Comparative Analysis of Hybrid Deep CNN Models in the Classification of Pulmonary Diseases

Md. Muhaiminul Kabir muhaiminulkabir32@gmail.com
Nishi Kanta Paul nishikantapaul108@gmail.com
Md. Shihabul Islam Shovo shihabul900@gmail.com
Kazi Atiqur Rahman iamatiq7@gmail.com

***Abstract*—The global impact of the COVID-19 pandemic emphasizes the critical need for accurate diagnostic tools to distinguish it from other respiratory conditions. Deep learning, particularly Convolutional Neural Networks (CNNs), is highlighted for its ability to automatically extract complex patterns from medical images, enhancing diagnostic accuracy and facilitating early detection. The integration of deep learning in healthcare reduces human error, provides consistent results, and ensures robust performance across diverse datasets. Previous research demonstrates the effectiveness of CNNs in pulmonary disease classification through the analysis of chest X-rays and CT scans. The paper employs various CNN models on a large-scale combined dataset to classify pulmonary diseases, aiming to enhance accuracy and efficiency in disease classification with various class distributions. The experiments highlights great results of individual models, like DenseNet201 and InceptionV3 but poor performance for ensemble models.**

***Index Terms*—Neural Networks, Convolution Neural Networks, Ensemble models, Pulmonary diseases**

I. Introduction

The global impact of the coronavirus pandemic, which started in December 2019 [1], has disrupted economies, industries, and agriculture worldwide. The outbreak initially reported in Wuhan, China, led to a widespread halt in development as people prioritized self-protection. As COVID-19 evolved into a pandemic, declared by the WHO on March 11, 2020 [1], the challenge emerged in distinguishing it from other lung diseases. This was particularly tricky due to similarities in symptoms with typical pneumonia. The visual resemblance in lung opacities on X-rays resulted in misdiagnoses, inadvertently spreading the highly contagious disease. This highlights the urgent need for accurate diagnostic tools, like advanced technologies, to help healthcare providers differentiate COVID-19 from other respiratory conditions, ultimately enhancing patient care.

The integration of deep learning in the classification of pulmonary diseases brings forth several benefits to healthcare. Particularly, deep learning models, particularly Convolutional Neural Networks (CNNs), demonstrate a remarkable ability to automatically extract complex patterns from medical images, aiding in precise disease identification [2], [3]. This not only enhances diagnostic accuracy but also facilitates early detection of pulmonary conditions. The application of deep learning reduces the potential for human error and subjectivity in diagnoses, providing more consistent and reliable results. Moreover, these models are capable of handling large and diverse datasets, ensuring robust performance across various patient demographics. The efficiency and speed of deep learning algorithms streamline the diagnostic process, offering healthcare professionals valuable tools for swift and accurate pulmonary disease classification. Ultimately, the adoption of deep learning in healthcare contributes to improved patient care and treatment outcomes.

Previous research in the field of pulmonary disease classification has demonstrated the effectiveness of CNN in enhancing diagnostic accuracy [4]–[6]. Studies have utilized CNNs to analyze medical imaging data, such as chest X-rays and CT scans [7], showcasing their ability to automatically extract relevant features for disease identification.

However, many of these researchers used a small amount of data [4] to train their deep learning structures. A shortage of data can cause generalization issues while classifying pulmonary diseases. To solve this issue, the use of pre-trained CNN models emerged as an effective solution. [6]–[8].

In this paper, We collected a large amount of Pulmonary X-ray images and we employed several popular CNN architectures to classify pulmonary diseases using a large amount of X-ray images. The study focuses on evaluating the performances of pre-trained models on X-ray images of lungs. Additionally, a combination of multiple datasets was used, aiming to enhance the accuracy and generalized classification of pulmonary diseases. This research contributes to this specific domain as we used a combination of multiple lung X-ray image datasets which can determine the maximum learning capability of deep learning structures.

Section II focuses on relevant studies, and Section III demonstrates the characteristics of the datasets we used in our experiments. Section IV con- includes the methodology we used in the proposed research. In section V, we assess the CNN models' performance using the combined dataset, and section VI provides our conclusion.

II. Related works

We conducted a literature review and provided a detailed analysis of the most recent developments in machine learning and deep learning.

A. Machine Learning

Boban and Megalingam [10] classified lung disorders using a variety of machine learning techniques; KNN achieved 99.2% accuracy, MLP 98%, and SVM 70.45%. They used a

collection of 400 lung illness images. The study was limited by the small dataset and its focus on a few specific lung diseases.

Spathis and Vlamos [11] investigated machine learning methods for diagnosing respiratory conditions, achieving accuracy rates of 97.7% and 80.3% for different conditions. Utilizing a dataset of 132 patient entries, the research concentrated on variables such as age, smoking status, and respiratory metrics.

Westcott et al. [12] developed a machine-learning model using CT texture analysis to predict lung ventilation heterogeneity in patients with respiratory conditions. They trained a model on 67 subjects and tested it on 27, resulting in 88% accuracy and an AUC of 0.82. To establish ground truth, they merged specialized MRI ventilation maps with CT scan data. The study's limitations involved concentrating on more advanced cases and omitting certain lung areas to simplify data alignment.

B. Deep Learning

Mukherjee et al. [13] introduced a neural network to identify COVID-19 using chest X-rays and CT images. The nine-layer model, composed of alternating convolutional and dense layers, was optimized with specific activations and dropout methods to minimize overfitting. Using diverse radiological data, the model achieved an accuracy of 96.28% in identifying COVID-19.

FethyaSeid Yimer et al. [14] proposed a deep learning method to categorize various lung diseases. The study, which focused on six diseases, used more than 12,000 X-ray scans from regional and NIH datasets. The model's performance reached 97.3%. The study addresses limitations such as limited data (295 images) and imbalanced classes.

Arias-Londono et al. [15] presented a study evaluating the application of deep learning models to detect COVID-19 from X-ray images. The model takes 224x224-pixel center-cropped X-ray images as input. The study developed a CNN model achieving 91.5% accuracy and 87.4% recall. However, the classification is limited to only three classes.

Horry et al. [16] investigated deep-learning models to detect COVID-19 from chest X-ray scans. The study tests popular models like deep learning models on small publicly available datasets of X-ray images. Despite the limited dataset, VGG19 outperformed other models with 83% precision and an F1 score of 80%, while deeper models like ResNet50 and Xception underperformed due to data constraints.

Given these works, we can summarise that there are a lot of areas to work on in this field especially evaluating CNN models with large datasets. In this work, we used a combination of multiple datasets to increase the learning capability of the pre-trained CNN models. The proposed study includes the evaluation of three popular CNN architectures: ResNet50, InceptionV3, and Densnet201 on the combined dataset to improve pulmonary disease classification.

III. Datasets

This section highlights the features of the datasets used in our study. We integrated a combination of seven datasets in our work to achieve a well-balanced distribution among the classes.

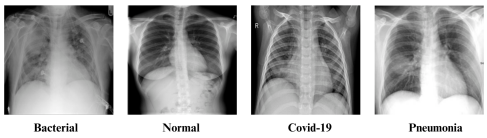

Fig. 1. Images of different classes

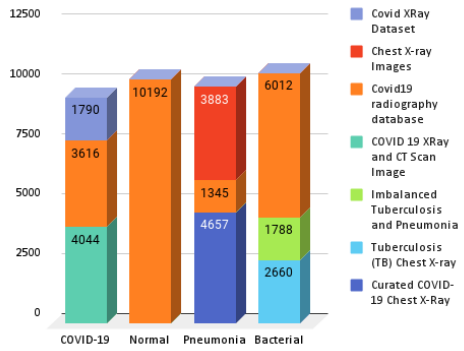

Fig. 2. Combined Dataset

In our work we merged lung X-ray images from 'Chest X-ray dataset' [18], 'COVID X-ray dataset' [19],' Curated COVID-19 Chest X-Ray Dataset' [20], 'COVID-19 XRay and CT Scan Image' [21], 'Tuberculosis (TB) Chest X-ray Cleaned Database' [22] and 'Imbalanced Tuberculosis and Pneumonia dataset' [23] and the 'Covid-19 Radiography Database' [17]. In our research, we selected 4 types of X-ray images for classification as shown in Fig. 1. Among them, three classes are disease-based and one class is for normal lung X-ray images. The classes are: Covid-19, Pneumonia, Bacterial infection and Normal. To create the combined datasets, we used both binary and multiclass datasets. When utilizing multiclass datasets, we only used the classes that were not balanced in the combined dataset. The 'COVID X-ray dataset' and the 'Chest X-ray dataset', are both binary class datasets. There are 3,883 images of pneumonia-affected lung X-ray images from the 'Chest X-ray dataset' and 1,790 images of COVID-19-affected lung X-ray images were merged from the COVID-19 X-ray dataset. From the 'COVID-19 X-ray and CT Scan Image dataset', only the X-ray images of lungs(4,044 images) were used to balance the COVID-19 class. We used the 'Curated COVID-19 Chest X-Ray Dataset' to balance the pneumonia class. The pneumonia class was balanced by merging pneumonia-affected images of the 'Curated COVID-19 Chest X-ray dataset' (4,657 images). To balance the Bacterial Infection class we used Tuberculosis affected lung X-rays from the 'Tuberculosis (TB) Chest X-ray Cleaned Database'(2,660 images) and the 'Imbalanced Tuberculosis and Pneumonia dataset'(1,788 images) as Tuberculosis is caused by bacteria [24]. Normal lung images were taken from the Covid-19 Radiography Database(10,192).

Fig. 2 demonstrates the class distribution of the combination. The combination results in a balanced distribution of 9,450 COVID images, 10,192 normal images, 9,885 pneumonia images, and 10,460 images related to bacterial infection.

IV. Methodology

This section covers the step-by-step approach to our proposed work. We start with pre-processing X-ray images and move on to data splitting, model training and classification for optimal results as shown in Fig. 3. Everything is summarized briefly for a straightforward understanding of our methodology further in this section.

A. Preprocessing

Before training our models, we prepared the combined datasets by resizing them to a standard size of (224 x 224). In the combined dataset, some images were RGB but the majority of the images were grayscaled. To match the dimension of each data, we converted the RGB X-ray images to grayscale images.

B. Splitting

In our experiment, we split the combined dataset into training and testing sets, with 80% allocated for training and 20% for testing. We set up our dataset and divided it for training and testing. The training set was used to train the pre-trained models and the test set was utilized to obtain classification results.

C. Model Training

In our research, we utilized pre-trained deep CNN models on the ImageNet dataset, specifically DenseNet201, InceptionV3, and ResNet50, as fundamental architectures for our work. We removed the top layers and froze the feature extraction sections. Subsequently, we introduced additional fully connected hidden layers and a linear output layer, initializing the hidden layer parameters randomly with a rectified linear unit activation function. To enhance generalization, we incorporated dropout layers into the models. Given the four target classes in our combined dataset, we configured the output layer with four neurons.

Our training process involved 80% of our X-ray images. We chose the SGD (Stochastic Gradient Descent) optimizer with a low learning rate to stabilize training and facilitate the convergence of the fully connected layers towards optimal solutions. Looking specifically at the selected architectures:

- **ResNet50**: With 50 layers, ResNet50 tackles common issues in deep networks by using "shortcut" connections, improving learning through the inclusion of initial information at the end. Its use of "residual blocks" ensures a smooth flow of information, making it particularly effective for image recognition tasks.
- **InceptionV3**: This intelligent neural network uses "inception modules" to capture diverse details in images through clever filter manipulations. With a deep architecture featuring multiple modules, InceptionV3 excels in recognizing complex visual patterns, making it highly efficient for tasks such as image classification.
- **DenseNet201**: Designed for image tasks, DenseNet201 stands out by fostering direct connections between all layers, promoting efficient information sharing through its "dense blocks" within its 201-layer architecture. This design choice makes DenseNet201 a powerful and effective option for tasks involving image recognition.

Alongside these deep learning models, we used ensemble architecture by using a combination of two models from the three CNN models. After applying the architecture, we evaluated two more models Ensemble-01 and Ensemble-02, which were also evaluated utilizing the combined dataset. Ensemble-01 is the combination of DenseNet201 and ResNet50. On the other hand, Ensemble-02 is configured by an ensemble of DenseNet201 and InceptionV3.According to the architecture demonstrated in Fig. 4, we ensembled the output of fine-tuned pre-trained models and attached a top layer consisting of a hidden layer and an output layer with 4 output neurons.

In summary, our research utilized these pre-trained models as baselines, adapting them to our specific task through modification of their architectures and training on our dataset, ultimately showcasing the performances of ResNet50, InceptionV3, and DenseNet201 in the context of image-related tasks.

D. Classification

The classification results were obtained using the trained CNN models on combined datasets. As mentioned previously, the combined dataset was separated into testing and training sets. In the process of classification, 20% images were used to test the models after training to classify 4 pulmonary disease classes. We used a variety of metrics to evaluate model performance in addition to accuracy, including precision, recall, and F1-score. Our evaluation framework includes a detailed analysis of the models' categorization report. Obtained results of the experiment are discussed in section V.

V. Experiment Results

The performance of the chosen models is shown in this part based on the outcomes of testing and training the pre-trained models. For effective model training and analysis, we used high-performance systems with GPU support servers for our experiments. Model performance was compared for both imbalanced and balanced distributions as part of the evaluation process. The models' capacity for categorization was examined using metrics like F1-score, recall, accuracy, and precision in classification.

Hyperparameter adjustment was crucial to our efforts to improve model performance. In particular, we concentrated on changing the batch sizes during training—a critical factor affecting the convergence and effectiveness of the model. Using 4, 8, 32, and 128 batch sizes, we carefully evaluated their effect on the overall performance of the model. Based on our experimentation, we found that a batch size of 32

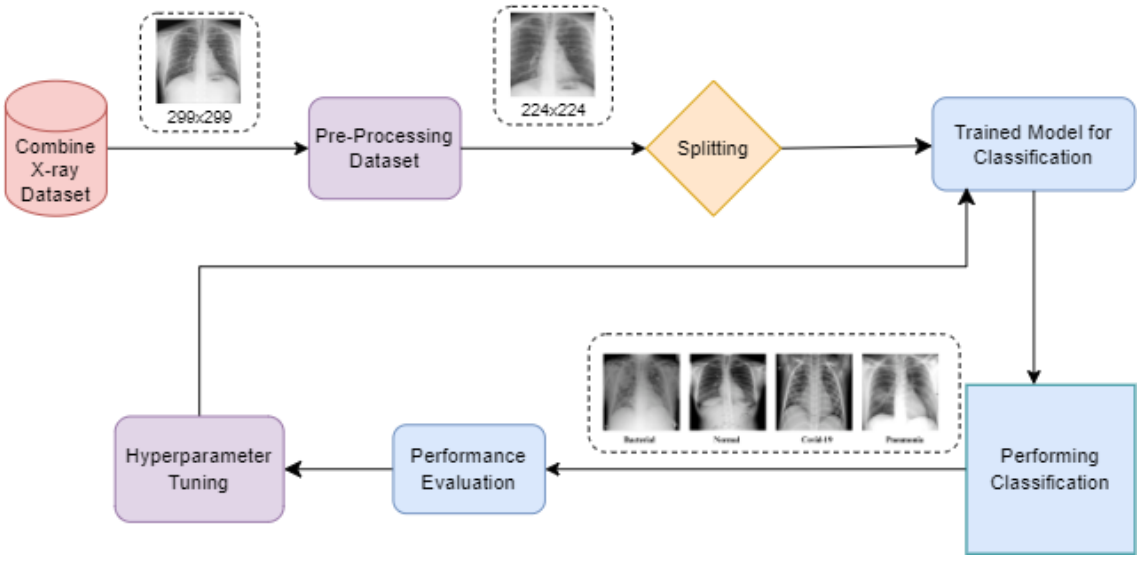

Fig. 3. Overall work process of Pulmonary Diseases Classification

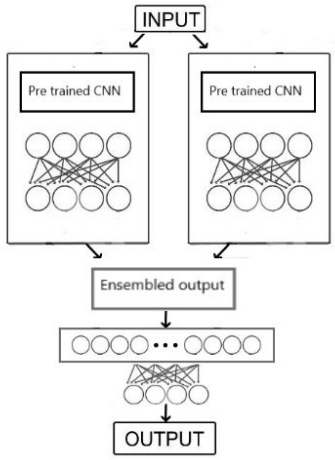

Fig. 4. Ensemble model architechture

consistently produced the best results out of all the sizes evaluated.

TABLE I demonstrates the class-wise Classification report obtained by utilising deep CNN models and Fig. 5 illustrates the overall classification accuracy of each CNN model for imbalanced and balanced classification.

VI. Result Analysis

The analysis of various CNN models for respiratory conditions reveals distinctive performance characteristics across different categories. For the Covid-19 class, InceptionV3 exhibits superior precision (96%) but a low F1-score (75%), outperforming ResNet50 and DenseNet201 in precision. In the Normal class, all models demonstrate robust performance, with InceptionV3 consistently leading in precision, recall, and F1-score. Notably, DenseNet201 performs average in the Pneumonia class with precision (77%) and competitive recall (71%), while ResNet50 and InceptionV3 achieved higher recall. In the Bacterial class, DenseNet201 proves highly effective with the highest precision (96%) and recall (92%).

On the other hand, comparative analysis between ensemble models (Ensemble-01 and Ensemble-02) and individual models (ResNet50, InceptionV3, DenseNet201), shows certain trends across disease classes. In the Covid19 class, Ensemble-01 and Ensemble-02 exhibit precision values of 65% and 64%, respectively, lower than ResNet50 (87%) and InceptionV3 (96%), but their recall and F1-scores remain competitive. For the Normal class, both ensembles show comparable performance, with Ensemble-02 excelling in recall with a value of 88%. In the Pneumonia class, Ensemble-01 maintains precision and recall close to individual models, while Ensemble-02 slightly lags in recall with a value of 75%. In the Bacterial class, both ensembles demonstrate good precision (Ensemble-01: 75%, Ensemble-02: 83%) and recall (Ensemble-01: 83%, Ensemble-02: 75%).

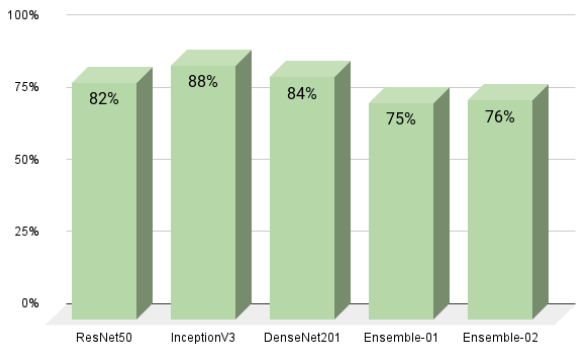

Fig. 5. The Comparison Study of Different Models Accuracy

In the overall accuracy assessment in Fig 5, ResNet50 achieves 82%, InceptionV3 attains 88%, and DenseNet201

TABLE I
COMPARISON OF DIFFERENT MODELS (P= PRECISION, R= RECALL, F= F1-SCORE)

Models/Class	Covid19			Normal			Pneumonia			Bacterial		
	P(%)	R(%)	F(%)	P(%)	R(%)	F(%)	P(%)	R(%)	F(%)	P(%)	R(%)	F(%)
ResNet50	87	55	67	84	89	87	74	98	84	88	84	86
InceptionV3	96	62	75	94	95	94	74	98	85	94	93	94
DenseNet201	69	76	73	92	96	94	77	71	74	96	92	92
Ensemble-01	65	59	62	83	78	83	77	80	78	75	83	79
Ensemble-02	64	60	62	79	88	81	76	80	79	83	75	79

scores 84%. The ensembles, Ensemble-01 and Ensemble-02, contribute with 75% and 76% accuracy, respectively. These results provide a summary of the models' comprehensive performance across various respiratory conditions, indicating InceptionV3's strong overall accuracy and the ensembles offering slightly lower accuracy values.

VII. CONCLUSION AND FUTURE WORKS

This study compared ResNet50, InceptionV3, and DenseNet201 in pulmonary disease classification, revealing distinct strengths and weaknesses. The analysis highlights great results of individual models, like DenseNet201 and InceptionV3, which show high recall, precision, and F1 scores in a variety of disease classes. It is important to point out, that there is a performance trade-off with the ensemble models, Ensemble-01 and Ensemble-02, which show somewhat lower precision and recall values. Models have distinct advantages and limitations, underscoring the need to select the appropriate architecture depending on the dataset. To enhance the performance of deep learning models, addressing class imbalance is crucial. Increasing dataset size and diversifying examples, especially for medical image classification can improve generalization. In this research, we collected lung X-ray images from different sources and merged them together to achieve a well-balanced dataset which was utilized to determine the best performances of the models. In the process of merging different datasets to create a balanced dataset, the reliability and authenticity of the data were ensured. Exploring complex CNN architectures should be prioritized for future research. Continuous efforts in refining models and datasets are essential for advancing pulmonary disease classification.

REFERENCES

[1] World Health Organization. 2020. WHO Announces COVID-19 Outbreak a Pandemic. [Online] Available at: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020

[2] Rahaman, Md & Li, Chen & Yao, Yudong & Kulwa, Frank & Rahman, Mohammad & Wang, Qian & Qi, Shouliang & Kong, Fanjie & Zhu, Xuemin & Zhao, Xin. (2020). Identification of COVID-19 samples from chest X-Ray images using deep learning: A comparison of transfer learning approaches. Journal of X-ray science and technology. 28. 10.3233/XST-200715.

[3] Albahri OS, Zaidan AA, Albahri AS, Zaidan BB, Abdulkareem KH, Al-Qaysi ZT, Alamoodi AH, Aleesa AM, Chyad MA, Alesa RM, Kem LC, Lakulu MM, Ibrahim AB, Rashid NA. Systematic review of artificial intelligence techniques in the detection and classification of COVID-19 medical images in terms of evaluation and benchmarking: Taxonomy analysis, challenges, future solutions and methodological aspects. J Infect Public Health. 2020 Oct;13(10):1381-1396. doi: 10.1016/j.jiph.2020.06.028. Epub 2020 Jul 1. PMID: 32646771; PMCID: PMC7328559.

[4] S. Liu and W. Deng, "Very deep convolutional neural network based image classification using small training sample size," 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), Kuala Lumpur, Malaysia, 2015, pp. 730-734, doi: 10.1109/ACPR.2015.7486599.

[5] Khan, Asif & Shah, Junaid & Bhat, Mohammad. (2020). CoroNet: A Deep Neural Network for Detection and Diagnosis of COVID-19 from Chest X-ray Images. Computer Methods and Programs in Biomedicine. 196. 105581. 10.1016/j.cmpb.2020.105581.

[6] Alruwaili, Madallah & Shehab, Abdulaziz & Abd ElGhany, Sameh. (2021). COVID-19 Diagnosis Using an Enhanced Inception-ResNetV2 Deep Learning Model in CXR Images. Journal of Healthcare Engineering. 2021. 1-16. 10.1155/2021/6658058.

[7] Purohit, Kiran & Kesarwani, Abhishek & Kisku, Dakshina Ranjan & Dalui, Mamata. (2020). COVID-19 Detection on Chest X-Ray and CT Scan Images Using Multi-image Augmented Deep Learning Model. 10.1101/2020.07.15.205567.

[8] Zargari, Abolfazl & Heidari, Morteza & Shariati, Shujaat. (2021). COVID-Classifier: an automated machine learning model to assist in the diagnosis of COVID-19 infection in chest X-ray images. Scientific Reports. 11. 10.1038/s41598-021-88807-2.

[9] Wang, Xiaosong & Peng, Yifan & Lu, Le & Lu, Zhiyong & Bagheri, Mohammadhadi & Summers, Ronald. (2017). ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. arXiv:1705.02315.

[10] Boban, Binila Mariyam, and Rajesh Kannan Megalingam. "Lung diseases classification based on machine learning algorithms and performance evaluation." 2020 International Conference on Communication and Signal Processing (ICCSP). IEEE, 2020.

[11] Spathis D, Vlamos P. Diagnosing asthma and chronic obstructive pulmonary disease with machine learning. Health Informatics Journal. 2019;25(3):811-827. doi:10.1177/1460458217723169

[12] Westcott, A., Capaldi, D. P., McCormack, D. G., Ward, A. D., Fenster, A., & Parraga, G. (2019). Chronic obstructive pulmonary disease: thoracic CT texture analysis and machine learning to predict pulmonary ventilation. Radiology, 293(3), 676-684.

[13] Mukherjee H, Ghosh S, Dhar A, Obaidullah SM, Santosh KC, Roy K. Deep neural network to detect COVID-19: one architecture for both CT Scans and Chest X-rays. Appl Intell (Dordr). 2021;51(5):2777-2789. doi: 10.1007/s10489-020-01943-6. Epub 2020 Nov 6. PMID: 34764562; PMCID: PMC7646727.

[14] Yimer, Fethyaseid & Tessema, Abel & Simegn, Gizeaddis. (2021). Multiple Lung Diseases Classification from Chest X-Ray Images using Deep Learning approach. International Journal of Advanced Trends in Computer Science and Engineering. 10. 2936-2946. 10.30534/ijatcse/2021/021052021.

[15] Arias-Londono JD, Gomez-Garcia JA, Moro-Velazquez L, Godino-Llorente JI. Artificial Intelligence Applied to Chest X-Ray Images for the Automatic Detection of COVID-19. A Thoughtful Evaluation Approach. IEEE Access. 2020 Dec 14;8:226811-226827. doi: 10.1109/ACCESS.2020.3044858. PMID: 34786299; PMCID: PMC8545248.

[16] Horry, M. J., Chakraborty, S., Paul, M., Ulhaq, A., Pradhan, B., Saha, M., & Shukla, N. (2020). X-ray image based COVID-19 detection using pre-trained deep learning models.

[17] https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database

[18] https://www.kaggle.com/datasets/tolgadincer/labeled-chest-xray-images

[19] https://www.kaggle.com/datasets/ahemateja19bec1025/covid-xray-dataset

[20] https://www.kaggle.com/datasets/francismon/curated-covid19-chest-xray-dataset

[21] https://www.kaggle.com/datasets/ssarkar445/covid-19-xray-and-ct-scan-image-dataset

[22] https://www.kaggle.com/datasets/tawsifurrahman/tuberculosis-tb-chest-xray-dataset

[23] https://www.kaggle.com/datasets/roshanmaur/imbalanced-tuberculosis-and-pnuemonia-dataset

[24] Tuberculosis [Online] Available at: https://www.who.int/news-room/fact-sheets/detail/tuberculosis