Multi-modal Classification of Lung Diseases using Deep Learning

Divyam Vijay
Dept of Computer Science
PES University
Bengaluru, India
PES1UG20CS136@pesu.pes.edu

Druva M Hegde
Dept of Computer Science
PES University
Bengaluru, India
PES1UG20CS155@pesu.pes.edu

Govinda Madhava
Dept of Computer Science
PES University
Bengaluru, India
PES1UG20CS155@pesu.pes.edu

Kaushal Suresh
Dept of Computer Science
PES University
Bengaluru, India
PES1UG20CS201@pesu.pes.edu

Prof. Preethi P
Dept of Computer Science
PES University
Bengaluru, India
preethip@pesu.edu

***Abstract*—This research introduces a comprehensive diagnostic system utilizing a multimodal approach for the analysis of Chest-X-ray images and Lung sound data. The primary focus involves the differentiation of prevalent respiratory conditions, including pneumonia and Covid-19, for which a multimodal approach combining both image and audio data is employed.**

The effectiveness of the system is underscored by the exploration of various multimodal methods, reflecting a nuanced approach to the fusion of image and audio features. This exploration contributes to the model's capability to discern intricate patterns within the data, enhancing its diagnostic accuracy. The studied multimodal methods provide valuable insights into the synergy of different data modalities for improved respiratory disease diagnosis.

I. INTRODUCTION AND BACKGROUND

Respiratory diseases pose a significant global health challenge, necessitating swift and precise diagnoses. This research introduces a multimodal diagnostic system that integrates Chest-X-ray images and Lung sound data. Focusing on prevalent conditions like pneumonia and Covid-19, our approach combines image and audio data, aiming to streamline and enhance the diagnostic process. By exploring various multimodal methods, we seek to develop a system that not only improves diagnostic precision but also eases the workload of healthcare professionals. This work contributes to advancing medical diagnostics, offering a more efficient tool for accurate respiratory disease identification.

II. LITERATURE SURVEY

In the work cited [1], the research paper presents a sophisticated deep learning model designed for the multi-class classification of common lung disorders, leveraging the Xception architecture. The model's training utilized a dataset comprising 11,716 chest X-ray photos from the National Institute of Health dataset repository, supplemented by an additional 443 local chest X-ray images from Jimma University Medical Centre. Notably, the study employed various denoising techniques, including the Wiener filter, Gaussian filter, Median filter, and Mean filter. Among these, the Median filter emerged as particularly effective in reducing salt and pepper noises, as well as speckle and Poisson noises in X-ray images.

To enhance image quality, the research incorporated the Contrast Limited Adaptive Histogram Equalisation (CLAHE) technique, selected for its ability to address the limitations associated with other histogram equalization methods. The study tackled data imbalance by implementing rotation-based data augmentation. The Xception model served a dual purpose, performing both feature extraction and image classification, ultimately achieving impressive accuracy levels. Noteworthy strengths of the proposed methodology include the utilization of a deep convolutional neural network architecture, adept noise removal and image enhancement techniques, a robust data augmentation strategy, and coverage of a diverse range of lung diseases.

However, the study is not without limitations. It employed a relatively small dataset, relied on inadequate pre-processing techniques, lacked segmentation (extraction of Region of Interest), and primarily focused on chest X-ray images. Despite these constraints, the proposed method demonstrated high accuracy and sensitivity in lung disease classification. Its potential impact on assisting healthcare professionals in making more informed diagnosis and treatment decisions for patients with lung diseases is evident.

The proposed strategy in the work cited [5] makes use of lung sound spectrograms to categorize them as crackles, wheezes, normal, or wheezes with crackles. The spectrogram pictures are then processed using a pretrained CNN model to extract deep characteristics. To improve the classification accuracy, the CNN model employs a parallel-connected pooling structure to retain the distinctive information lost during the max-pooling process. The collected deep features are then fed into the LDA classifier, and the RSE approach is used to boost classification performance even more. The findings reveal an average accuracy score of 71.15%, with the ‘normal’ label achieving the highest score and the ‘wheezes’ label achieving the lowest score. One advantage of the suggested technology is that it is less computationally expensive and complicated compared to Wavenet and CRNNs. Spectrograms give a concise representation of the audio signal, making it easier to extract essential characteristics. The approach, however, has drawbacks. Some of the most common lower-tract respiratory diseases cannot be diagnosed solely using Lung Sounds and their spectrogram images. There is a loss of temporal information since the link between the audio signal and the appropriate time intervals is no longer directly retained. Spectrograms have a fixed resolution in the frequency domain and may not catch fine-grained fluctuations in the audio stream.

The research cited in [6] proposes a respiratory sound classification approach for diagnosing COPD utilizing the wavelet transform and transfer learning algorithm. Breathing noises are classified as breathing cycles. These breathing cycles are analyzed using numerous wavelet families and spectral modifications to describe the data, followed by categorization using machine learning methods. In this paper, spectroscopy is utilized to extract signal features. Machine learning VGG16[18], Resnet50, and InceptionV3 were employed in this work to classify data on respiratory disorders. The suggested technique divides breathing sounds into breathing cycles to analyze them at a granular level, possibly capturing more nuanced and useful information for categorization. The limitations of the proposed method are that breathing sounds alone may not be sufficient for accurate diagnosis of respiratory diseases, other pre-trained models could potentially give better results, and there is a lack of reporting on false-negative and false-positive rates which are important performance metrics in medical diagnosis. Additionally, the reported high classification accuracy could indicate overfitting.

The proposed technique in this study cited [8] leverages the U-Net architecture as the foundation for an automated lung segmentation model. Comprising five coding layers and five decoding layers, the model integrates an EfficientNet-b4 encoder pre-trained on ImageNet. The decoder consists of five blocks, including a dropout layer, a two-dimensional convolution and padding layer, two residual blocks, and a LeakyReLU activation function. The use of residual blocks and LeakyReLU aims to mitigate issues related to ”vanishing gradient” and ”neuron death.” For mask generation, a 1*1 convolution layer is employed, followed by a ”Sigmoid” activation function.

Comparisons were made with the classic U-Net model and other state-of-the-art lung segmentation methods. Performance evaluation utilized metrics such as accuracy, specificity, sensitivity, Dice coefficient, and Jaccard index. The results revealed that the proposed technique outperformed the classic U-Net model and achieved comparable results to other advanced methods. Notably, the proposed technique exhibited lower standard deviation, indicating more consistent segmentation performance.

The decision to employ EfficientNet-b4 as the encoder and Residual blocks as the decoder facilitated effective feature extraction from the lung field, contributing to enhanced segmentation accuracy. However, the automated lung segmentation model exhibited limitations in processing images depicting certain conditions, such as pulmonary consolidation, lung effusion, lung edema, and atelectasis. Addressing these constraints necessitates further research to refine and improve the model’s performance.

In addressing the challenges of multimodal integration in the work cited [9], the work highlighted several key issues. Practical challenges, such as varying noise levels and conflicts between modalities, impede the full utilization of different data sources. Additionally, signals from diverse modalities often contain complementary information about various aspects of an object, event, or activity. Learning-based methods that effectively combine such information have the potential to yield more robust inferences. The work also underscores the need to tackle scenarios where the received data modality is weak, presenting a unique challenge in multimodal systems.

Existing multimodal techniques were explored, including early and late fusion, hybrid fusion, model ensemble, joint training methods, modality-specific decisions, and a multiplicative combination approach. Early and late fusion involve combining features before training and using vectors for each modality, respectively. Hybrid fusion is a combination of these methods. Model ensemble employs training multiple models on different modalities, later combining their outputs. Joint training methods, based on deep neural networks, simultaneously model features from multiple modalities. A unique approach involves making independent decisions on each modality using modal-specific models. The proposed multiplicative combination approach stands out by combining modalities in a differentiable and multiplicative fashion.

The work delves into multimodal deep learning, emphasizing its advantages and disadvantages. Multimodal deep learning offers superior performance, computationally tractable representation capability in vector format, and domain-specific models for each modality, aggregated by another function $p()$. However, it does not assume reliability in different modality inputs and faces challenges in training and regularization due to real-world data constraints.

III. METHODS

A. *Single Modality: Image Classification*

For the image classification task focusing on a single modality (images), a comparative study was conducted using three distinct models:

1) ***Inception Resnet****:* The Inception Resnet model was employed to leverage the benefits of a deep convolutional architecture. Its ability to capture intricate features in images was crucial for accurate classification.

2) ***MobileNet****:* MobileNet, known for its efficiency in terms of computational resources, was another model under consideration. Its lightweight architecture made it suitable for real-

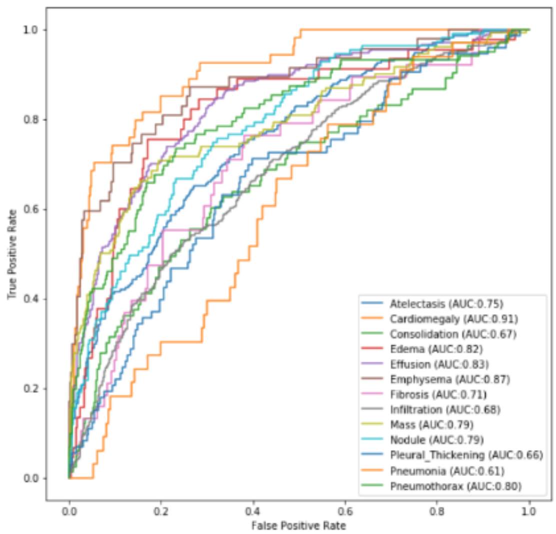

Fig. 1. Inception Resnet ROC curve

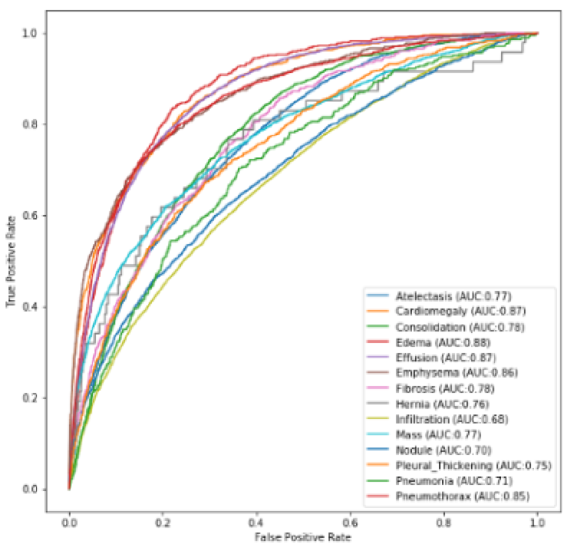

Fig. 2. MobileNet ROC curve

time applications, and we explored its performance in the image classification task.

*3) **Decision Tree from Scratch Classification**:* In addition to the utilization of pre-trained models, we undertook the development of a Decision Tree model from the ground up. This process involved the meticulous creation of a dataset through image segmentation. Initially, a U-Net model with pre-existing weights was applied for segmentation on a distinct dataset. Subsequently, these weights underwent fine-tuning for our specific task using the NIH chest X-ray dataset. In addressing segmentation challenges, manual segmentation was performed utilizing image editing software.

However, the manual segmentation approach encountered limitations, particularly in scenarios where diseases such as Atelectasis and Infiltration relied on features extending beyond mere shapes. Conditions of this nature involved nuanced factors such as the presence of white fluid, necessitating a more comprehensive approach.

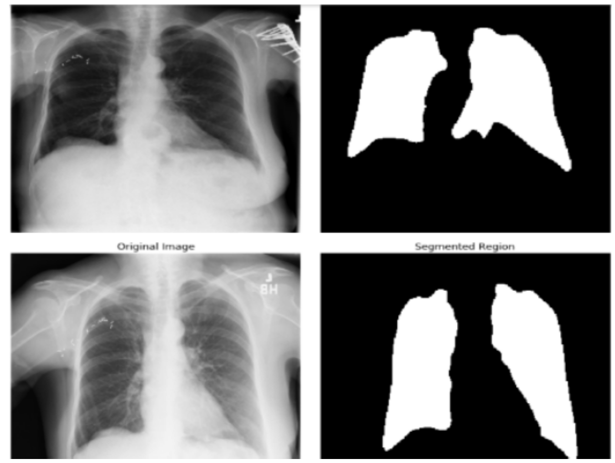

Fig. 3. MobileNet ROC curve

*4) **CNN approaches**:*

a) Approach 1:- CNN + Normalized Images: In this approach, a Convolutional Neural Network (CNN) is employed for classification using normalized images. Several key techniques are applied to enhance the model's performance:

- **Regularization**: To mitigate overfitting, regularization techniques are implemented, ensuring a balance between model complexity and generalization.
- **Weight Initialization**: Optimal weight initialization strategies are employed to set the initial weights, enhancing the convergence speed and avoiding the vanishing or exploding gradient problems.
- **Weight Constraints**: Constraints are imposed on the weights during training, preventing them from reaching extreme values and contributing to a more stable learning process.
- **Dropout**: Dropout layers are integrated into the CNN architecture, randomly deactivating a fraction of neurons during training to enhance the model's robustness.

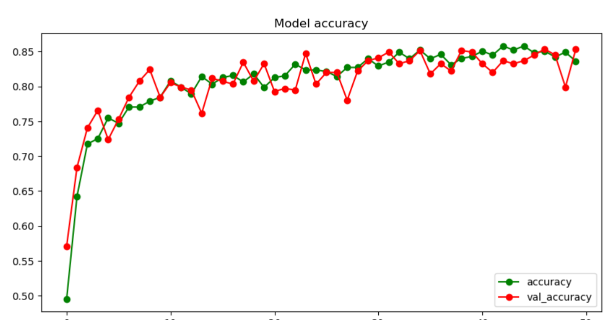

Fig. 4. Accuracy Graph for Train-Test

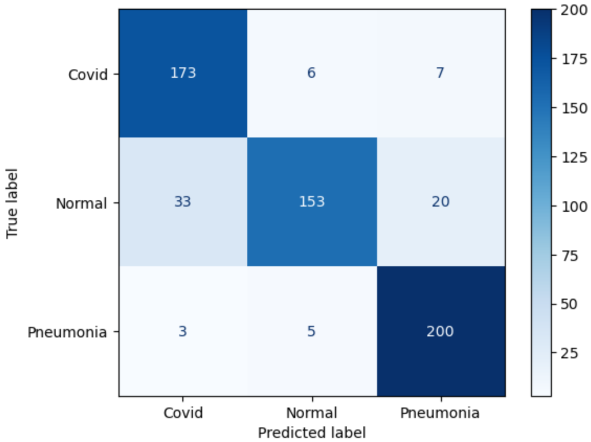

Fig. 5. Confusion Matrix

b) Approach 2: CNN + Reduced HOG: This approach involves the utilization of a CNN for classification, incorporating Reduced Histogram of Oriented Gradients (HOG) features. The approach includes the following steps:

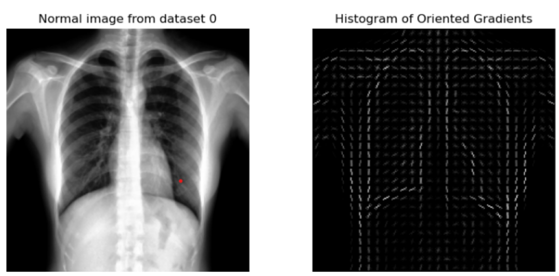

Fig. 6. HOG Feature Extraction

- **Normalized Image Preprocessing:** Adjusting pixel values to a standard range for consistent model training.
- **HOG Feature Extraction:** Capturing image structure through gradient orientation distributions.
- **PCA Feature Reduction:** Reducing dimensionality using principal component analysis.
- **CNN Classification:** Utilizing convolutional neural networks for image classification tasks.

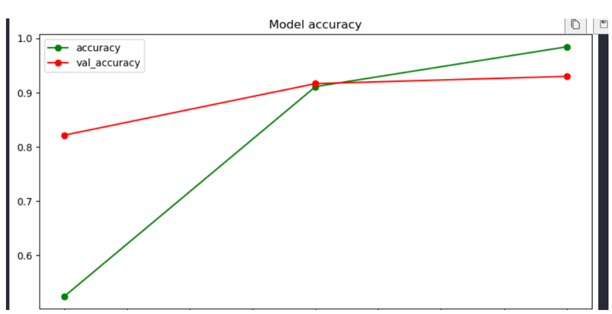

Fig. 7. Train-Test Validation Accuracy

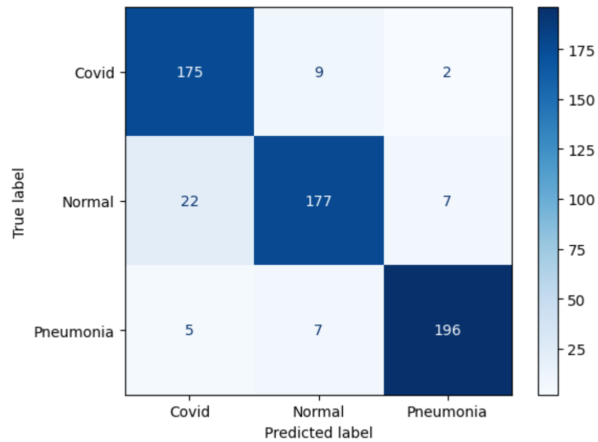

Fig. 8. Confusion Matrix

c) Approach 3: CNN + Reduced LPB: Similar to the second approach, this method involves CNN classification with reduced Local Binary Pattern (LPB) features. The workflow comprises:

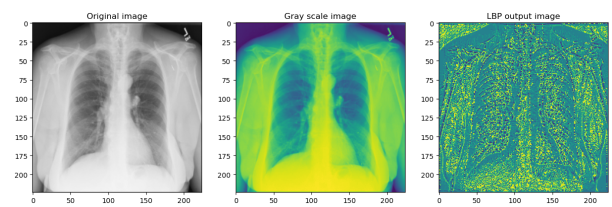

Fig. 9. LBP Feature Extraction

- **Normalized Image Preprocessing:** Adjusting pixel values to a standard range for consistent model training.
- **LPB Feature Extraction:** Describing local patterns in pixel intensity variations for texture analysis.
- **PCA Feature Reduction:** Reducing dimensionality using principal component analysis.
- **CNN Classification:** Utilizing convolutional neural networks for image classification tasks.

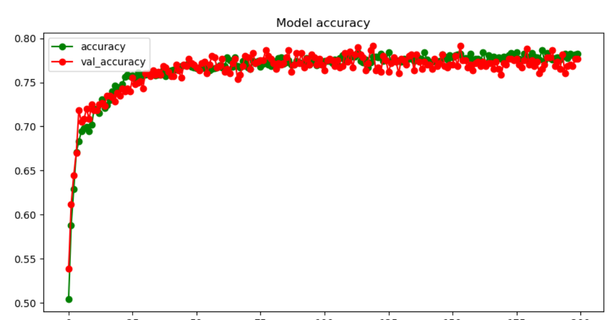

Fig. 10. Train-Test Validation Accuracy

*5) **Conclusion of Image Model Analysis**:* These approaches collectively form a comparative study, evaluating

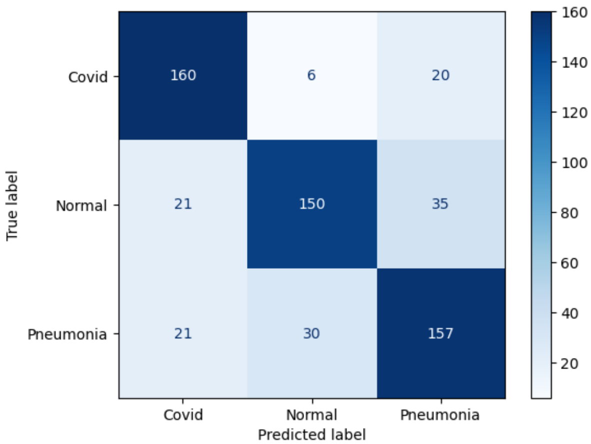

Fig. 11. Confusion Matrix

the effectiveness of different feature extraction techniques in conjunction with CNN's for image classification. The visual representations facilitate a nuanced understanding of each approach's strengths and limitations.

*a) **Advantages/Disadvantages of Methods used**:* Normalization techniques offer advantages such as scale invariance and accelerated convergence during model training, yet they come with the drawbacks of potential information loss and sensitivity to outliers in X-ray data. Feature reduction methods like HOG and LBP present trade-offs; HOG excels in capturing lesion boundaries through edge detection but may lose texture information, while LBP is proficient in identifying texture patterns crucial for disease identification but lacks sensitivity to pixel intensity variations. Among the model architectures discussed, CNN's demonstrate hierarchical feature learning, capturing both basic edges and complex patterns associated with respiratory diseases, though they may struggle with unrelated data. Random Forest, an ensemble learning approach, exhibits feature diversity with normalized, HOG, and LBP features, yet it may face challenges in representing complex spatial relationships in chest X-ray images. Mobile Net is praised for its efficiency in model size, suitable for resource-constrained environments, while Inception-ResNet-V2 focuses on fine-grained feature extraction but introduces computational complexity.

B. Single Modality: Audio Classification

*1) **Data Preparation**:* To prepare the image data for the lung sound classification model, we first merge three diverse datasets: the Lung Sounds Dataset, Respiratory Sound Database, and CoronaHack Respiratory Sound Dataset. We ensure a uniform representation of common class labels across the datasets, such as healthy and various respiratory ailments. The final dataset is curated to include a comprehensive range of respiratory sounds, facilitating a holistic approach to training the image model for accurate classification of different respiratory conditions.

Table 7.1: Accuracies of all models used for comparison

CNN Models a)Normalized b)HOG Feature Reduced c)LBP Feature Reduced	 79.802% 93.014% 72.224%
Random Forest a)HOG Feature Reduced b)LBP Feature Reduced	 77.23% 86.22%
MobileNet	77.87%
Inception-ResNet-V2	76.2%
Decision Tree from scratch	52%
Pretrained VGG16	82.009%

Fig. 12. Comparative Analysis of Images

In the preprocessing stage, we utilize augmentation techniques such as:

- **Speed Up**: Playback speed is increased by a randomly selected factor within the range of 0.8 to 1.2.
- **Speed Down**: Playback speed is decreased by a randomly selected factor within the same range.
- **Pitch Shift**: The audio is subjected to a random semitone shift, with the shift amount chosen from a predefined range of -2 to 2 semitones.
- **White Noise**: White noise is introduced to the audio signal, enhancing its robustness. The amplitude of the noise is controlled by the whitenoiselevel parameter.

Leveraging functions like speed up, speed down, pitch shift, and white noise injection, we diversify the dataset to improve model robustness.

Normalization techniques, such as maintaining a constant amplitude, are applied to ensure consistency across audio samples, contributing to a more effective training process for the image model aimed at classifying respiratory sounds.

*2) **Model Training**:* For the model selection, we compared two approaches:

a) ANN using Audio Features: The list of features we extracted from the audio files are:

- **Spectral Centroid**: Spectral centroid is the center of mass of the spectrum. It indicates where the "average" frequency of the sound is located.Spectral centroid can provide insights into the dominant frequency content of the lung sound. Changes in spectral centroid may indicate shifts in the pitch or tonal characteristics of the sound, which can be indicative of different respiratory conditions.
- **Spectral Roll-Off**: Frequency below which a certain percentage (e.g., 85%) of the total spectral energy lies.Spectral rolloff can help identify the cutoff frequency of the lung sound's spectral content. Changes in spectral rolloff may suggest alterations in the higher-frequency components of the sound, which could be associated with different lung conditions.
- **Root Mean Square (RMS) Energy**: Measures the magnitude of the audio signal's waveform. It quantifies the overall loudness and energy. RMS energy reflects the intensity and loudness of the lung sound. Changes in

RMS energy could indicate variations in the strength and intensity of the sound, potentially revealing important information about the underlying condition.

- **Zero-Crossing Rate**: Rate at which the audio waveform changes polarity (crosses the zero axis).Zero crossing rate can indicate the noisiness or abrupt changes in the lung sound. Some respiratory conditions may lead to increased or irregular zero crossing rates, which could be captured as features for classification.
- **MFCCs**: The Mel scale is a perceptual scale of pitches that corresponds more closely to how humans perceive the differences in frequency. It is non-linear and emphasizes lower frequencies where human hearing is more sensitive. Mel-scale filtering is used to convert the linearly spaced frequencies in the Fourier domain to the Mel scale. Also, MFCCs are more robust to noise variations making it a viable feature for this use-case.

The model architecture consists of sequential layers, starting with a dense layer of 512 units with a rectified linear unit (ReLU) activation function. Subsequent layers include 256 and 128 units with dropout layers to mitigate overfitting. The final layer, with softmax activation, outputs probabilities for different respiratory sound classes. The model is serialized to JSON format for architecture storage and HDF5 format for weight storage.

The neural network model is compiled using the Adam optimizer, which is known for its efficiency in training deep learning models. The chosen loss function is 'sparse categorical crossentropy,' suitable for multi-class classification tasks where the target labels are integers. The model is trained on the training data for 25 epochs with a batch size of 256. The 'accuracy' metric is employed to monitor the model's performance during training.

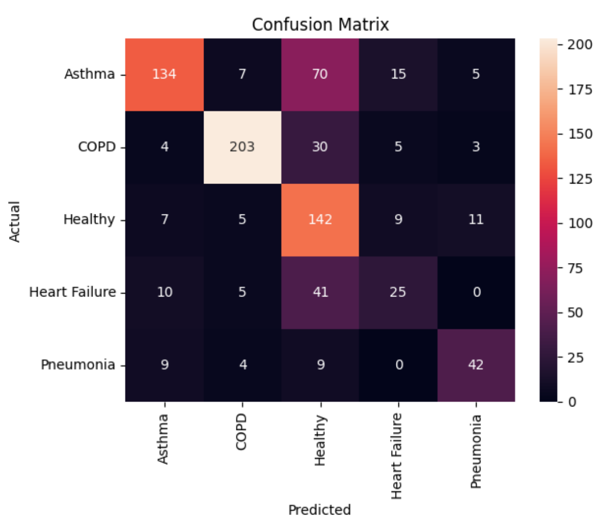

Fig. 13. ANN Confusion Matrix

b) CNN using Spectrogram Images and Transfer Learning: In this implementation, the fundamental concept revolves around generating spectrograms from audio files. Spectrograms serve as visual representations of the frequency spectrum of a signal over time, offering insights into the temporal variations of audio data. This process involves converting the audio signal into a time-frequency representation, which is then transformed into a logarithmic scale for enhanced interpretability. We utilize the Short-Time Fourier Transform (STFT) of an audio file, due to its ability to capture the frequency content of a signal as it evolves over time. Unlike the regular Fourier Transform, which provides a static representation of the entire signal, the STFT breaks the signal into short, overlapping segments and performs a Fourier Transform on each segment. This results in a time-varying representation of the signal's frequency content.

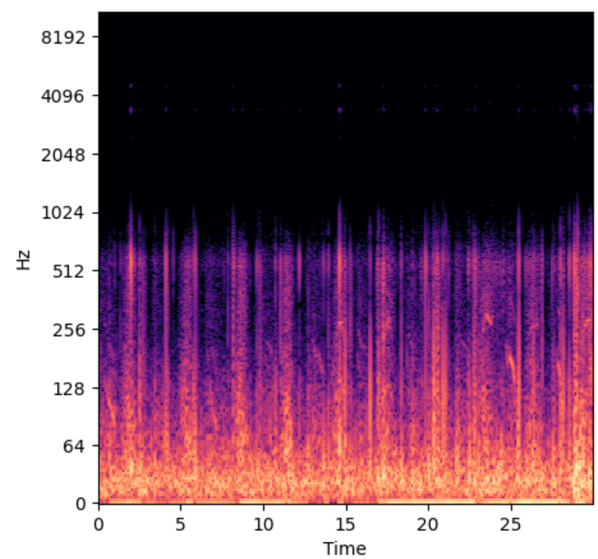

Fig. 14. Spectrogram Image of Audio

The VGG16 model is utilized as a base model, pre-trained on the ImageNet dataset. Leveraging a pre-trained model allows the network to capture generic features learned from a vast dataset, aiding in the effective extraction of relevant features for the new task.

A custom top model is added to the pre-trained VGG16 base. This top model includes densely connected layers with dropout for regularization. The final layer has a softmax activation function, suitable for multi classification tasks.

The model is compiled with specific modifications: The learning rate is decreased to 0.00001 for more controlled weight updates during fine-tuning. Categorical cross-entropy loss is chosen, indicating that the model is designed for multi-class classification. The batch size is increased to 32 for more efficient training.

The training process utilizes k-fold cross-validation, splitting the data into training and validation sets for each fold. This helps assess the model's performance across different subsets of the data, reducing the risk of overfitting to a particular set.

Early stopping and learning rate scheduler callbacks are employed to enhance training efficiency. Early stopping prevents overfitting by stopping training when the validation loss plateaus, and the learning rate scheduler adjusts the learning rate during training to optimize convergence.

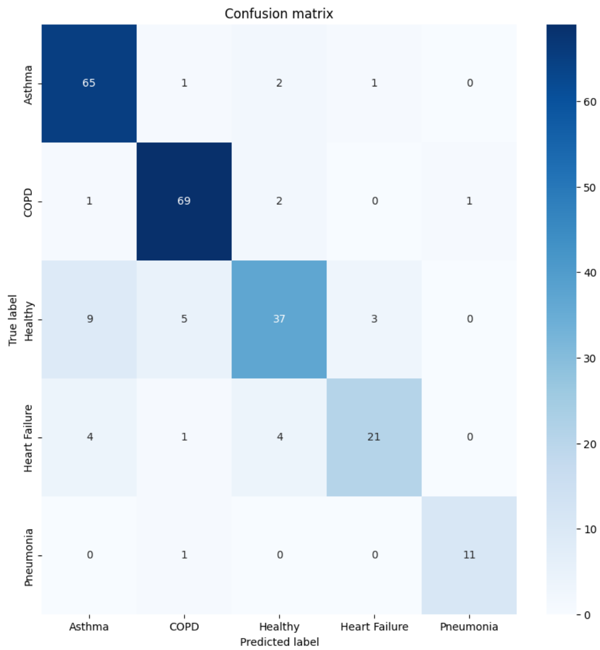

Fig. 15. Confusion Matrix for the VGG16+Transfer Learning Model

*c) **Advantages/Disadvantages of Methods used**:* In the comparative analysis of the two models, the Artificial Neural Network (ANN) exhibited a lower accuracy of 68.67%, revealing its limitations in capturing intricate patterns within the audio data. Notably, the ANN struggled with misclassifying critical respiratory conditions such as Asthma, COPD, and Heart Failure as Healthy. This suboptimal performance suggests that the inherent complexity of respiratory sound patterns may not be adequately captured by the chosen ANN architecture or that further tuning and feature extraction may be required to enhance its discriminative capabilities.

Conversely, the VGG16 + Transfer Learning model, pre-trained on spectrogram images of audio files, demonstrated significantly improved performance with an accuracy of 88.57%. The utilization of transfer learning, leveraging the knowledge gained from a pre-existing model, proved beneficial in extracting relevant features from the spectrogram images. The VGG16 + Transfer Learning model exhibited a higher capacity to discern subtle variations in respiratory sound patterns, leading to more accurate predictions. This notable performance difference emphasizes the efficacy of leveraging pre-trained models and specialized architectures, tailored to the intricacies of audio data, in addressing the challenges posed by respiratory sound classification tasks. Further exploration and fine-tuning of transfer learning models could offer avenues for even greater accuracy in the classification of respiratory conditions.

*C. **Multi Modality Classification***

We have used of the three multi-modal model techniques based on different approaches: accuracy scores, decision fusion (soft voting), and probability scores fed into an artificial neural network (ANN).

a) Accuracy Scores: The initial approach to evaluating multi-modal models involves assessing their individual accuracy scores. This method provides a straightforward comparison of the models' overall performance in terms of correctly predicting outcomes. Higher accuracy scores imply better model performance. However, this approach may overlook the nuances of decision-making, as it doesn't consider the confidence or certainty associated with each prediction.:

b) Decision Fusion (Soft Voting): Employing a decision fusion technique, such as soft voting, enhances the decision-making process by considering the confidence levels of both models. Instead of relying solely on accuracy, soft voting allows for a more nuanced approach by considering the probability distribution of each model's predictions. This approach is particularly effective when dealing with uncertain or ambiguous instances, providing a more robust and reliable decision-making mechanism.:

c) Probability Scores Fed into ANN: The third approach involves taking the probability scores generated by individual models and feeding them into an artificial neural network (ANN). This method leverages the power of neural networks to learn complex patterns and relationships from the probability distributions. By doing so, the model can capture subtle interactions between different modalities, potentially improving overall performance compared to a simple accuracy-based evaluation. This approach is especially useful when dealing with intricate and non-linear relationships within the data.

IV. RESULTS AND DISCUSSION

We are getting 98% accuracy for the multimodal approach, by which we can conclude that our model is working well. The results are the same as expected, as we are getting better accuracy for multimodality as compared to single modality. So it is satisfying the alternate hypothesis.

We also observe that more weightage is given to image probabilities, as evidenced by looking at multiple input cases. When both probabilities are strong, the output is as evidenced. But when they are disagreeing, images are considered because disease diagnosis is more accurate from chest x-ray images, since they are different for men and women, and different age groups. Only when the image probabilities are indecisive, the audio file is given more weightage.

V. CONCLUSION

The findings support the premise that combining multiple data modalities improves overall diagnostic accuracy for respiratory illnesses. Notably, the multimodal model placed a deliberate emphasis on picture probabilities, recognising the inherent differences in chest X-ray images across groups. This observation emphasizes the significance of demographic considerations in multimodal medical diagnoses.

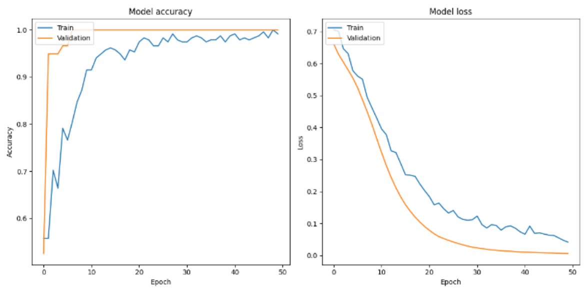

Figure 9.8: Plots of histories of accuracy and loss after multimodal model training

Classification Report:

	precision	recall	f1-score	support
Healthy	1.00	0.97	0.98	32
Pneumonia	0.96	1.00	0.98	27
accuracy			0.98	59
macro avg	0.98	0.98	0.98	59
weighted avg	0.98	0.98	0.98	59

Figure 9.9: Classification Report for multimodal model

Fig. 16. Result of MultiModal Classification

As far as future work is concerned, there are avenues for further enhancement and exploration. For example, we can extend the multimodal framework to include additional modalities such as Spirometry Data. The integration of diverse data sources could provide a more comprehensive understanding of respiratory health, potentially leading to more accurate and nuanced diagnoses. We can also investigate and implement advanced late model fusion technologies. This could involve exploring state-of-the-art techniques and architectures to further refine the integration of information from different modalities, ultimately improving the overall system performance.

Because the scope of our project was limited, we were able to integrate the functionality of classifying only a few diseases. This can be enhanced if additional patient data for both picture and audio for more disorders becomes available. The goal is to contribute to the evolution of diagnostic tools through ongoing refinement and research, ensuring that disease diagnosis utilising multimodal data remains at the forefront of innovation in respiratory healthcare.

ACKNOWLEDGEMENT

We extend our sincere gratitude to Prof. Preethi P, Department of Computer Science and Engineering, PES University, for her unwavering guidance, invaluable assistance, and constant encouragement throughout the development of this UE19CS390A - Capstone Project Phase – 1.

Our appreciation extends to the project coordinator, Dr. Priyanka H., the esteemed panel members, and the dedicated supporting staff for their efforts in organizing, managing, and facilitating the entire process.

Special thanks are due to Dr. Shylaja S S, Chairperson, Department of Computer Science and Engineering, PES University, for the wealth of knowledge and support provided.

We would like to express our gratitude to Dr. M. R. Doreswamy, Chancellor, PES University, Prof. Jawahar Doreswamy, Pro Chancellor – PES University, Dr. Suryaprasad J, Vice-Chancellor, and Dr. B.K. Keshavan, Dean of Faculty, PES University, for their continuous support, providing us with various opportunities, and offering enlightenment at every step of our journey.

Lastly, we acknowledge the unwavering support and encouragement from our families and friends, without which this project would not have been possible.

REFERENCES

[1] Yimer, F., Tessema, A., Simegn, G. (2021). Multiple Lung Diseases Classification from Chest X-Ray Images using Deep Learning approach. *International Journal of Advanced Trends in Computer Science and Engineering*.

[2] Fraiwan, M., Fraiwan, L., Alkhodari, M., et al. (2022). Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory. *J Ambient Intell Human Comput*, 13(4759–4771). DOI: 10.1007/s12652-021-03184-y.

[3] Pattrapisetwong, P., Chiracharit, W. (2016). Automatic lung segmentation in chest radiographs using shadow filter and local thresholding. In *2016 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, Chiang Mai, Thailand. DOI: 10.1109/CIBCB.2016.7758113.

[4] Goyal, S., Singh, R. (2021). Detection and classification of lung diseases for pneumonia and Covid-19 using machine and deep learning techniques. *J Ambient Intell Human Comput*. DOI: 10.1007/s12652-021-03464-7.

[5] Demir, F., Ismael, A. M., Sengur, A. (2020). Classification of Lung Sounds With CNN Model Using Parallel Pooling Structure. In *IEEE Access*, 8, 105376-105383. DOI: 10.1109/ACCESS.2020.3000111.

[6] Han-Trong, T. (2022). Diagnosing Chronic Obstructive Pulmonary Disease Based on Breathing Sound Using Machine Learning. In *2022 7th National Scientific Conference on Applying New Technology in Green Buildings (ATiGB)*, Da Nang, Vietnam. DOI: 10.1109/ATiGB56486.2022.9984119.

[7] Bharati, S., Podder, P., Mondal, M.R.H. (2020). Hybrid deep learning for detecting lung diseases from X-ray images. *Inform Med Unlocked*, 20, 100391. DOI: 10.1016/j.imu.2020.100391.

[8] Liu, W., Luo, J., Yang, Y., Wang, W., Deng, J., Yu, L. (2022). Automatic lung segmentation in chest X-ray images using improved U-Net. *Sci Rep*, 12(1), 8649. DOI: 10.1038/s41598-022-12743-y.

[9] *Source Information*. DOI: https://doi.org/10.48550/arXiv.1805.11730.