning for COVID-19 pneumonia diagnosis on chest X-ray image analysis using Vision Transformer." Medical Biological Engineering Computing (2023): 1-14

[16] Kiliçarslan, Serhat, et al. "Detection and classification of pneumonia using novel Superior Exponential (SupEx) activation function in convolutional neural networks." Expert Systems with Applications 217 (2023): 119503

[17] Modak, Sudipta, Esam Abdel-Raheem, and Luis Rueda. "Applications of Deep Learning in Disease Diagnosis of Chest Radiographs: A Survey on Materials and Methods." Biomedical Engineering Advances (2023): 100076.

[18] Sharma, Shagun, and Kalpna Guleria. "A Deep Learning based model for the Detection of Pneumonia from Chest X-Ray Images using VGG-16 and Neural Networks." Procedia Computer Science 218 (2023): 357-366.