 e-ISSN: 2614-8404
p-ISSN: 2776-3234

Improved Lung Disease Classification Using Bagging and Averaged Ensemble Models

Hana Ali Tayib[1*], Azar Abid Salih[2]
[1, 2] Duhok Polytechnic University, Technical College of Administration, Department of Information Technology Management – Kurdistan Region – Iraq
Email: [1]hanaa.tayib@dpu.edu.krd, [2]azar.abid@dpu.edu.krd

Abstract – One of the essential medical imaging tasks for early diagnosis and treatment planning is categorizing lung diseases from chest X-ray (CXR) images. This work constructs a strong ensemble learning platform on a variety of deep models for boosting diagnosis performance to detect and identify lung disease. Three prtrianed CNN models InceptionV3, ResNet50, and EfficientNetV2M were trained on a CXR dataset, motivated by the complementary architectural features and the success demonstrated in medical imaging problems, such as chest X-rays. These three networks belong to different families of the CNNs and therefore make different contributions for diversity and stability in the ensemble. The models were then ensembled in two methods: averaging (soft voting) and bagging with hard voting (maximum bootstrap aggregation) in the first method. Various sets of pre-trained models were experimented with for the averaged ensemble. According to experimental results, the soft voting (averaged) ensemble between EfficientNetV2M and InceptionV3 performed better than the other models' combinations and achieved the highest accuracy of 93.00% in classification. This was followed by the combination of EfficientNetV2M and ResNet50 with an accuracy of 92.09%, then InceptionV3 and ResNet50 with a value of 91.75%, and the complete ensemble of the three models with an accuracy of 92.14%.The bagging hard voting strategy was somewhat with lower accuracy, but the InceptionV3 based bagging ensemble attained 90.56%, EfficientNetV2M attained 91.00%, and ResNet50 attained 88.00%. It is evident from the results that soft voting strategy, InceptionV3 and EfficientNetV2M ensemble provides the best optimal and stable classification performance among all the configurations that were attempted. The study proves that ensemble learning improves the accuracy of lung disease classification models, and choosing the right architectures is essential, with EfficientNetV2M and InceptionV3 showing improved performance, resulting in early diagnosis and improved patient outcomes.

Keywords – Lung Diseases, CNN, Ensemble Learning, X-ray, InceptionV3 Model, ResNet50 Model, EfficientNetV2M, Soft Voting.

I. INTRODUCTION

Lung diseases remain a serious global health threat due to their high morbidity and mortality rates. An accurate and timely diagnosis is essential for diseases such as COVID-19, TB, pneumonia, idiopathic pulmonary fibrosis, and chronic obstructive pulmonary disease (COPD). This is necessary in order to better manage treatment and avoid complications. Chest X-ray (CXR) technology is the most prevalent diagnostic imaging modality among all other modalities currently in use [1]. This is mainly because of its clinical utility in lung abnormality detection, low cost, and ease of use. CT scanning using deep learning has significantly improved detection and disease classification of infections like COVID-19, pneumonia, and lung cancer. Convolutional neural networks can segment infected regions, detect disease types, and classify severity with extremely high accuracy. But then there remain problems like false positives, data set variability, interpretability, and large datasets and annotation [2].

Magnetic Resonance Imaging (MRI) offers more advantages over CT in detecting and classifying lung diseases. It avoids ionizing radiation, making it valuable for vulnerable populations. MRI provides structural and functional imaging, allowing for differentiation of benign versus malignant lesions, inflammatory foci, or disease states. In some scenarios, MRI demonstrates comparable accuracy to CT, especially for larger pulmonary nodules. Combining MRI with deep learning can improve image quality, contrast, and precise quantification of disease severity, improving classification and diagnosis [3].

Deep Learning (DL) and, more notably, Convolutional Neural Networks (CNNs) have seen significantly better performance in medical image processing over the last few years. Self-discovery of complex features from raw pixel data sets directly is the strategy that is employed in order to achieve the above goal [4]. Although when applied in variable or imbalanced data, DL models can be prone to architectural bias, overfitting, and bad generalizability [5], which is especially so when the sets of data are applied in unbalanced datasets.

Ensemble learning approaches have gained popularity as a way of mitigating the effects of these issues [6]. These approaches entail the combination of many models to better maintain prediction stability, accuracy, and robustness compared to single models [7].

through the use of ensemble approaches, the use of computed tomography (CXR) images to identify lung conditions has been significantly boosted [8],[9] These are capable of detecting more types of discriminative features and decision patterns because they use different CNN architectures and then fuse their outputs by applying a range of techniques such as soft voting, hard voting, or stacking [10]. This allows them to learn a wider range of discriminative features and decision patterns. Ensembles learned on balanced datasets tend to reduce class bias, which will allow for more balanced performance across a

e-ISSN: 2614-8404
p-ISSN: 2776-3234

range of sickness classes [11]. Further, ensembles learned on balanced datasets are more accurate [12].

The article here introduces a robust ensemble DL model that has been designed with focus towards automatic identification and classification of various lung conditions. Through the use of the strength of various pre-trained CNN models combined with ensemble methods, the new model attempts to offer better diagnostic performance and reliability and act as a robust tool in clinical decision support for pulmonary disease screening. The main contributions of this study can be summarized as follows:

• Using a new dataset, we achieved superior results compared to existing methods

• This study proposes two ensemble approaches bagging and averaging to enhance lung disease classification and compare their effectiveness. It introduces a deep bagging ensemble framework using three advanced CNN models: EfficientNetV2M, InceptionV3, and ResNet50, aiming to improve classification accuracy and robustness.

• The models were restored to the weights corresponding to the highest validation accuracy achieved during training, ensuring optimal performance.

This paper is structured as follows: section 2 provides an overview of related works conducted about previous studies. Section 3 describes the proposed method, which consists of data description, image enhancement, image classification using ResNet, EfficientNetV2M, and Inception v3 architectures, in addition to explaining the assembly methods and evaluation. Section 4 describes the research results and discussions. Section 5 present comparison with previous studies Section 6 concludes this paper with future works.

II. METHODOLOGY

This article proposes a model based on deep learning for five-class multi-class chest X-ray image classification. The whole process involves pre-processing of image data, the development of three independent convolutional neural networks (CNNs), the separate training of each, and finally their ensemble with an ensemble method to improve the accuracy of classification as shown in Figure1.

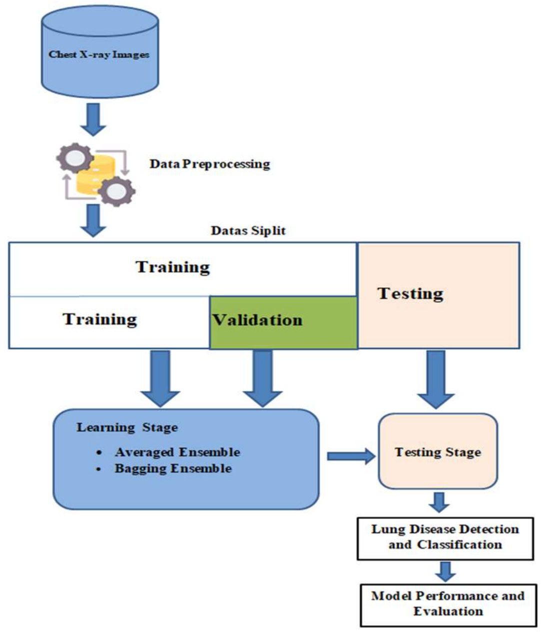


Figure1. The main stages of the proposed methodology

a. Dataset and Label Preparation

The data used within this study is a labeled dataset of chest X-ray. The dataset was obtained from Kaggle https://www.kaggle.com/datasets/omkarmanohardalvi/lungs-disease-dataset-4-types, an open-source platform that provides publicly available medical imaging datasets for research purposes, contain 10095 images,(6054) for training, (2016) for validation and (2025) for testing. Images to five clinically significant categories: Bacterial Pneumonia, Coronavirus Disease, Tuberculosis, Viral Pneumonia and Normal X-ray as shown in Figure2. Data are divided into train, validation, and test sets, and labels are derived from directory structure. Automatic label assignment was performed through uniform mapping schema to map class names to numeric identifiers. To prevent order-bias in training, the samples were randomly shuffled prior to consumption by the models [21].

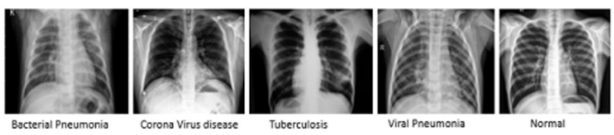


Figure2. Lung Diseases Chest X-Ray Images [21]

b. Data split

Quality, quantity, and how data are processed are the key elements in effective deep learning models, particularly in health-related processing like medical imaging [22]. Convolutional Neural Networks (CNNs) use bulk data for their classification by them. Splitting data into training, validation, and testing is the manner in which effective learning takes place, overfitting is avoided, and very good generalization of the model occurs, particularly in health-related tasks like medical imaging, as given in Table1.

Table 1. The Dataset Split

e-ISSN: 2614-8404
p-ISSN: 2776-3234

Class	Training	Validation	Testing
Viral Pneumonia	1204	401	403
Tuberculosis	1220	406	408
COVID-19	1218	406	407
Bacterial Pneumonia	1205	401	403
Normal	1207	402	404

c. Data Preprocessing

To gain consistency among all the neural network models, all the images were re-sized to a uniform resolution of 224 × 224 pixels. An image processing pipeline of high performance was established to perform decoding, resizing, batching, and prefetching operations. The training dataset was randomized with a sufficient buffer for better randomization, but the validation and test datasets were given a fixed order for the sake of evaluation consistency. These preprocessing steps resulted in optimal memory and training performance and also avoided architecture-specific preprocessing for normalizing inputs to all models [23].

d. Model Architecture Design

The model leverages three pre-trained convolution neural network models namely InceptionV3, EfficientNetV2-M, and ResNet50, all of which were transferred via transfer learning from large natural images. For each of the models, the initial classification layers were removed and replaced with a global pooling layer and a fully connected classification head for five-class classification. This adaptation allowed for reuse of generalized feature extraction capability and conversion of the model into classification-specific to a domain. The models execute independently and produce a condensed distribution over the five target categories.

1. InceptionV3

InceptionV3, which is a deep convolutional neural network model, is effective in image classification tasks like the diagnosis of lung diseases. It utilizes parallel convolutional layers to learn local and global features of chest X-ray images. Techniques like factorized convolutions, auxiliary classifiers, and batch normalization are used to improve training performance and model generalization. Here, it was fine-tuned for multiple lung diseases[24],[25].

2. EfficientNetV2M

EfficientNetV2M is a convolution neural network architecture that equilibrates for the given amount of accuracy, training time, and computational cost. It utilizes a compound scaling technique to learn high-quality feature representations from chest X-ray images at a low computational cost. The model was fine-tuned on a balanced chest X-ray database for the development of the model to enhance the accuracy of classification and to extract deep features [26],[27].

3. ResNet50

ResNet50 is one of the deep convolutional neural network architectures that has become popular due to its strong performance in image classification, including lung disease detection. It uses residual learning to mitigate the vanishing gradient problem and contains 50 layers. ResNet50 is particularly useful in chest X-ray classification when it comes to detecting subtle hierarchical features, increasing diagnostic capacity in various classes of lung disease [28],[29].

e. Ensemble Bootstrap Sampling "Bagging"

The process of training in this work is built on a bagging ensemble learning method aimed at enhancing the generalization and robustness of lung disease classification from chest X-ray images. Training proceeds by first generating five independent training subsets using bootstrap sampling, each made up of 3,000 random samples (with replacement) of the original training set. For each subset, a deep learning model based on the EfficientNetV2M architecture is constructed and trained separately for 30 epochs. This results in the creation of five diverse models, each of which is trained on a separate distribution of the data. In inference, all five trained models predict for the input image, The predictions of all the models are combined, and the predicted class is determined based on the highest agreement (in hard voting). This ensemble decision is the most confident classification of the models. Following the decision of training and ensemble strategy, the system proceeds to evaluate performance on standalone test dataset employing metrics such as accuracy, precision, recall, and F1-score to assess the performance of the merged model As shown in Figure 3.

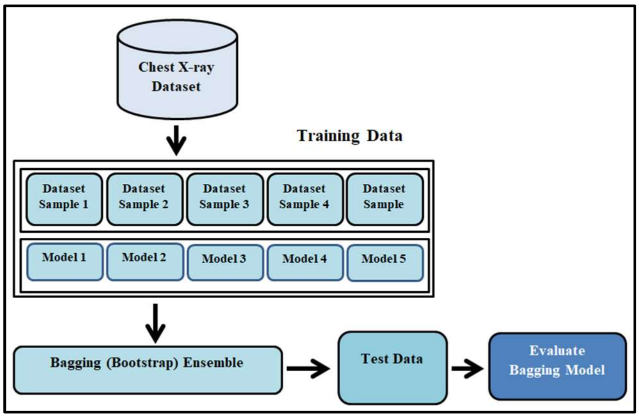


Figure 3. Bagging Ensamble (EfficientNetV2M, InceptionV3, and ResNet50)

f. Averaged Ensemble

Soft voting ensemble inference was applied. Individual class probability distributions for test samples were produced by each of the three separately trained models. All model predictions were averaged by taking probabilities per class. Each sample's final prediction was

e-ISSN: 2614-8404
p-ISSN: 2776-3234

obtained by selecting the class with the maximum mean probability.
This models ensemble was chosen due to its efficacy in breaking the bounds of a single model and to aggregate the strength of each of multiple disparate architectures. Use of heterogeneous models improves generalizability and significantly enhances robustness in classification, particularly in challenging diagnosis issues such as discrimination of pneumonia subtypes. Figure 4. displays models ensemble with soft voting.

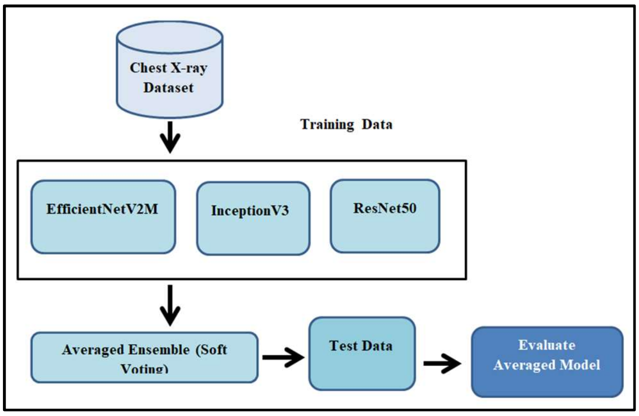


Figure 4. Averaged Ensamble Models (EfficientNetV2M, InceptionV3, and ResNet50)

g. Training Strategy

Averaged Ensemble Model Architecture All models employed in the ensemble training method begin with an input layer to receive resized chest X-ray images of equal 224×224 pixels size with three color channels. The input is fed through an existing pre-trained convolutional neural network, i.e., EfficientNetV2M, InceptionV3, or ResNet50, as the base models. These models are used without their topmost initial layers (include_top=False) to allow for lung disease classification transfer learning. A Global Average Pooling layer is used following the convolutional layers to decrease the spatial dimension of feature maps and the generation of a fixed-length feature vector for classification. Finally, a Dense output layer with softmax activation is added to predict the probability distribution over the target classes (i.e., two or three disease classes, depending on which dataset is utilized). During training, each model is compiled with the Nadam optimizer under a Triangular Cyclical Learning Rate (TCLR) schedule for the sake of better convergence. The models are separately trained with the same data under Sparse Categorical Crossentropy loss function and tested with a test set. The trained models are utilized for ensemble in prediction by selecting the class with highest average of predicted probabilities as the final prediction. This method employs architectural diversity and aggregation to enhance the stability and accuracy of classification, as shown in Table 2.

Table 2. The architecture of Averaged Ensemble Model Layers and parameters Settings

Layer type	Output shape	parameters
input_1 (InputLayer)	[(None, 224, 224, 3)]	0
lambda (Lambda)	(None, 224, 224, 3)	0
inception_v3 (Functional)	(None, 2048)	21802784
dense (Dense)	(None, 5)	10245

Total params: 21813029 (83.21 MB)
Trainable params: 21778597 (83.08 MB)
Non-trainable params: 34432 (134.50 KB)

Table 3: indicates the hyperparameters used Averaged Ensemble Model during training. The hyperparameters have a critical role to play in the tuning of models' performance in diagnosing lung disease using chest X-ray images. The parameters were carefully tuned with the aid of easy tuning, which was suitable and efficient for the task.

Table3. The hyperparameters employed in Averaged Ensemble Model during training.

Hyperparameter	Value
Loss function	Sparse_Categorical_Crossentropy
Classification function	SoftMax
Batch size	32
epoch	30
Learning rate range (cyclical)	1e-6 to 1e-3
Optimizer	Nadam

The training approach of all three ensemble models EfficientNetV2M, InceptionV3, and ResNet50 is a standard architectural pattern with an input layer which is capable of accepting chest X-ray images of dimensions 224×224×3. All models employ a pre-trained convolutional base (EfficientNetV2M, InceptionV3, or ResNet50) with initialization of weights as ImageNet where the top classification layers are removed (include_top=False) for enabling custom fine-tuning. According to the base model, there is Global Average Pooling layer usage to reduce high-dimensional feature maps into high-density feature vectors for generalization and dimensionality reduction. Next, a Dense output layer with softmax activation is added for five-class lung disease classification. The models are trained on Nadam optimizer and Sparse Categorical Crossentropy loss function. In addition to that, an adaptive training learning rate is achieved by using a Triangular Cyclical Learning Rate (TCLR) scheduler. For promoting diversity and robustness, each model type is trained five times with different bootstrap samples of the training data (3,000 images per subset) to produce a bagging ensemble. After being trained, ensemble models are merged using hard voting, where the prediction is the mode of the five models that were each trained independently on the data. The technique improves generalization and classification performance via model diversification as well as data variability. As evident from table 4.

e-ISSN: 2614-8404
p-ISSN: 2776-3234

Table 4. The architecture of Bagging Ensemble Model layers and parameters settings

Layer type	Output shape	parameters
input_1 (InputLayer)	[(None, 224, 224, 3)]	0
lambda (Lambda)	(None, 224, 224, 3)	0
efficientnetv2-m (Functional)	(None, 1280)	53150388
dense (Dense)	(None, 5)	6405

Total params: 53156793 (202.78 MB)
Trainable params: 52864761 (201.66 MB)
Non-trainable params: 292032 (1.11 MB)

Table 5: presents the hyperparameters used in Bagging Ensemble Model during training. These hyperparameters are crucial for optimizing the models' performance in lung disease detection from chest X-ray images. The parameters were carefully optimized through convenient tuning, which proved to be suitable and effective for the task.

Table 5: Gives the hyperparameters used in Bagging Ensemble Model for training. The hyperparameters were used to improve the performance of the models in classifying lung diseases in chest X-ray images. The parameters were easy to optimize, which was effective and suitable for the task.

Table5: The hyperparameters employed in Bagging Ensemble Model during training.

Hyperparameter	Value
Loss Function	Sparse_Categorical_Crossentropy
Classification function	SoftMax
Batch size	32
epoch	30
Learning rate range (cyclical)	1e-6 to 1e-3
Optimizer	Nadam

h. Performance Evaluation

The system's performance was evaluated using the following simple measures:
• Accuracy: Ratio of the correctly predicted samples to all samples.
• Precision (Weighted): Weighted average of positive predictive values for classes.
• Recall (Weighted): Weighted average of true positive rates for classes.
• F1-Score (Weighted): Harmonic mean of recall and weighted precision.
Confusion matrix plot was also done to see how the predicted and actual labels are distributed such that discriminative capacity of the model between disease classes can be realized.

III. Results and Discussion

The figures in the table show the power of the Ensemble Bootstrap Sampling technique, or "Bagging," when separately used on three contemporary deep neural network models EfficientNetV2M, InceptionV3, and ResNet50 each of which was trained as an ensemble of five models for 30 iterations. Among the models, EfficientNetV2M ensemble bagging scored highest with 91% accuracy, 91.01% precision, 91.16% recall, and 91.05% F1 score. It is extremely sensitive and specific balance-wise and thus extremely robust in multi-class lung disease classification. For a comparison, InceptionV3 ensemble yielded lower values of 90.56% accuracy and 90.50% F1 score, which means similar predictability but lack of robustness. The worst was ResNet50 ensemble with 88% accuracy and 87.63% F1 score, which means relatively poor generalization ability with this ensemble configuration. These outcomes validate that ensemble bagging greatly improves the classifiability and that model architecture exerts an excellent impact on the performance of the overall ensemble. EfficientNetV2M, owing to its scalability and parameterization, is most appropriate to use in ensemble bagging for medical image classification tasks like lung disease detection, as evident from Table 3.

Table 3. Results of bagging ensambling learning model results

Models	Accuracy	Precision	Recall	F1 Score
EfficientNet V2M	91	91.01	91.16	91.05
InceptionV3	90.56	90.50	90.56	90.50
ResNet50	88	87.60	87.80	87.63

Figure 5. Displays EfficientNetV2M model, after bagging via the bagging ensemble technique, shows high classification accuracy and generalizability in multi-class Corona Virus Disease, Normal, Tuberculosis, and Viral Pneumonia detection from chest X-ray images. The model shows excellent accuracy in Corona Virus Disease, Normal, Tuberculosis, and Viral Pneumonia classification with 332 correct classifications. However, there were 65 wrongly classified samples that were identified as Viral Pneumonia, an indication of the challenge of identifying pneumonia subtypes due to radiographic overlap.

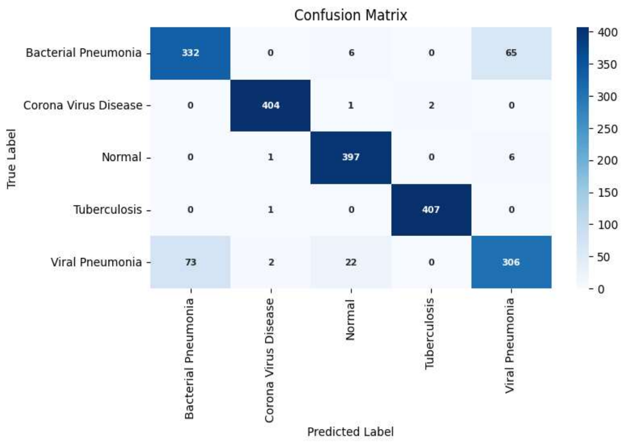


Figure5. Confusion matrix of Ensemble Bootstrap Sampling “Bagging” for EfficientNetV2M model

Figure 6. Shows the InceptionV3 model, Ensemble Bootstrap Sampling "Bagging," accurately identifies five classes of lung diseases from chest X-ray images: Bacterial Pneumonia, Corona Virus Disease, Normal, Tuberculosis, and Viral Pneumonia. There are few misclassifications performed in the class of Viral Pneumonia, suggesting that the model is struggling to separate healthy from bacterial lung images. Despite that, the good performance of the model supports ensemble-based systems for detecting lung diseases.

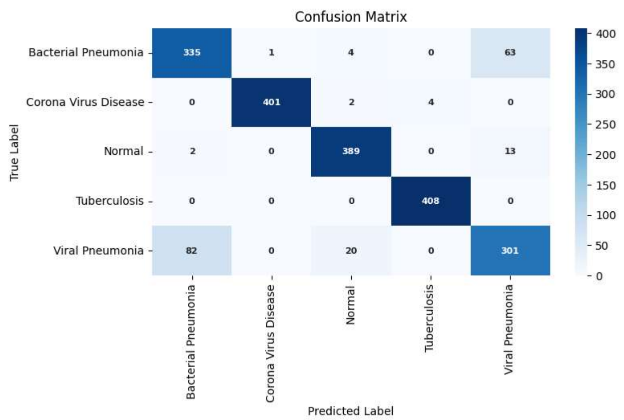


Figure6.Confusion matrix of Ensemble Bootstrap Sampling “Bagging” for InceptionV3 model

Figure7. Depicts The ResNet50 model, employing the Ensemble Bootstrap Sampling "Bagging" method, accurately distinguishes between cases of lung disease into the five categories: Bacterial Pneumonia, Corona Virus Disease, Normal, Tuberculosis, and Viral Pneumonia. It doesn't distinguish, however, between Viral Pneumonia and Bacterial Pneumonia since these have similar radiographic appearances. Such a restriction aside, the bagging method enhances model robustness and maximizes performance across several disease classes.

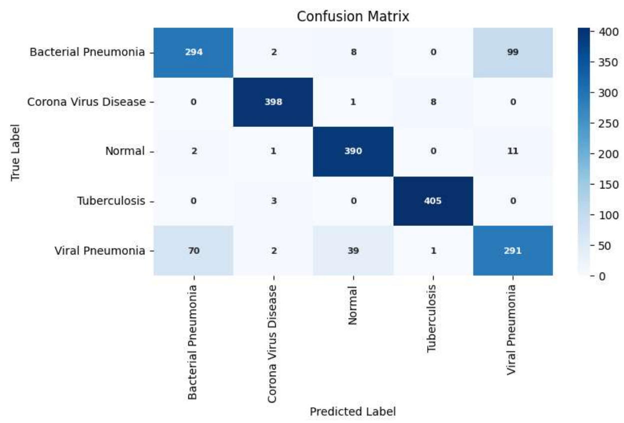


Figure7.Confusion matrix of Ensemble Bootstrap Sampling “Bagging” for ResNet50 model

Averaged ensembling voting approach, or averaged ensembling, is an approach that takes averages of prediction probabilities from different models to make the final prediction. Comparison of performance between different combinations of ensemble of deep learning models for classification of lung disease is reported in tabular form. The best performing combination was EfficientNetV2-M and InceptionV3 with 93% accuracy, 92.5% precision, 93% recall, and 92.5% F1-score. The ensemble of three models i.e., InceptionV3, EfficientNetV2-M, and ResNet50 gave respectable best with 92.14% accuracy and almost similar results for precision and recall. The ensemble of EfficientNetV2-M and ResNet50 gave the accuracy of 92.09% with a slight decline in precision and F1-score. InceptionV3-ResNet50 had the poorest metrics among the ensembles that were run with 91.75% accuracy, precision, and recall. EfficientNetV2-M removal caused a sharp decline in efficiency of classification. as evident from table 4.

Table 4. Averaged Ensemble Model results

Models	Accuracy	Precision	Recall	F1 Score
EfficientNetV2-M + InceptionV3	**93**	**92.5**	**93**	**92.5**
InceptionV3+ EfficientNetV2-M+ ResNet50	**92.14**	**92.09**	**92.14**	**92.09**
EfficientNetV2-M+ ResNet50	**92.09**	**92.05**	**92.09**	**92.04**
InceptionV3+ ResNet50	**91.75**	**91.75**	**91.75**	**91.71**

Figure8. Presents an ensemble model involving two pre-trained convolutional neural networks, EfficientNetV2-M and InceptionV3, for lung disease classification from chest X-ray images indicates the highest classifying accuracy for all the five classes of diseases. Corona Virus Disease, Normal, and Tuberculosis are distinctly classified with highest feature representation and decision fusion.

Bacterial Pneumonia class, however, indicated good accuracy with 339 correct classifications due to overlapping radiographic features leading to misclassifications.

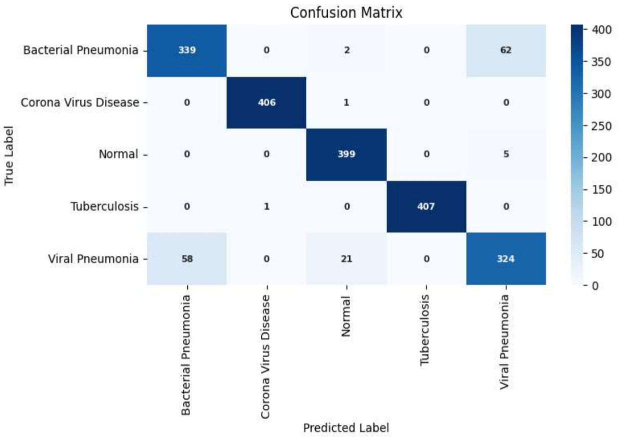


Figure8. Confusion matrix ensemble two pre-trained convolutional neural network models: InceptionV3, EfficientNetV2-M

Figure9. Represents a collection of three pre-trained convolutional neural networks (EfficientNetV2-M, InceptionV3, and ResNet50) for detection of lung diseases from chest X-rays to have enhanced overall performance in the accurate detection of Corona Virus Disease, Normal, and Tuberculosis. 73 cases of Bacterial Pneumonia are, however, labeled as Viral Pneumonia, causing ambiguity in both pneumonia conditions. Although all these problems, the ensemble approach reasonably integrates the strengths of individual models and verifies the performance of ensemble learning in improving the diagnosis accuracy of challenging medical image tasks.

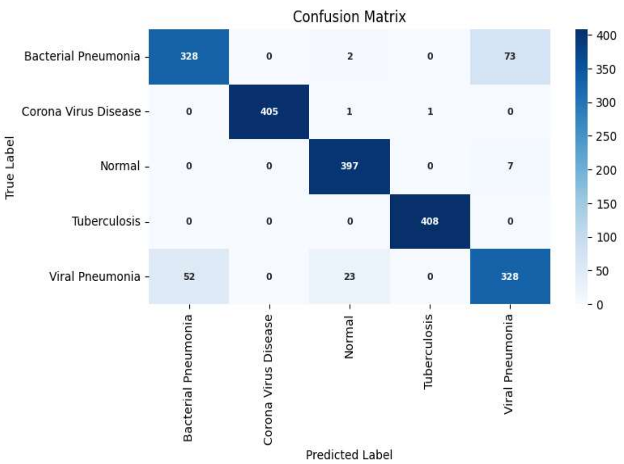


Figure9. Confusion matrix ensemble three pre-trained convolutional neural network models: InceptionV3, EfficientNetV2-M, and ResNet50

Figure10. Illustrates An ensemble model of EfficientNetV2-M and ResNet50 for lung disease classification from chest X-ray images illustrates better accuracy in Tuberculosis, Normal cases, and Corona Virus Disease diagnosis. Bacterial pneumonia was, however, misclassified in 325 cases, which suggests radiological feature overlap. The model utilizes both models' strength judiciously to enhance generalization and stable diagnosis, and would be a valuable addition to automated systems for lung disease diagnosis. This reflects the challenge of differentiating viral from bacterial infections and normal findings in medical imaging.

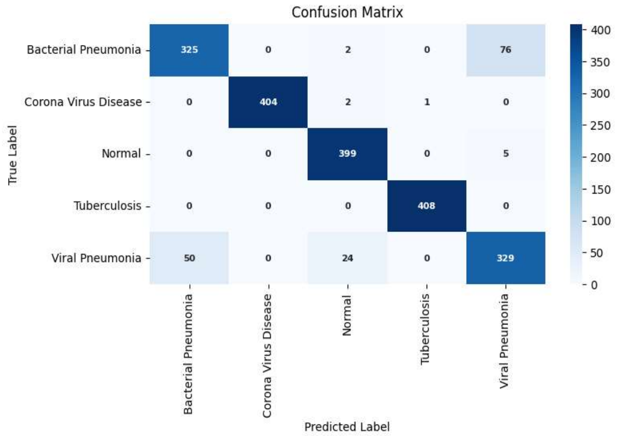


Figure10. Confusion matrix ensemble two pre-trained convolutional neural network models: EfficientNetV2-M, ResNet50

Figure11. Illustrates An ensemble model with InceptionV3 and ResNet50 for lung disease category classification from chest X-ray images showed high diagnostic performance with ideal Tuberculosis classification and close-to-ideal Corona Virus Disease and Normal classification. Bacterial Pneumonia was accurately classified in 322 cases but 79 were misclassified as Viral Pneumonia. Despite such misclassifications, the ensemble method effectively leverages InceptionV3's multi-scale feature extraction capacity and ResNet50's residual deep learning capacity, confirming its proficiency in improving computer-aided lung disease detection systems in clinical practice.

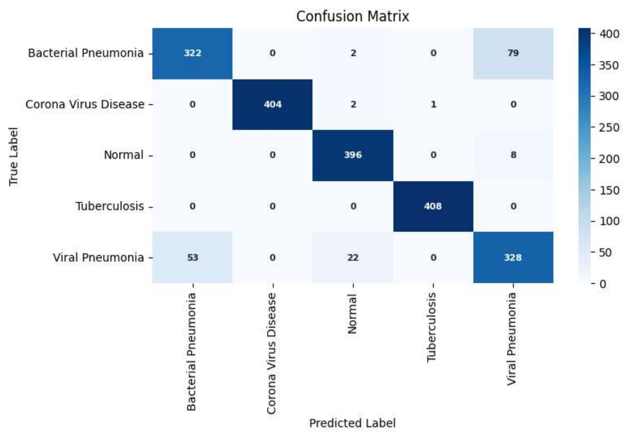
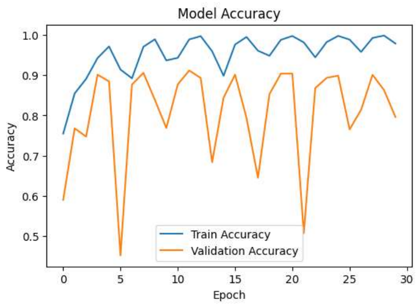


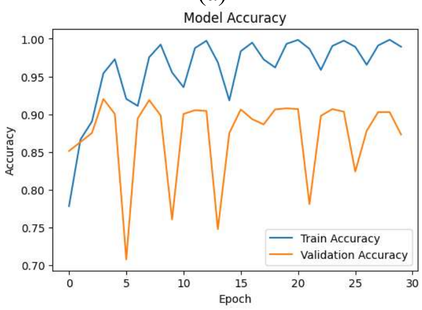
Figure11. Confusion matrix ensemble two pre-trained convolutional neural network models: InceptionV3, ResNet50

e-ISSN: 2614-8404
p-ISSN: 2776-3234

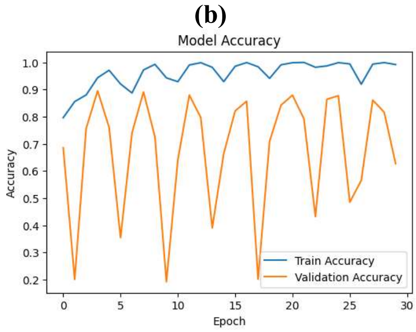
Figure12. Demonstrates the converging training curves of accuracy for the three deep learning models: InceptionV3, EfficientNetV2-M, and ResNet50 high capacity to learn, all converging towards 1.0 values. The validation accuracy is very distinct, especially in ResNet50. The performance of both InceptionV3 and EfficientNetV2-M is stable, with EfficientNetV2-M even more stable. The oscillations are caused by overfitting, data imbalance, or batch selection sensitivity of the model. The plot points state that further regularization or fine-tuning can be performed whereby validation accuracy will be more stable and generalization to new data will be better.



(a)



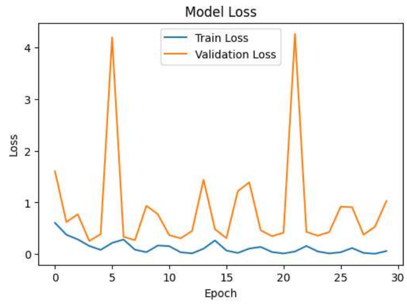
(b)



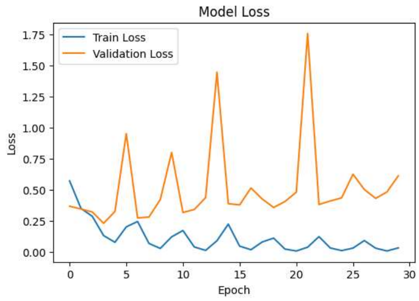
(c)

Figure12. Training Model Accuracy (a) InceptionV3 Model, (b) EfficientNetV2-M Model and (c) ResNet50 Model

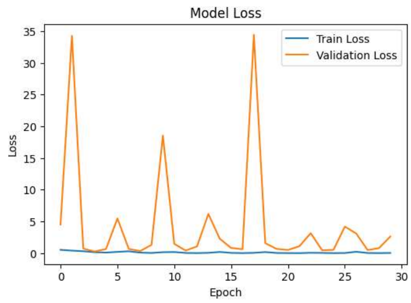
Figure13. Shows the study contrasted training loss and validation loss patterns of three deep models, namely InceptionV3, EfficientNetV2-M, and ResNet50. The networks all experienced low training loss, indicating good learning. Validation loss became unstable with sudden spikes to symbolize instability. InceptionV3 experienced spikes above 4.0, and EfficientNetV2-M had comparatively more regulated fluctuations. ResNet50 experienced the most instability, with spikes more than 30 for certain epochs. The spikes are indicative of possible overfitting, and this can result from bootstrap sampling sensitivity as well as validation set noise. EfficientNetV2-M experienced the most constant loss pattern.



(a)



(b)



(c)

Figure13. Training Model Loss (a) InceptionV3 Model, (b) EfficientNetV2-M Model and (c) ResNet50 Model

IV. COMPARATIVE ANALYSIS

Table 5. offers a comparative The comparison outlined by the table reflects the accuracy of the suggested ensemble methods against state-of-the-art models used for lung disease classification from images of chest X-ray. Previous work offered varying degrees of accuracy depending on the ensemble method, number of classes, and pre-trained models utilized. For example, CNN-LSTM ensemble with VGG16, MobileNet, DenseNet, and InceptionV3 had 89.31% accuracy on six disease classes (Nair & Singh, 2025), while entropy-based ensembles and stacking models of two-class classifications have 81.16% to 92% accuracy (Abad et al., 2024). Hybrid and weighted averages with different architectures such as DenseNet, ResNet, and Inception provided better results with the highest of 91.62% accuracy (Abad et al., 2024). On the contrary, the approach used in this research averaged EfficientNetV2-M and InceptionV3 average ensemble and bagging ensemble of EfficientNetV2-M to be better, particularly in the more difficult five-class case. The accuracy at its best with

e-ISSN: 2614-8404
p-ISSN: 2776-3234

averaged ensemble was 93%, while for bagging ensemble it was a 91%, and this is better than most two- and three-class models reported in the literature. These results exhibit the strength of the proposed ensemble approaches for learning dense feature representation and generalization, especially for multi-class lung disease detection from chest X-ray images.

V. CONCLUSION

The study compared two ensemble learning models to classify lung disease from chest X-rays. While the Averaged Ensemble Model employed the soft voting mean of three pre-trained models, the Bagging Ensemble Learning Model employed multiple replicates of the identical model trained over different data sets. The Averaged Ensemble approach was found to outperform due to its architecturally intricate architecture that allowed the models to learn complementary elements and identify different patterns in medical images. An ensemble learning model for accuracy and robustness improvement for lung disease identification was built holistically. The approach made use of the strength of three of the latest pre-trained CNN models EfficientNetV2M, InceptionV3, and ResNet50 coupled with Averaged Multi-Mdel ensemble (Soft Voting) and bagging ensembling strategies. Soft combination of EfficientNetV2M and InceptionV3 produced the best accuracy of 93.00%, supporting the need of the complementarity of models to ensure diagnostic reliability. The precision of future lung disease detection ensemble learning and clinical utility can be enhanced through the use of attention mechanisms, domain-specific augmentation techniques, synthetic data augmentation techniques, hybrid ensembling techniques, and explainable AI techniques.

e-ISSN: 2614-8404
p-ISSN: 2776-3234

Table 5. Comparison results with previous models

Ref.	Ensemble Method	Pretrained Model	Classes	Accuracy	Data Type
[13]	(CNN-LSTM)	CNN-LSTM, VGG16, MobileNet, DenseNet,and InceptionV3	6	89.31%	X-Ray
[14]	Entropy	ResNet50, DenseNet121, Inception,ResNet-v2	2	81.16	X-Ray
[15]	Stacking CNN Models	InceptionResNetV2	2	90.87	X-Ray
[16]	(Hybrid deep learning model)-NN	VGG16, VGG19, ResNet,EfficientNetB0	2	92%	X-Ray
[17]	CheXNet model weighted Averaging	DenseNet, InceptionV3, Xception, Resnet50	2	91.50	X-Ray
[18]	Weighted average ensembling	ResNet-18 and DenseNet-121	2	RSNA 86.85	X-Ray
[19]	weighted average ensembling	DenseNet201, Resnet50V2, Inceptionv3	2	91.62%	X-Ray
[20]	Stacking approach	Custom WRN (Wide Residual Network) VGG-16, InceptionV3, Xception,DenseNet-121, MobileNet-V2	3	90.97	X-Ray
Proposed Method	Averaged ensemble	**EfficientNetV2-M+ InceptionV3**	5	93	X-Ray
Proposed Method	Bagging	**Efficient**	5	91	X-Ray

e-ISSN: 2614-8404
p-ISSN: 2776-3234

REFERENCES

[1] M. H. Al-Sheikh, O. Al Dandan, A. S. Al-Shamayleh, H. A. Jalab, and R. W. Ibrahim, "Multi-class deep learning architecture for classifying lung diseases from chest X-Ray and CT images," *Sci Rep*, vol. 13, no. 1, p. 19373, Nov. 2023, doi: 10.1038/s41598-023-46147-3.

[2] K. Y. Win, N. Maneerat, S. Sreng, and K. Hamamoto, "Ensemble Deep Learning for the Detection of COVID-19 in Unbalanced Chest X-ray Dataset," *Applied Sciences*, vol. 11, no. 22, p. 10528, Nov. 2021, doi: 10.3390/app112210528.

[3] V. Ravi, V. Acharya, and M. Alazab, "A multichannel EfficientNet deep learning-based stacking ensemble approach for lung disease detection using chest X-ray images," *Cluster Comput*, vol. 26, no. 2, pp. 1181–1203, Apr. 2023, doi: 10.1007/s10586-022-03664-6.

[4] E. Kotei and R. Thirunavukarasu, "Ensemble Technique Coupled with Deep Transfer Learning Framework for Automatic Detection of Tuberculosis from Chest X-ray Radiographs," *Healthcare*, vol. 10, no. 11, p. 2335, Nov. 2022, doi: 10.3390/healthcare10112335.

[5] D. J. Rana and K. Rana, "SEMLCC: A Stacked Ensemble Model with Transfer Learning for High-Accuracy Lung Cancer Classification from CT Images," *Procedia Computer Science*, vol. 258, pp. 2584–2596, 2025, doi: 10.1016/j.procs.2025.04.520.

[6] Department of CSE, M.M. Engineering College, Maharishi Markandesh (Deemed to be University) Mullana, Ambala, India, K. Agrawal, R. Kumar, and S. Jain, "An Efficient Ensemble Model for Diagnosing Covid-19 and Pneumonia Using Chest X-Ray Images," *IJST*, vol. 15, no. 38, pp. 1900–1906, Oct. 2022, doi: 10.17485/IJST/v15i38.1269.

[7] Research Scholar, Instrumentation and Control Engineering, Gujarat Technological University, Gujarat, India, M. Patel, and M. Shah, "Deep Ensemble Learning Model for Diagnosis of Lung Diseases from Chest X -Ray Images," *IJST*, vol. 17, no. 8, pp. 702–712, Feb. 2024, doi: 10.17485/IJST/v17i8.3151.

[8] T. T. Ifty, S. A. Shafin, S. M. Shahriar, and T. Towhid, "Explainable Lung Disease Classification from Chest X-Ray Images Utilizing Deep Learning and XAI," Apr. 17, 2024, *arXiv*: arXiv:2404.11428. doi: 10.48550/arXiv.2404.11428.

[9] "ELREI: Ensemble Learning of ResNet, EfficientNet, and Inception-v3 for Lung Disease Classification based on Chest X-Ray Image," *IJIES*, vol. 16, no. 5, pp. 149–161, Oct. 2023, doi: 10.22266/ijies2023.1031.14.

[10] Y. H. Bhosale and K. S. Patnaik, "PulDi-COVID: Chronic obstructive pulmonary (lung) diseases with COVID-19 classification using ensemble deep convolutional neural network from chest X-ray images to minimize severity and mortality rates," *Biomedical Signal Processing and Control*, vol. 81, p. 104445, Mar. 2023, doi: 10.1016/j.bspc.2022.104445.

[11] N. K. Chowdhury, M. A. Kabir, M. M. Rahman, and N. Rezoana, "ECOVNet: An Ensemble of Deep Convolutional Neural Networks Based on EfficientNet to Detect COVID-19 From Chest X-rays," *PeerJ Computer Science*, vol. 7, p. e551, May 2021, doi: 10.7717/peerj-cs.551.

[12] S. Ali *et al.*, "A Soft Voting Ensemble-Based Model for the Early Prediction of Idiopathic Pulmonary Fibrosis (IPF) Disease Severity in Lungs Disease Patients," *Life*, vol. 11, no. 10, p. 1092, Oct. 2021, doi: 10.3390/life11101092.

[13] R. R. Nair and T. Singh, "Exploring Ensemble Architectures for Lung X-Ray Multi-Class Image Classification using CNN-LSTM," *Procedia Computer Science*, vol. 258, pp. 852–861, 2025, doi: 10.1016/j.procs.2025.04.317.

[14] M. Abad, J. Casas-Roma, and F. Prados, "Generalizable disease detection using model ensemble on chest X-ray images," *Sci Rep*, vol. 14, no. 1, p. 5890, Mar. 2024, doi: 10.1038/s41598-024-56171-6.

[15] N. H. Pham and G. S. Tran, "Apply a CNN-Based Ensemble Model to Chest-X Ray Image-Based Pneumonia Classification," *JAIT*, no. 11, pp. 1205–1214, 2024, doi: 10.12720/jait.15.11.1205-1214.

[16] M. Abdullah, F. B. Abrha, B. Kedir, and T. Tamirat Tagesse, "A Hybrid Deep Learning CNN model for COVID-19 detection from chest X-rays," *Heliyon*, vol. 10, no. 5, p. e26938, Mar. 2024, doi: 10.1016/j.heliyon.2024.e26938.

[17] L. Devnath *et al.*, "Deep Ensemble Learning for the Automatic Detection of Pneumoconiosis in Coal Worker's Chest X-ray Radiography," *JCM*, vol. 11, no. 18, p. 5342, Sep. 2022, doi: 10.3390/jcm11185342.

[18] 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, Sep. 2021, doi: 10.1371/journal.pone.0256630.

[19] A. K. Das, S. Ghosh, S. Thunder, R. Dutta, S. Agarwal, and A. Chakrabarti, "Automatic COVID-19 detection from X-ray images using ensemble learning with convolutional neural network," *Pattern Anal Applic*, vol. 24, no. 3, pp. 1111–1124, Aug. 2021, doi: 10.1007/s10044-021-00970-4.

[20] S. Rajaraman, S. Sornapudi, P. O. Alderson, L. R. Folio, and S. K. Antani, "Interpreting Deep Ensemble Learning through Radiologist Annotations for COVID-19 Detection in Chest Radiographs," Jul. 16, 2020. doi: 10.1101/2020.07.15.20154385.

[21] A. Tripathi, T. Singh, R. R. Nair, and P. Duraisamy, "Improving Early Detection and Classification of Lung Diseases With Innovative MobileNetV2 Framework," *IEEE Access*, vol. 12, pp. 116202–116217, 2024, doi: 10.1109/ACCESS.2024.3440577.

e-ISSN: 2614-8404
p-ISSN: 2776-3234

[22] H. A. Khater and S. A. Gamel, "Early diagnosis of respiratory system diseases (RSD) using deep convolutional neural networks," *J Ambient Intell Human Comput*, vol. 14, no. 9, pp. 12273–12283, Sep. 2023, doi: 10.1007/s12652-023-04659-w.

[23] M. Z. I. Noman, K. Sati, M. A. Yousuf, S. Aloteibi, and M. A. Moni, "LungCT-NET: An explainable transfer learning-based robust ensemble model for lung cancer diagnosis," *Knowledge-Based Systems*, vol. 324, p. 113854, Aug. 2025, doi: 10.1016/j.knosys.2025.113854.

[24] M. Thapa and R. Kaur, "An Explainable Deep – Learning based Multi-Label Image Classification for Chest X-Rays," *Procedia Computer Science*, vol. 258, pp. 2425–2434, 2025, doi: 10.1016/j.procs.2025.04.505.

[25] S. Rajaraman and S. K. Antani, "Modality-Specific Deep Learning Model Ensembles Toward Improving TB Detection in Chest Radiographs," *IEEE Access*, vol. 8, pp. 27318–27326, 2020, doi: 10.1109/ACCESS.2020.2971257.

[26] N. Ullah, M. Marzougui, I. Ahmad, and S. A. Chelloug, "DeepLungNet: An Effective DL-Based Approach for Lung Disease Classification Using CRIs," *Electronics*, vol. 12, no. 8, p. 1860, Apr. 2023, doi: 10.3390/electronics12081860.

[27] K. El Asnaoui and Y. Chawki, "Using X-ray images and deep learning for automated detection of coronavirus disease," *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 10, pp. 3615–3626, Jul. 2021, doi: 10.1080/07391102.2020.1767212.

[28] Md. N. Islam *et al.*, "Interpretable Differential Diagnosis of Non-COVID Viral Pneumonia, Lung Opacity and COVID-19 Using Tuned Transfer Learning and Explainable AI," *Healthcare*, vol. 11, no. 3, p. 410, Jan. 2023, doi: 10.3390/healthcare11030410.

[29] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks," *Pattern Anal Applic*, vol. 24, no. 3, pp. 1207–1220, Aug. 2021, doi: 10.1007/s10044-021-00984-y.