IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Multi Class Classification Of Respiratory Diseases Based On A Stacked Ensemble Of Alexnet And Inceptionv3 Using Chest X-Ray Images

Arumbaka Naveen [1] Dr.K. Jhansi Rani [2]

[1]M. Tech, Computer and communication, ECE, JNTU Kakinada, Andhra Pradesh, India

[2]Assistant Professor, Dept of ECE, JNTU KAKINADA, Andhra Pradesh, India

ABSTRACT

Ensemble learning is a machine learning technique that combines the predictions of multiple models to improve overall performance and robustness, often outperforming individual models by leveraging their collective strengths. This work presents a robust deep learning framework for multi-class classification of respiratory disease COVID-19, Pneumonia, Tuberculosis, and Normal from chest X-ray images using a novel stacking ensemble approach. A curated Combined Dataset comprising 4017 images was constructed from multiple public sources, addressing class imbalance through offline and real-time data augmentation techniques. Preprocessing steps included standardization to 224×224×3 RGB format and pixel normalization. The proposed model classification integrates two convolutional neural networks (CNNs) a custom-built AlexNet and a fine-tuned InceptionV3 to extract complementary features from medical images. Their softmax outputs are combined via a logistic regression-based stacking ensemble, leveraging the strengths of both models. Experimental results demonstrate that the stacking ensemble model significantly outperforms individual models, achieving 99.54% accuracy, precision, recall, and F1-score, compared to 93.25% for AlexNet and 83.67% for InceptionV3. The model's high performance highlights its potential for reliable and automated diagnosis of respiratory conditions in clinical settings.

Keywords: Chest X ray images, AlexNet, InceptionV3, Pneumonia, Tuberculosis, COVID-19, Stacking Ensemble, Deep Learning, Machine Learning.

1. INTRODUCTION

Respiratory illnesses like COVID-19, Pneumonia, and Tuberculosis continue to impact millions of people around the world, putting immense pressure on healthcare systems. Chest X-ray imaging plays a vital role in diagnosing these conditions due to its speed, availability, and affordability. However, reading and interpreting these images accurately requires trained radiologists, and even then, human error and fatigue can lead to inconsistent results. This has led researchers to explore deep learning-based solutions to support and enhance medical decision-making.

Deep learning, especially convolutional neural networks (CNNs), has shown remarkable progress in the field of medical imaging. Yet, building a reliable model for multi-class classification of respiratory diseases remains challenging due to factors like uneven class distributions, differences in image sizes, and subtle similarities between disease patterns. To overcome these issues, created the Combined Dataset by combining chest X-rays from various trusted public sources, representing four key classes: COVID-19, Pneumonia, Tuberculosis, and Normal. The dataset was carefully processed and augmented to ensure that each class was fairly represented and that models trained on it could generalize well to real-world cases.

AlexNet is a pioneering deep convolutional neural network (CNN) architecture introduced in 2012 by Alex Krizhevsky and colleagues. Designed for image classification tasks, it played a key role in advancing computer vision by winning the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). AlexNet consists of eight layers five convolutional layers followed by three fully connected layers incorporating techniques like ReLU activation for faster training, dropout to prevent overfitting, and data augmentation to enhance generalization. Its ability to extract hierarchical features from images, such as edges and textures, makes it effective for medical imaging tasks like classifying chest X-rays. AlexNet's innovative design laid the foundation for modern deep learning models, offering a balance of depth and computational efficiency suitable for automated diagnostic systems.

InceptionV3 is an advanced convolutional neural network (CNN) architecture developed by Google in 2015, designed for efficient and high-performing image classification. It builds on the Inception family, emphasizing computational efficiency through factorized convolutions and smaller filter sizes to reduce parameters while maintaining deep feature extraction. The model features "Inception modules" that process input data at multiple scales concurrently, capturing diverse patterns like fine details and broader structures in images. With 42 layers, InceptionV3 leverages techniques like batch normalization, label smoothing, and auxiliary classifiers to enhance training stability and accuracy. Its ability to extract complex features makes it well-suited for medical imaging tasks, such as classifying chest X-rays for respiratory disease detection, offering robust performance in automated diagnostic applications.

Stacking ensemble, also known as stacked generalization, is a machine learning technique that enhances predictive performance by combining the outputs of multiple base models using a meta-learner. Unlike simple ensemble methods that average or vote on predictions, stacking trains a secondary model, such as logistic regression, to learn how to optimally blend the outputs of diverse base models, like convolutional neural networks. This approach leverages the unique strengths of each model, capturing complementary patterns in data, such as varied features in medical images. By integrating predictions in a structured manner, stacking improves accuracy and robustness, making it highly effective for complex tasks like classifying respiratory diseases from chest X-rays in diagnostic systems.

The paper is structured as follows: Section 2 provides a literature review of related studies. Section 3 describes the methodologies employed. Section 4 discusses the experiments conducted and their findings. Finally, Section 5 offers the conclusion of the work.

2.RELATED STUDY

Charan et al. developed a transfer learning approach for classifying tuberculosis, COVID-19, pneumonia, and normal cases using 4,028 chest X-ray images from diverse public archives. Images were preprocessed by resizing to 224×224 pixels, converting to RGB, and augmenting with zooming, rotations, and flips. Local Binary Patterns extracted textural features, integrated with pre-trained CNNs (MobileNetV2, VGG16, InceptionNet, ResNet50, EfficientNet). Models used frozen layers, global average pooling, and softmax output with Adam optimizer. ResNet50 achieved 97.1% accuracy; EfficientNet reached 96.3% with textural features. On unseen data, ResNet50 scored 93% accuracy. Integrating textural features enhances diagnostic accuracy, but larger datasets are needed [1].

A. Naveen and K. Jhansi Rani (2025) presented a survey on ensemble learning techniques for medical image classification. They highlighted the limitations of individual CNNs in handling overfitting, noise, and class imbalance. The survey reviewed ensemble methods such as bagging, boosting, and stacking, focusing on their application to chest X-ray images. It also compared performance in multi-class disease detection, including

COVID-19, pneumonia, and tuberculosis. The authors conclude that stacking offers superior accuracy and robustness, though challenges of interpretability and computational cost remain [2].

Chokchaithanakul et al. explored adaptive preprocessing and augmentation for tuberculosis screening on chest X-ray datasets, addressing out-of-domain data challenges. The study compiled 6,168 radiographs from five sources, using lung BCET and augmentations like CLAHE, optimized with ResNet50 and EfficientNetB0 via transfer learning. Performance, evaluated with AUC and Grad-CAM, showed lung BCET achieving 0.9253 (in-domain), 0.6240 (BT), and 0.7978 (Maesot), with combined augmentations reaching 0.9446. The conclusion highlights improved robustness, though segmentation accuracy and dataset diversity pose limits. [3].

Berliana and Bustamam investigated stacking ensemble learning for classifying COVID-19 using chest X-ray and CT scan images. The study employed a dataset of 1,140 chest X-ray images and 2,400 CT images, balanced between COVID-19 and non-COVID-19 cases. The methodology utilized a twolevel stacking ensemble model, with Support Vector Classification (SVC), Random Forest (RF), and K-Nearest Neighbors (KNN) as base learners at the first level and SVC as the meta-learner at the second level. The model achieved 99% accuracy, 98% precision, and 100% recall for X-ray images, and 97% accuracy, 97% precision, and 97% recall for CT images. The conclusion highlights that the stacking ensemble learning model, leveraging diverse base learners, outperforms individual models, enhancing rapid and accurate COVID-19 diagnosis from medical imaging [4].

Abimannan et al. explored the design of ensemble-based deep learning systems that integrate multiple feature types to enhance model performance across complex tasks. Their approach combined different deep learning models—such as CNNs and RNNs—with ensemble techniques like bagging, boosting, and stacking. These systems processed multimodal inputs, including text, images, and audio, and used fusion strategies like concatenation and attention mechanisms to merge features before final decision-making. The architecture was applied in various domains, including image recognition, object detection, medical image analysis, speech recognition, and natural language tasks. Key challenges such as interpretability, computational load, and privacy concerns were addressed through strategies like optimized ensemble combinations and federated learning. Overall, the findings showed that combining diverse models and features leads to better generalization and more robust performance compared to relying on single models, especially in data-rich, multimodal environments [5].

Dhar introduced a multistage ensemble learning model (MSEN) with weighted voting and genetic algorithm (GA) optimization for early detection of Chronic Obstructive Pulmonary Disease (COPD) using the Exasens dataset. The methodology created two pools of four classifiers each: Pool 1 (XGB, ET, RF, GB) and Pool 2 (LR, SVC, NuSVC, KNN). Each pool generated a weighted ensemble model via a weighted voting strategy, and these were combined to form the MSEN model. GA optimized classifier hyperparameters, while grid search tuned weights for classifiers and ensembles. Evaluated on 239 samples (160 healthy, 79 COPD), the MSEN model outperformed individual and genetically optimized machine learning models. The conclusion states that MSEN achieves superior performance (accuracy 98.20%, precision 98.00%, recall 96.00%, F1-measure 96.67%, AUC 99.12%) compared to benchmarks like XGBoost (92.05%), offering a reliable tool for COPD detection, though its complexity and training time [6].

Rahman et al. developed a deep learning-based approach for detecting tuberculosis (TB) using chest X-ray images, focusing on lung segmentation and visualization to improve diagnostic reliability. The methodology utilized two U-Net variants (original and modified) for lung segmentation, followed by TB classification using nine pre-trained convolutional neural networks, including CheXNet, DenseNet201, and ResNet. The dataset comprised 3500 normal and 3500 TB chest X-ray images from NLM, Belarus, NIAID, and RSNA datasets. The original U-Net outperformed the modified variant, and segmented lung images were classified alongside non-segmented ones. Score-CAM visualization confirmed that decisionmaking focused on lung regions in segmented images, while t-SNE highlighted feature distinctions. The conclusion underscores that segmentation markedly enhances classification accuracy (98.6 with DenseNet201 on segmented images versus 96.47% with CheXNet on non-segmented), offering a robust CAD tool for TB screening, though misclassifications suggest potential for patch-based analysis [7].

Ali et al. developed an ensemble deep learning model for lung segmentation in chest radiographs to support computer-aided diagnosis for diseases like tuberculosis and COVID-19. The methodology utilized the Shenzhen Hospital dataset, comprising 566 chest radiographs with manually segmented lung masks. The model extended DeepLabV3+, using backbones including ResNet18, ResNet50, Mobilenetv2, Xception, and InceptionResnetV2, with adjusted atrous spatial pyramid pooling dilation rates (4, 8, 16). The ensemble combined Xception and InceptionResnetV2 outputs, averaged into a probability matrix for final segmentation. Performance was evaluated using intersection-over-union (IoU) and accuracy metrics. The conclusion demonstrates that the ensemble model achieves high segmentation accuracy (IoU: 0.97, global accuracy: 98.91) on the test set, outperforming single-backbone models and benchmarks, offering robust lung segmentation for CAD systems, though limited dataset size suggests further validation [8].

Xu and Yuan developed a convolutional neural network model with coordinate attention (CoordAttention) for automated pulmonary tuberculosis (TB) detection in chest X-ray images to enhance diagnostic efficiency. The methodology utilized the Shenzhen dataset, comprising 662 chest X-rays (326 normal, 336 TB), and employed the VGG16-CoordAttention model, integrating CoordAttention into VGG16 with transfer learning and five-fold cross-validation. The conclusion shows the model achieves high performance (accuracy: 92.73 %, AUC: 97.71%, precision: 92.73%, recall: 92.83%, F1-score: 92.82%), outperforming methods like ConvNet and B-CNN, offering robust, lightweight TB detection, though further validation is needed [9].

Elhanashi et al. investigated deep learning techniques for classifying and localizing multiple abnormalities, including COVID-19, on chest X-ray images to enhance diagnostic accuracy. The methodology utilized two datasets: the COVID-19 Radiography Database (21,165 images) for multi-classification and the SIIM-FISABIO-RSNA dataset (6,334 images) for object detection. Multi-classification employed pre-trained convolutional neural networks (VGG16, VGG19, ResNet50, Xception) to categorize images into normal, COVID-19, lung opacity, and viral pneumonia. Object detection combined EfficientNet, YOLOv7, and Faster R-CNN using weighted box fusion to localize abnormalities with bounding boxes. Performance was evaluated using accuracy, precision, recall, F1-score, and mean average precision (mAP). The conclusion indicates that VGG16 achieves the highest multi-classification accuracy (92.1 %) with an F1-score of 0.92 for COVID-19, while the ensemble model (EfficientNet+YOLOv7+Faster R-CNN) attains an mAP of 0.612, outperforming single models. The approach enhances diagnostic efficiency, though class imbalance and computational complexity remain challenges for future research [10].

Militante et al. developed a deep learning approach for detecting COVID-19 and pneumonia using chest X-ray images to assist radiologists in diagnosing respiratory diseases. The methodology employed a VGG-16 convolutional neural network model with six convolutional layers, batch normalization, maxpooling, and dense layers, using a dataset of 15,798 chest X-ray images comprising normal, bacterial pneumonia, viral pneumonia, and COVID-19 cases from public sources. The conclusion states that the VGG-16 model achieves a 95% accuracy in detecting COVID-19, bacterial, and viral pneumonia, offering significant support for automated medical diagnosis [11].

Jibril and Adeshina developed a deep learning framework for detecting COVID-19 in chest X-ray images, leveraging convolutional neural networks enhanced by generative adversarial networks (GANs) to address limited data availability. The methodology utilized a dataset of 4025 postanterior X-ray images (700 COVID-19, 1345 viral pneumonia, 1341 normal) from a public repository, with GANs generating synthetic COVID-19 images to balance the dataset, resulting in 1339 COVID-19 samples. A pre-trained VGG-19 model with frozen convolutional layers and additional dense layers extracted features and classified images. The conclusion highlights that the framework achieves a 91.3% accuracy, outperforming other GAN-enhanced models, and supports radiologists in low-income regions like Nigeria [12]

Nefoussi et al. investigated deep learning for detecting COVID-19 in chest X-ray images, emphasizing the role of radiological imaging in screening during the global health crisis. The methodology involved combining three public and one private dataset with 518 COVID-19 samples and training eight pretrained convolutional neural networks, including VGG16, Xception, and MobileNet, to classify images as normal, pneumonia, or COVID-19. The conclusion states that Xception achieves the best performance across accuracy (0.94), precision, recall, and F-score for COVID-19 detection, demonstrating improved generalization through diverse datasets despite challenges with limited data [13].

3.METHODOLOGY

The Combined Dataset integrates multiple interrelated datasets to offer a comprehensive collection of chest X-ray images. Ensemble Techniques are applied to this dataset to accurately classify various lung disorders, including COVID-19, Pneumonia, Tuberculosis, and Normal.

This illustration depicts a suggested Flowchart for a machine learning model that begins with a dataset that is preprocessed and augmented to increase quality and diversity in the data. This preprocessed data has completed preprocessing and is separated into AlexNet and InceptionV3 neural networks for training and validation. The predicting models from each neural network are fused together in ensemble predictions. This ensemble prediction would then be used in a stacking ensemble process to further refine the predictions and predictions are evaluated in a performance evaluations stage. The Flowchart is designed to take advantage of both AlexNet and InceptionV3 and to merge the predictions from both networks through ensemble methods for a better performing model overall.

3.1 DATASET CREATION AND INTEGRATION

A comprehensive collection of chest X-ray images created to support the multi-class classification of respiratory diseases COVID-19, pneumonia, tuberculosis, and normal makes up the Combined Dataset created for this work. The complete dataset creation process, preprocessing pipeline, and data augmentation methods used to address class imbalance and strengthen models are presented in this section. To guarantee reproducibility and facilitate further research in automated medical image processing, the dataset is made publicly available at

[https://www.kaggle.com/datasets/arumbakanaveen/combined-dataset].

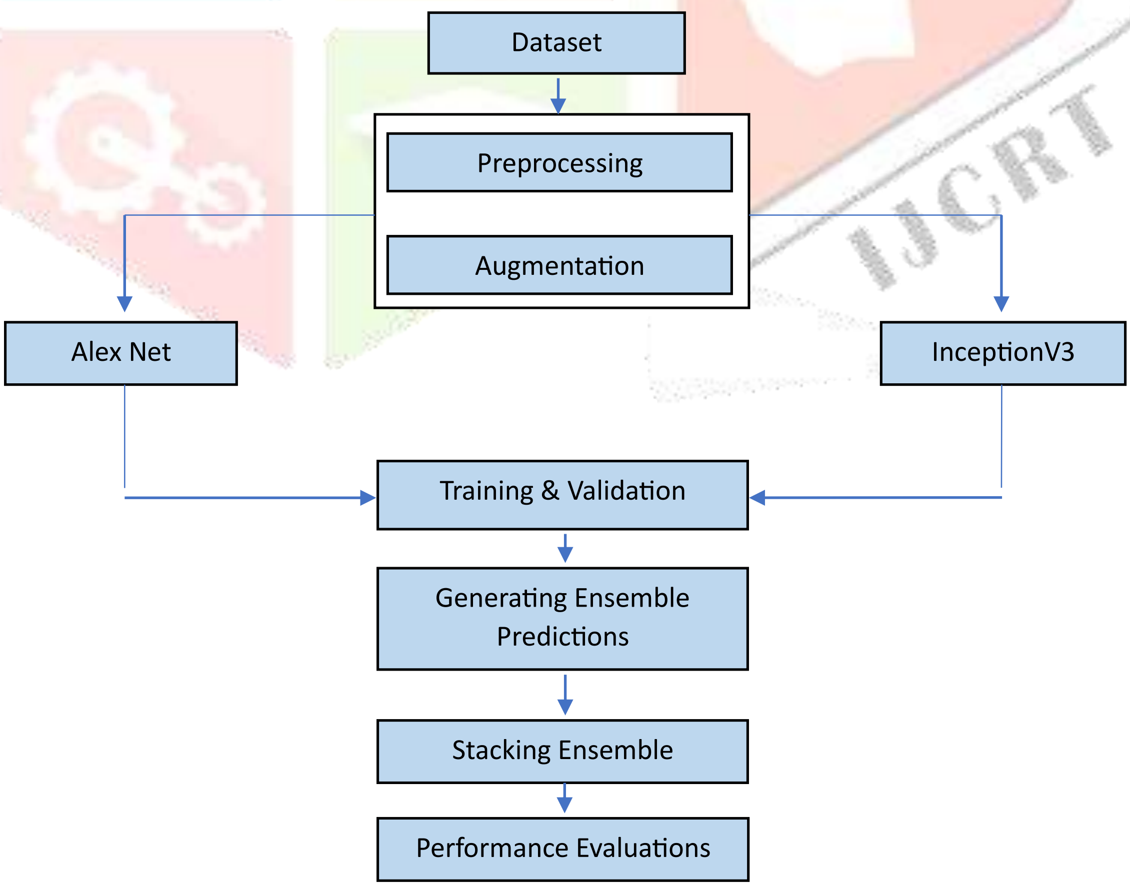

Figure 1: Flowchart For the Proposed Model

3.1.1 COMBINED DATASET

The Combined Dataset was curated by aggregating chest X-ray images from multiple publicly available sources, addressing the challenge of dataset heterogeneity in medical imaging. The dataset comprises four classes, each representing distinct clinical conditions:

COVID-19: 1009 CT grayscale images of size 2000×2000×1 is downloaded from the public repositories like COVID-19 Radiography Database [14]. These images depict typical radiological features of COVID-19 including ground-glass opacities and consolidation that had been confirmed on clinical diagnosis. Normal: 1017 grayscale images of size 224×224×1, collected from datasets such as ChestX-ray8 [15]. These are the images of normal lungs without any visible pathological issues, this is a category serving as a healthy control. Pneumonia: 1010 grayscale images of size 712includes bacterial pneumonia as well as viral pneumonia cases showing symptoms of lobar consolidation and interstitial patterns. Tuberculosis: 700 grayscale images of size 512×512×1, originally collected from repositories such as the TBX11 dataset [16]. The counting after off line data augmentation is increased to 981 images to mitigate class imbalance.

The total dataset size is 3736 images before augmentation, as summarized in Table I. The Combined Dataset balances class representation while capturing diverse radiological patterns, making it suitable for training robust deep learning models as shown in table 1.

Table 1: Combined Dataset Composition

CLASS	INITIAL IMAGES	DIMENSIONS
Covid-19	1009	2000×2000×1
Normal	1017	224×224×1
Pneumonia	1010	712×439×1
Tuberculosis	700	512×512×1
Total	3736	--------------

3.2 PREPROCESSING

Preprocessing refers to the initial steps taken to prepare raw data for analysis or modelling, ensuring it is clean, structured, and suitable for use in machine learning or other computational tasks. It involves transforming data to remove inconsistencies, handle missing values, and enhance its quality for better algorithm performance. Common techniques include normalizing data to a standard scale, removing outliers to reduce noise, balancing class distributions to avoid bias, and augmenting data to increase dataset size or diversity. For example, in image analysis, preprocessing might involve resizing images or adjusting contrast, while in numerical data, it could mean filling in missing entries or standardizing values. These steps are crucial to improve model accuracy and efficiency by making data more consistent and relevant to the task.

To prepare the Combined Dataset for deep learning-based classification, a preprocessing pipeline was applied to all images to address variations in tripod height and pixel intensity. Images were uniformly resized to 224x224x3 dimensions to ensure consistency. For grayscale images with a single channel, the grayscale data was duplicated across all three RGB channels to align with the standard three-channel input required by most convolutional neural network models. Pixel intensities were normalized to a [0, 1] range by dividing by 255, stabilizing numerical computations during training, enhancing convergence speed, and ensuring uniform feature scaling across the dataset.

3.3 DATA AUGMENTATION

Data augmentation is a fundamental technique in deep learning aimed at increasing the diversity of training data by generating modified versions of existing samples. This process plays a crucial role in improving the generalization ability of neural networks, especially when dealing with limited or imbalanced datasets. Common augmentation techniques include geometric transformations such as rotation, scaling, and translation, as well as adjustments to brightness and contrast. These operations maintain the semantic content of the image while providing variations that help the model become more robust to real-world variability.

In image classification and medical recognition tasks, data augmentation simulates natural variations that may occur due to changes in viewing angle, illumination, or scale. By incorporating such variations into the training data, models are better equipped to handle test data that deviates from the original distribution. This is particularly valuable in scenarios where data collection is constrained or where certain classes are underrepresented. Augmentation also reduces overfitting by limiting the model's ability to memorize specific image patterns, forcing it to learn more general features.

Two types of augmentation strategies were implemented: off-line and real-time. In off-line augmentation, the dataset is expanded prior to training by applying fixed transformations. For example, in cases where a class contains fewer samples than others, synthetic images can be generated using scaling, shifting, and brightness alterations to help balance the dataset. This approach ensures that all classes contribute equally to the learning process, thus improving model fairness and consistency during training.

Real-time augmentation, on the other hand, applies random transformations to images during training. Libraries such as TensorFlow's ImageDataGenerator offer parameters including rotation range, width and height shifts, shear transformations, zoom, and flips. These augmentations are applied dynamically in each training epoch, creating different versions of the same input image on-the-fly. This technique enhances the model's robustness and adaptability, especially in domains like medical imaging, where models must generalize across diverse patient populations, imaging conditions, and equipment settings.

As shown in Table 2, the dataset consists of four classes with varying image sizes initially, all resized to a uniform dimension of 224×224×3, and only the Tuberculosis class underwent image augmentation.

Table 2: Dataset Characteristics Before and After Preprocessing and Augmentation

CLASS	BEFORE PREPROCESSING SIZE	AFTER PREPROCESSING SIZE	IMAGES BEFORE AUGMENTATION	IMAGES AFTER AUGMENTATION
Covid-19	2000×2000×1	224×224×3	1009	1009
Normal	224×224×1	224×224×3	1017	1017
Pneumonia	712×439×1	224×224×3	1010	1010
Tuberculosis	512×512×1	224×224×3	700	981

3.4. PROPOSED MODEL

The proposed model implements an efficient image classification model that is built via stacking an ensemble of convolutional neural networks (CNNs), AlexNet and InceptionV3, thereby increasing prediction performance.

The proposed models must also have class weights and training parameters that account for biases that may exist in the data. AlexNet was developed from scratch to extract important image features, whereas InceptionV3 was developed using weights compared to ImageNet, to take advantages of feature representation that were learned previously. The proposed model was build using a logistic regression method that takes the output of both models, thus creating a stacking ensemble, as a method that both models can jointly perform classification. The proposed model includes a method of training that uses a custom learning rate schedule and evaluation method. The proposed model as shown in figure 2.

3.4.1. ALEXNET

The AlexNet model was created for the classification problem. AlexNet works on RGB images of 224x22 4x3 size and these are normalized to the interval [0,1]. The model begins with an input layer called 'image_input' on which the image is passed directly to the input of the model. The first convolutional layer (conv1) applies 96 filters of shape 11x11 with a stride of 4, valid padding, ReLU activation, and L2 weight decay regularization with a 0.005 weight decay. The output of conv1 is 54x54x96. After this layer, the model applies a BatchNormalization layer to help stabilize the training. The model applies a MaxPooling layer with a size of 3x3 and a stride of 2. After this layer, the output size is reduced to 26x26x96. The second convolution layer (conv2) uses 256 filters (5x5, same padding, ReLU activation and L2 weight decay). The output of conv2 is 26x26x256. Like before, a BatchNormalization layer followed by a MaxPooling layer with a size of 3x3 and a stride of 2, this time produce an output size of 12x12x256.

The third and fourth convolution layers (conv3 and conv4) use 384 filters each (3x3, same padding, ReLU activation and L2 weight decay). After each of conv3 and conv4, the output remains of size 12x12x384. Next, the fifth convolution layer (conv5). The fifth convolution layer uses 256 filters (3x3, same padding, ReLU activation and L2 weight decay), the output of conv5 is 12x12x256. The model then applies a BatchNormalization and a MaxPooling with 3x3 and a stride of 2, the output is reduced to 5x5x256 for the next operation. The output of conv5 is sent through a GlobalAveragePooling2D layer (see Section 2.4.4 - Global averages pooling) generates a 256-dimensional vector. The output of the GlobalAveragePooling2D layer is then connected to a Dropout layer (0.5 rate) to help minimize overfitting. Next is a Dense layer; this layer uses a softmax activation with L2 weight decay regularly, and outputs the log probabilities for the number of classes.

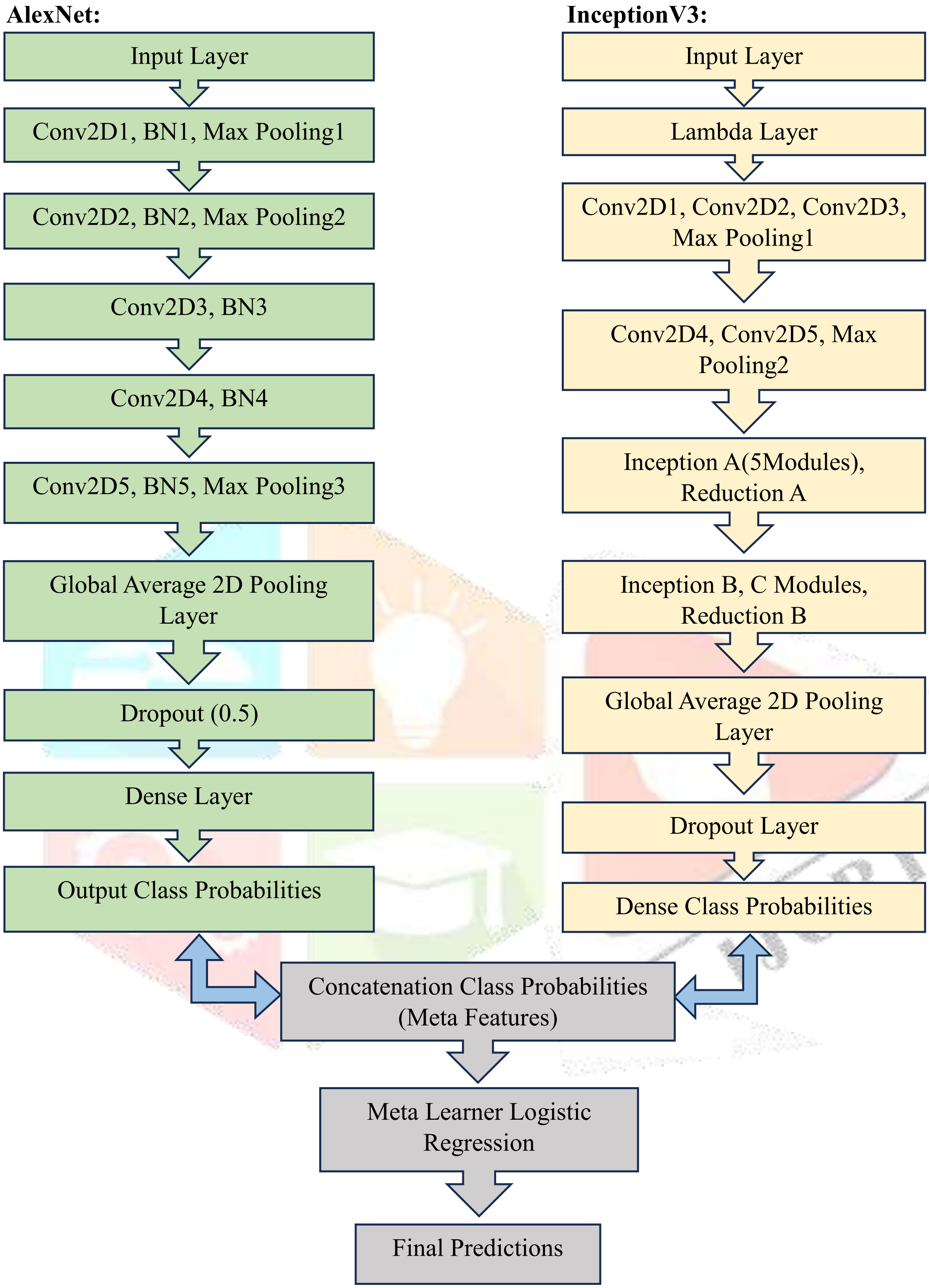

Figure 2: Proposed Model

3.4.2 INCEPTIONV3

The InceptionV3 model modified to solve the problem used ImageNet trained weights, with all the layers trainable, and will process RGB 224x224x3 images in the [0,1] range. The input layer, called 'image_input' connects to a Lambda layer that utilized the preprocessing function which is specific to InceptionV3 and converts the pixel values to [-1,1] values, consistent with the range of the ImageNet dataset while not changing the dimensions to 224x224x3 images. The InceptionV3 base model excluding the top dense layer, consists of several Inception modules. The Inception module performs multiple convolutional and pooling operations in parallel, as well as a concatenation operation, which enables the extraction of features at multiple scales and convenes the output of the Inception module to a final feature map of 5x5x2048.

These modules further include an auxiliary classifier to support learning of the features. Next, a GlobalAveragePooling2D layer, called 'avg_pool,' converts the output into a 2048-dimensional vector. A Dropout layer (0.5 dropped units, called 'dropout') is included as a strategy to help mitigate overfitting. The final layer is called output_dense which is a Dense layer that corresponds to the number of classes and a softmax activation with L2 regularization (0.005). This layer outputs the class probabilities which are learned from the combination of the pre-learned features and the specific to the task features and weights, to enhance classification.

3.4.3 STACKING ENSEMBLE METHOD

The stacking ensemble makes use of AlexNet and InceptionV3 predictive capabilities and combines them to improve classification performance. After training, each model produces probability output for the training and validation sets, which were collected in batches by using the respective data generators. For the validation dataset, the probability outputs were put into a feature matrix, using the concatenated outputs from AlexNet and InceptionV3, where each sample included the probability vectors from both models

The logistic regression method is constructed using Scikit-learn logical regression function with balanced class weights, the 'liblinear' solver option, a maximum of 1000 iterations, and a random state of 42 for consistency, will then be trained based on the concatenated validation feature matrix to predict what the true class labels are. A similar concatenated feature matrix will be created for the training set, and the produced logistic regression method predictions will also predict class labels for the training dataset. The stacking approach exploits Alex Net’s performance of feature extraction, while InceptionV3 predetermines the hierarchical features previously displayed pre-trained in InceptionV3, thereby enhancing overall classification performance through how to optimally combine class label predictions.

4.PERFORMANCE EVOLUTIONS

To assess the performance of the classifiers, we used four commonly used metrics: accuracy, precision, recall, and F1-score. Together, these metrics represent the classification performance of the models, which is especially important in multi-class medical image classification applications, where a misclassification error can have serious clinical implications. The formulas for the metrics are as follows:

I. ACCURACY

Accuracy: accuracy is a metric that assesses the proportion of correct predictions based on the total number of predictions by a model which mirrors how often the output from the model is equal to the actual outcome. (Measures the overall correctness of predictions).

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

II. PRECISION

Precision is the proportion of correctly identified positive instances relative to the total number of instances predicted as positive. Indicates how many predicted positives are actually correct.

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

III. RECALL

Recall is the proportion of correctly predicted positive instances relative to the total number of actual positive instances. Reflects how well the model captures actual positives.

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

IV. F1-SCORE

The F1 score is a harmonic mean of precision and recall, providing a balanced reflection of both metrics, making it a preferable choice over relying solely on precision or recall, especially when the distribution of classes is uneven.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

4.1 RESULTS

The proposed deep learning-based classification framework was implemented and trained using Google Colab, leveraging a T4 GPU environment for accelerated computation. Model development and experimentation were conducted using TensorFlow, Keras, and Scikit-learn libraries. Initial model prototyping and data preparation were performed locally on an HP Ryzen 3 laptop, with final training performed in the cloud. The Combined Dataset containing chest X-ray images classified into four categories (COVID-19, Pneumonia, Tuberculosis, and Normal) was divided using an 80:20 split for training and validation, respectively.

The stacking ensemble approach, which combines the outputs of AlexNet and InceptionV3, demonstrated exceptional performance during training, achieving a high accuracy of 99.54% within 30 epochs. The validation results also reflected strong generalization capabilities, with an accuracy of 98.76%, along with equally impressive precision, recall, and F1-score values.

In comparison, the individual models showed relatively lower performance. AlexNet achieved an accuracy of 93.25%, with precision, recall, and F1-score recorded at 94.44%, 94.45%, and 94.15% respectively, indicating a well-balanced performance. On the other hand, InceptionV3, despite utilizing pre-trained ImageNet weights, yielded a lower accuracy of 83.67%, along with a precision of 85.36%, recall of 83.67%, and F1-score of 83.45%.

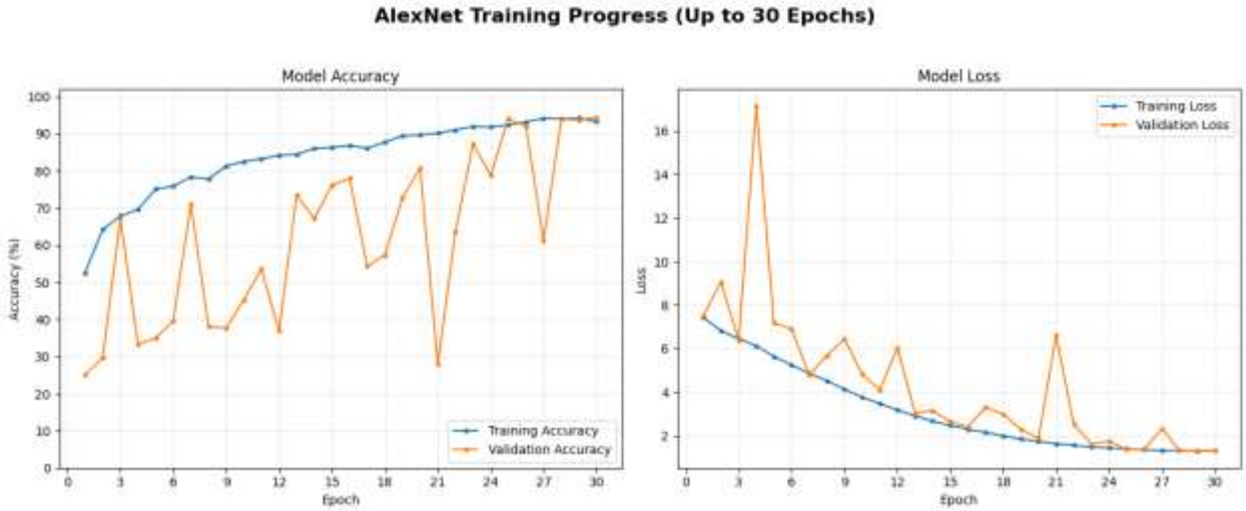

Figure 3: AlexNet Training Progress

The figure 3 illustrates the training and validation progress of the Individual AlexNet model, showing how its performance metrics likely accuracy and loss evolve over training epochs for the multi-class classification task. It provides a visual representation of AlexNet's learning behaviour on the Combined Dataset.

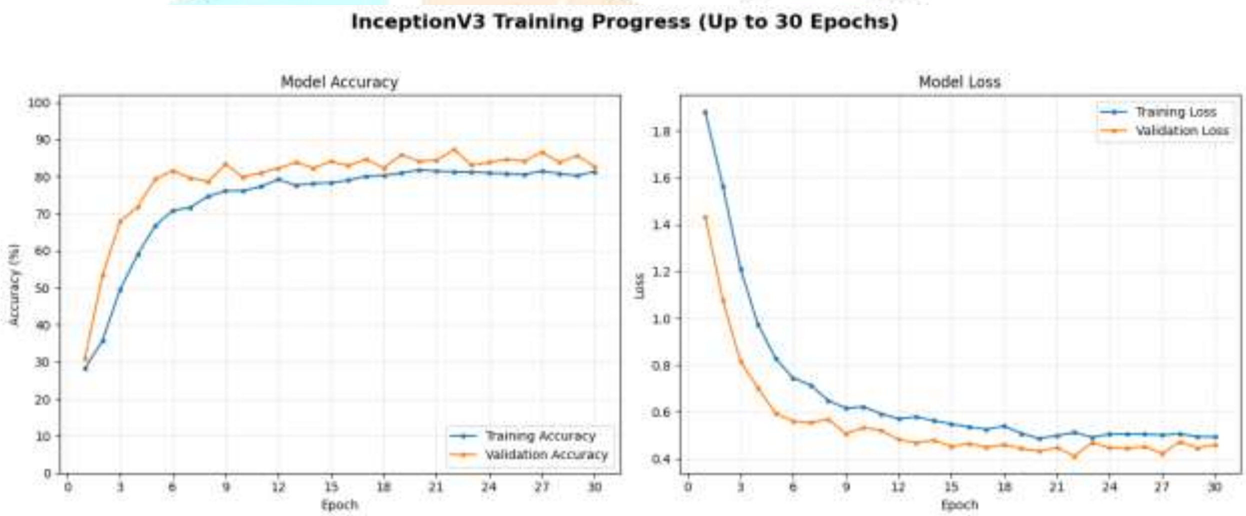

Figure 4: InceptionV3 Training Progress

The figure 4 depicts the training and validation progress of the Individual InceptionV3 model, showing metrics such as accuracy and loss across epochs. It reflects how InceptionV3 provides a visual representation of AlexNet's learning behaviour on the Combined Dataset.

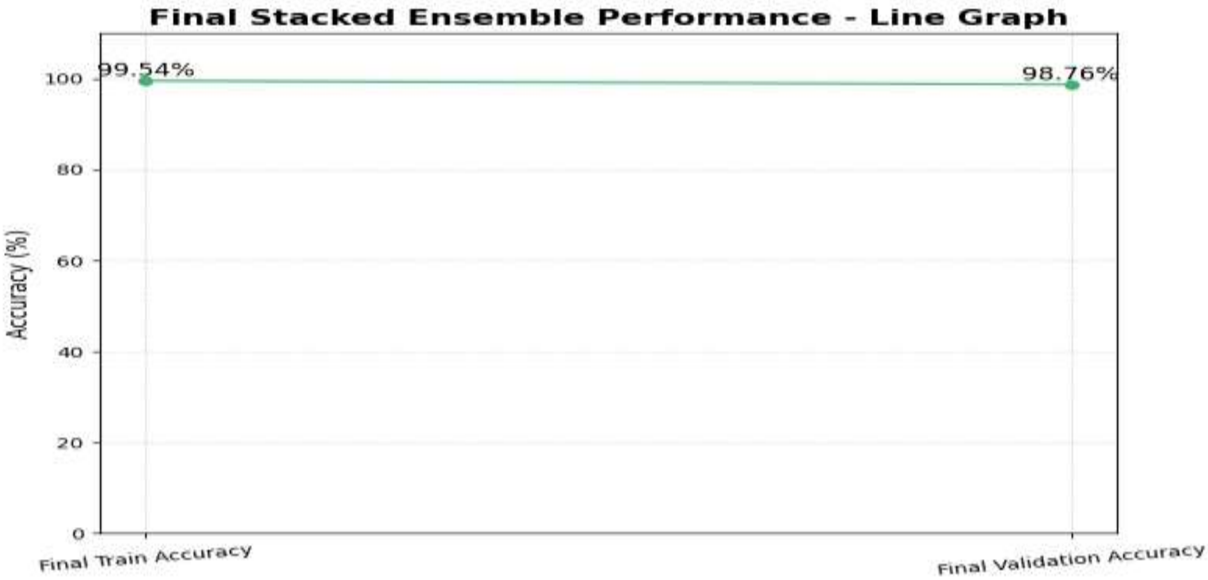

Figure 5: Proposed Model performance comparison

Figure 5 shows the final accuracy performance of the stacked ensemble model on both the training and validation dataset. The model achieved a training accuracy of 99.54% and a validation accuracy of 98.76%, indicating strong generalization ability with minimal overfitting. The near-horizontal line between the two points reflects consistent performance across combined datasets. These results confirm the effectiveness of the ensemble approach in accurately classifying respiratory classes with high reliability.

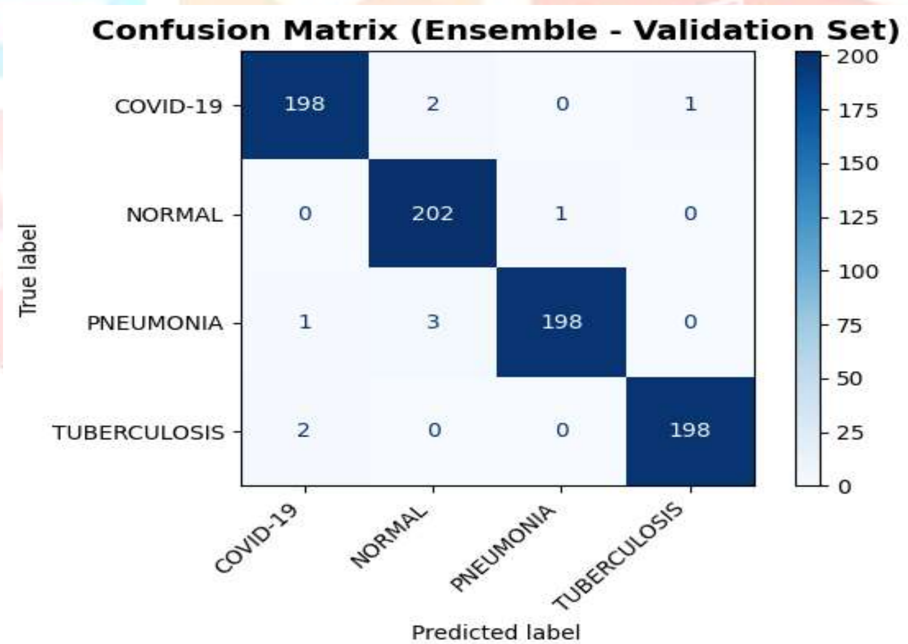

Figure 6: Ensemble Validation confusion matrix

The figure 6 presents the confusion matrix for the stacking ensemble model on the validation set, illustrating classification accuracy per class (COVID-19, Pneumonia, Tuberculosis, Normal). It shows how often the model correctly or incorrectly predicts each class, providing insight into its performance across different respiratory conditions.

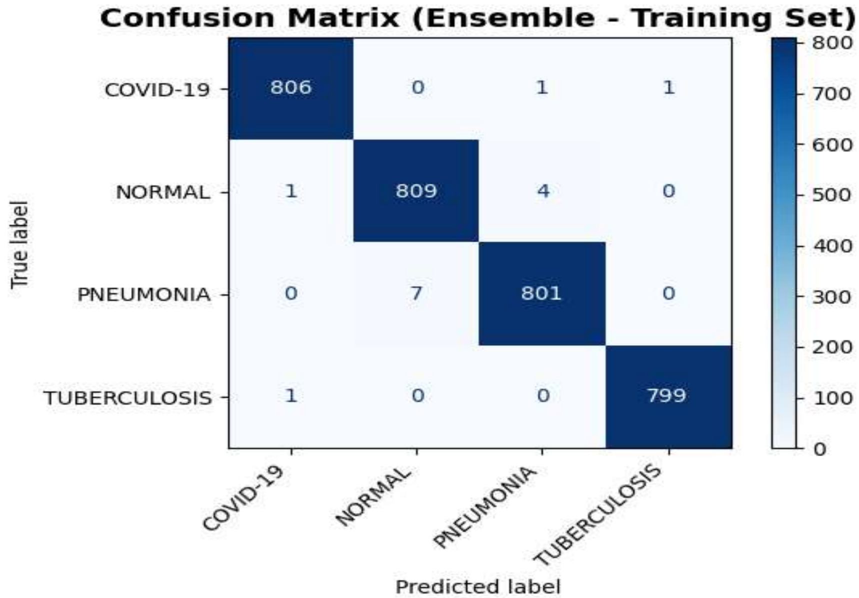

Figure 7: Ensemble Training Confusion Matrix

This figure 7 shows the confusion matrix for the stacking ensemble model on the training set, detailing the model's classification performance across the four classes during training. It complements Figure 6 by showing how well the model fits the training data, with a reported accuracy of 99.54%. The detailed breakdown of training and validation phase performance is summarized in Table 3.

Table 3: Performance Comparison of Individual Models and Proposed Model

MODEL	ACCURACY		PRECISION		RECALL		F1-SCORE	
	TRAIN	VAL	TRAIN	VAL	TRAIN	VAL	TRAIN	VAL
ALEX NET	93.25	94.44	94.44	94.46	94.45	94.44	94.15	94.44
INCEPTIONV3	83.67	83.70	85.36	83.75	83.67	83.70	83.45	83.66
PROPOSED MODEL ALEXNET+INCEPTIONV3	99.54	98.76	99.32	98.66	99.32	98.63	99.32	98.63

5.CONCLUSION:

This work presents an effective deep learning-based model for the multi-class classification of respiratory diseases COVID-19, Pneumonia, Tuberculosis, and Normal using chest X-ray images. A carefully curated Combined Dataset was developed from multiple public sources, and advanced preprocessing techniques, including data augmentation and class balancing, were applied to ensure robustness. The proposed stacking ensemble method, which combines AlexNet and InceptionV3 models, demonstrated superior performance by leveraging the strengths of both models. With a validation accuracy of 99.3% and consistent scores across precision, recall, and F1 metrics, the model proved highly accurate and reliable. The results underscore the potential of the proposed method to assist clinicians in fast, automated, and accurate diagnosis of respiratory conditions.

6.FUTURE SCOPE:

While the current model delivers promising results, there is ample scope for further improvement. Future work can focus on expanding the dataset with more diverse samples across age groups, imaging equipment, and geographic regions to enhance model generalizability. Incorporating explainable AI techniques such as Grad-CAM can improve model transparency and trust among medical professionals. Additionally, integrating patient metadata such as clinical history or symptoms with image features could enable more holistic and context-aware diagnosis. Optimizing the model for deployment on mobile or embedded devices can also support real-time screening in low-resource or remote healthcare settings. Lastly, the model could be adapted to other imaging modalities like CT scans or MRIs, broadening its utility across various diagnostic applications in medical imaging.

REFERENCES:

[1] K. S. Charan et al., "Transfer Learning Based Multi-Class Lung Disease Prediction Using Textural Features Derived From Fusion Data," IEEE Access, vol. 12, pp. 108248–108263, 2024.

[2] A.Naveen,K.Jhansi A Survey on Ensemble Learning Techniques for Medical image Classification. https://www.irjet.net/volume12-issue8

[3] Adaptive_Image_Preprocessing_and_Augmentation_for_Tuberculosis_Screening_on_Out-of-Domain_Chest_X-Ray_Dataset : https://ieeexplore.ieee.org/document/9901288

[4] A. U. Berliana and A. Bustamam, "Implementation of Stacking Ensemble Learning for Classification of COVID-19 using Image Dataset CT Scan and Lung X-Ray," 2020, pp. 148–153, https: //ieeexplore.ieee.org/document/9307691.

[5] S. Abimannan et al., "Ensemble Multifeatured Deep Learning Models and Applications: A Survey," IEEE Access, vol. 11, pp. 107194–107224, 2023, https://ieeexplore.ieee.org/document/ 10249304

[6] P. Dhar, "Multistage Ensemble Learning Model With Weighted Voting and Genetic Algorithm Optimization Strategy for Detecting Chronic Obstructive Pulmonary Disease," IEEE Access, vol. 9, pp. 49640–49656, 2021, https://ieeexplore.ieee.org/document/9385742.

[7] T. Rahman et al., "Reliable Tuberculosis Detection Using Chest X-Ray With Deep Learning, Segmentation and Visualization," IEEE Access, vol. 8, pp. 191586–191601, 2020, https://ieeexplore. ieee.org/document/9181631

[8] R. Ali, R. C. Hardie, and H. K. Ragb, "Ensemble Lung Segmentation System Using Deep Neural Networks," Proc. IEEE, 2020, https://ieeexplore.ieee.org/document/9201177.

[9] T. Xu and Z. Yuan, "Convolution Neural Network With Coordinate Attention for the Automatic Detection of Pulmonary Tuberculosis Images on Chest X-Rays," IEEE Access, vol. 10, pp. 86710–86718, 2022, https://ieeexplore.ieee.org/document/9849909.

[10] A. Elhanashi et al., "Classification and Localization of Multi-Type Abnormalities on Chest X-Rays Images," IEEE Access, vol. 11, pp. 83264–83275, 2023, https://ieeexplore.ieee.org/document/ 10194568.

[11] S. V. Militante, N. V. Dionisio, and B. G. Sibbaluca, "Pneumonia and COVID-19 Detection using Convolutional Neural Networks," 2020, pp. 1–6, https://ieeexplore.ieee.org/document/ 9270841.

[12] F. Jibril et al., "A Deep Learning Based COVID-19 Detection Framework," unpublished manuscript, available at IEEE Xplore, downloaded on June 03, 2025, https://ieeexplore.ieee.org/document/ 10541865.

[13] S. Nefoussi et al., "A Comparative Study of Deep Learning Networks for COVID-19 Recognition in Chest X-ray Images," 2020, pp. 237–241, https://ieeexplore.ieee.org/document/9377922.

[14] COVID-19 Radiography Dataset: https://www.kaggle.com/datasets/preetviradiya/covid19-radiography-d(Accessed May 22, 2021)

[15] Pneumonia Chest X-ray Dataset: https://www.kaggle.com/datasets/faisalahmed07/pneumonia-chestxray-da(Accessed November 2, 2023)

[16] Tuberculosis (TB) Chest X-ray Database: https://www.kaggle.com/datasets/tawsifurrahman/ tuberculosis-tb-chest-xray-dataset (Accessed June 14, 2021)