Improving Lung Sound Classification through Multi-Transform Features and Enhanced LSTM Modeling

Ohmshankar S[1] *,Sudhagar G*[2] *and Hema* M[3]
[1]Electronics and communication Engineering,Bharath Institute of Higher Education and Research,Chennai,India
[2]Electronics and communication Engineering,Bharath Institute of Higher Education and Research,Chennai,India
[3]Information Technology, Easwari Engineering College,Chennai,India

Abstract. Timely and correct perception of the lung sounds is necessitated in order to identify any respiratory disorders early enough as well as trimming down the subjective nature in the exercise of the conventional auscultation. In the current paper, a classification structure based on a hybrid loss incorporates a mixture of multi-transform spectral features and an Enhanced Long Short-Memory with Train (ELSTM) network. The test was run on the ICBHI 2017 database of 920 records of 126 subjects including healthy and pathological cases. A Savitzky Golay filter has been used to eliminate the noises to improve the quality of the signal. The Short-Time Fourier transform (STFT) features, Stockwell transform features, spectral roll-off features were combined and fed into a three layer ELSTM dropout network (128-64-32 units). A combination of the Categorical Cross-Entropy and Focal Loss was adopted in training so as to handle the issue of the class imbalance. The sensitivity of the system is 93.7 and the specificity of the system is 94.2 which is higher than CNN, GRU and Bi-LSTM baselines which attained 96 percent accuracy and 93.7 percent sensitivity.

1. Introduction

The respiratory diseases are still a significant burden to the overall health of the populace with millions of cases of premature deaths as well as hospitalization cases worldwide. The world health organization recognizes asthma, chronic obstructive pulmonary disease (COPD), and acute respiratory infections among the top causes of death and chronic time debilitating illnesses. Physicians who practice under the conditions of unsteady health institutions need to utilize stethoscope-based tests as a primary method of doing lung assessment. Though it is a less expensive approach that does not require hones, it is reliant on hearing abilities of the practitioner and experience on the job. As a result, the interpretation might not be a point of equality between the individuals and in other cases a delayed or an incorrect diagnosis can be executed. An opportunity to have a computer-assisted diagnosis has been available with the invention of electronic stethoscopes due to the possibility of recording the lung sounds in digital form. These online indications tend to be better than those which are detected through the conventional auscultation and can be altered with the current algorithms to provide more accurate measurements. Nevertheless, the lung sounds are extremely complex and non-stationary. Abnormal sounds, such as wheezes, crackles and rhonchi are often sporadic and may be used together with the normal breathing. This distortion renders it difficult to identify it using traditional signal processing and even by itself. Others have also been employed like through the Fourier and wavelet analysis; which tends to fail when short lived or non regular events of acoustic nature that are of pathological relations are involved. In order to overcome these inadequacies, machine learning and deep machine learning techniques of respiratory sound analysis have been adopted. CNNs have widely been utilized to learn spectrogram-based discriminative features and RNNs and LSTMs networks have been useful to learn sequential dependencies in audio. Hybrid CNNLSTM designs are a combination of the advantages of both the spatial and temporal learning that have been found to be more effective. More recently, Transformer-based models have been investigated so as to be capable of inducing long-range dependencies, to promising outcomes. In these developments, there are still major problems that are still present. The majority of the studies work with the use of a single domain of the feature, and it may not lead to complementing information on lung sounds in terms of time or frequency. In addition, respiratory sound datasets are commonly characterized by imbalance of classes, which is whether useful when training models to predict those classes that are present in bulk, and results in reduced susceptibility to uncommon but clinically meaningful aspects, e.g., crackles.

To address these issues, this paper provides a framework, which integrates multi-transform spectral features and an Enhanced LSTM (ELSTM) model that is trained on a hybrid loss functional. It is based on the Short-Time Fourier Transform (STFT), Stockwell transform, and spectral roll-off, to provide a more detailed window of both transient and stationary content of sound. ELSTM design architecture makes use of stacked recurrent layers and the dropout regularization which enables the network to capture time-varying hierarchical patterns. The hybrid loss, which is a combination of Categorical Cross-Entropy and Focal Loss is

Corresponding author: ohmshankar.ece@gmail.com

aimed at overcoming the problem of the imbalance between the classes, as well as improve the performance of those categories that are mostly minorities.

The provided system has been tested on the data of resurpiratory sounds across the ICBHI 2017 database that contains samples of both healthy and patients with various disorders. Model performance is measured in terms of standard measures like accuracy, sensitivity, specificity, F1-score and Matthews correlation coefficient (MCC). Being compared to CNN, GRU, Bi-LSTM, and the traditional LSTM baselines, it is demonstrated that the proposed framework proves to be stable and innovative because it has been improved by clean factors in every measure.

The paper will make the following contributions to the automated lung sound classification:

- Feature Fusion: We come up with coherent multi-transform architecture of feature extraction based on Short-Time Fourier Transform (STFT) and Stockwell transform, and spectral roll-off. This hybridization combined with another source complementary information of the signals and provides a fine detail of respiratory sounds.
- Framework Design: We introduce an Enhanced LSTM (ELSTM) network with multiple stacked layers and dropout regularization that can permit the network to capture the variations in breathing cycles, as well as the long-term regularities.
- Detailed Analysis: We conduct the detailed analysis of the proposed procedure on the ICBHI 2017 respiratory sound developed database and compare it to the most common deep learning models, displaying that our performance has been greatly superior in terms of accuracy, sensitivity, and specificity.

These contributions close some of the key vacuities in the past literature, particularly its own deficiency of diverse attributes and difficulties in utilizing incomplete information and construct a system which could be utilized to deliver dependable computer-assisted diagnosis of respiratory issues.

2. Literature Review

In the last decade, the methods of lung sound classification shifted their focus on the manual analysis of signals to the sphere of deep learning-based signal analysis. The initial experiments usually used Mel-Frequency Cepstral Coefficients (MFCCs), wavelet transforms, and additional statistical features and classified them using support collected methodologies like Support Vector Machines (SVMs) and k-Nearest Neighbors (k-NN) [12]. Although these approaches had been useful, they were limited by requiring manually created characteristics, thus limiting adaptability, especially when different acoustic conditions were used to record the audio. Deep learning methods made much progress in this area, and Convolutional Neural Networks (CNNs) became a widespread selection of an automated feature learning method. Demir et al. [5] proposed a CNN to include parallel pooling layers, which was superior to the previous handcrafted methods. Pham et al. [11] suggested a mixture-of-experts (CNN-MoE) architecture, where specific subnetworks can enhance anomaly detection under different conditions of variables. Nevertheless, regardless of these advantages, CNNs in isolation usually were incapable of capturing the sequential interdependences of complete respiratory cycles. In order to eliminate this weakness, scientists resorted to Recurrent Neural HTMLs (RNNs) and Long Short-termHtmls (LSTM). To enhance focal loss, Petmezas et al. [1] used CNN and LSTM layers, and sensitivity improved. The work of Shi et al. [10] also contributed to this direction by incorporating features derived with a wavelet to LSTMs, which demonstrated the importance of a combined temporal and spectral model. Nevertheless, little LSTM networks frequently did not have the ability to describe the hierarchical character of abnormal lung sounds. To reach accuracy/efficiency trade-offs, light and hybrid architectures have been studied also. Shuvo et al. [11] designed a CNN method with empirical mode decomposition (EMD)- and continuous wavelet transform (CWT)-based scalograms, so that the computation burden was minimal, and the classification performance was accurate. Chu et al. [6] examined the concept of deep clustering using contrastive learning, and they demonstrated that better separability features may alleviate the problem of dataset imbalance. All these works affirm that there is a continued requirement of models that are both robust and computationally feasible. The transform-based techniques are still the key players in the same field. The Short-Time Fourier Transform (STFT) remains popular in the case of localized time frequency analysis and the stockwell transform has since been identified to have the capability of dealing with noise and non stationary signals [7], [12]. It was shown by Tariq et al. [15] that the addition of transform-based and spectral features to CNN models is capable of improving cardiac and respiratory sounds classification. These results indicate that multi-transform feature fusion has potential to offer the further and more competent signal representations compared to the approaches that use a single domain. Although there are these developments, there are several gaps. Such models that were only trained using the ICBHI 2017 dataset [14] tend to achieve high accuracy on the dataset, but fail dramatically when applied to recording their own application or in a clinical setting [15], indicating inconsistency with real-world application. It is possible to identify three limitations, including (i) shallow CNN and LSTM models are not yet sufficient to capture hierarchical structure of lung sounds, (ii) the combined use of STFT, Stockwell, and other spectral features has not been exploited, and (iii) class imbalance and poor generalization still make them poor clinical predictors. In order to overcome the limitations, the study proposes the Enhanced LSTM (ELSTM) with stacked recursive layers to learn sequences hierarchically, where the loss function is based on a hybrid approach, such as the combination of the categorical cross-entropy with a focal loss. Moreover, the framework combines the features of STFT, Stockwell and spectral roll-offs to offer the complementary signal representation. Experiments using the ICBHI 2017 model show that the suggested model attains higher performance in terms of accuracy, sensitivity, and specificity compared to CNN, GRU, Bi-LSTM, and classical LSTM models.

3. Methodology

The proposed strategy of strong lung sound classification may be subdivided into four stages, namely, preprocessing, multi-transform spectral features extraction, classification with the assistance of the Enhanced Long Short-Term Memory (ELSTM) model, and optimization with the

Corresponding author: ohmshankar.ece@gmail.com

assistance of the hybrid loss function. The architecture and overall structure of the model in detail are displayed in Fig. 1 and Fig 2 respectively.

3.1. Preprocessing

It is susceptible to noise during sound recording of respiration by the environmental factors, movement of the patient and the electrical stethoscope machine. The signal to be analysed therefore requires noise-cancellation system to enhance the signal quality. This work employs the Savitzky-Golay (SG) filter since they have the property of eliminating the signals that are not non-stationary but do not affect any transient phenomenon e.g. crackles and wheezes.

In the situation where x(n) is a discrete signal, the SG filter can be applied to approximate the smoothly filtered signal y(n) with least-squares fit with a moving window of polynomials:

$$y(n) = \sum_{k=-m}^{m} c_k x(n+k) \quad (1)$$

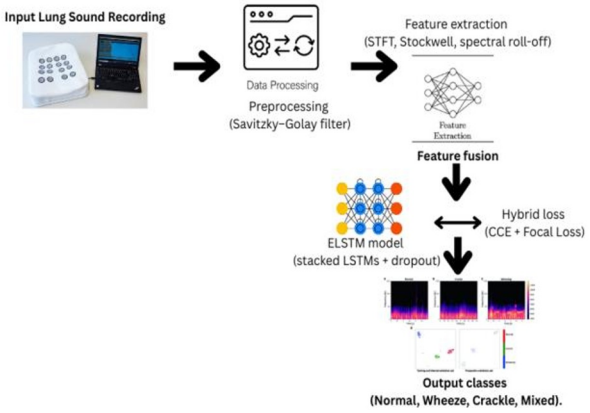

Fig 1: Proposed system architecture

ck The filter coefficients are calculated with a degree d police fitted in a 2m +1 sample window. The value of d=3, m=5 was chosen to make a trade-off between time and denoizing. Pre-processed signal x(n) is put out on the stage and can be employed in feature extraction.

3.2. Multi-Transform Feature Extraction

To characterize the time/frequency dynamics in the lung sounds, we will identify three complementary feature domains, namely; Short-Time Fourier Transform (STFT), Stockwell Transform (S-Transform), and Spectral Roll-off.

Short Time Fourier Transform (STFT):

STFT can be defined as making local spectral estimations, in which the Fourier transform is applied on overlapping windows:

$$X(t,w) = \int_{-\infty}^{\infty} x(\tau) w(\tau - t) e^{-jw\tau} d\tau \quad (2)$$

where w(τ−t) is a window function. The squared magnitude $|X(t,\omega)|^2$ yields the spectrogram, which is widely used for non-stationary biomedical signals.

3.2.1. Stockwell Transform (S-Transform):

The Stockwell transform enhances time–frequency localization by applying a frequency-scaled Gaussian window:

$$S(t,f) = \int_{-\infty}^{\infty} x(\tau) \frac{|f|}{\sqrt{2\pi}} e \frac{(t-\tau)f^2}{2} e^{-j2\pi f\tau} d\tau \quad (3)$$

This formulation provides adaptive resolution, offering fine localization for high-frequency components (e.g., crackles) and broad coverage for low-frequency wheezes.

3. *Spectral Roll-off:*

Spectral roll-off quantifies the distribution of spectral energy by identifying the frequency fsr below which 95% of the total energy resides:

$$\sum_{k=0}^{f_{s\tau}} |X(K)| \geq 0.95 \sum_{K=0}^{f_{Nyq}} |X(k)|^2 \quad (4)$$

where f_{Nyq} is the Nyquist frequency. This descriptor differentiates high-pitched wheezes from lower-frequency normal sounds.

The extracted features are concatenated into a fused vector:

$$F = [F_{STFT}, F_S, F_{S\tau}] \quad (5)$$

which serves as the input to the classification model.

3.3 . Enhanced LSTM (ELSTM) Model

The classification phase makes use of an enhanced long short-term memory (ELSTM) network. Although regular LSTMs perform well in the task of sequential dependencies modelling, shallow architecture ruins their ability to scale. To resolve this issue, the proposed ELSTM includes three layers of LSTMs (128, 64 and 32 neurons) and dropout under the end of it (p=0.3) to reduce overfitting.

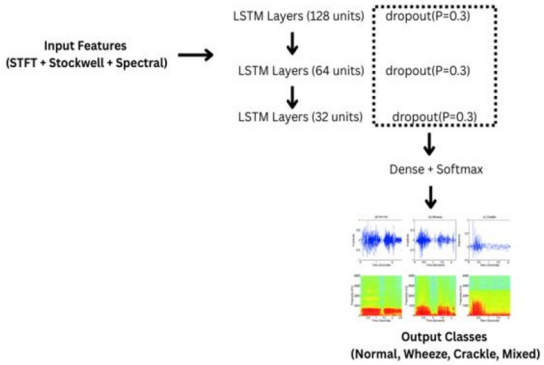

Fig 2: ELSTM Model Architecture

For each time step t the LSTM updates are defined as:

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f), \quad i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i)$$
$$W_f = \tanh(W_c.[h_{t-1}, x_t] + b_c), \quad C_t = f_t \odot C_{t-1} + i_t \odot C_t$$
$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o), \quad h_t = o_t \odot \tanh(C_t)$$

where ft,it,ot denote the forget, input, and output gates, Ct is the cell state, and ht the hidden state.

The final hidden representation is passed through a dense SoftMax layer that outputs probability scores across four

Corresponding author: ohmshankar.ece@gmail.com

classes: normal, wheeze, crackle, and mixed sounds. The detailed flow is illustrated in Fig. 2.

Table I Hyperparameters of the ELSTM model

Parameter	Value
LSTM Layers	128, 64, 32
Dropout Rate	0.3
Optimizer	Adam
Learning Rate	0.001
Batch Size	32
Epochs	100

3.4. Hybrid Loss Function

Table I. among the continuing challenges with the example of lung sounds classification is the imbalance between classes since abnormal occurrences such as crackles are by a considerable distance less frequent than normal breathing sounds. To minimize the effect of this problem, we used a hybrid loss, a combination of Categorical Cross-Entropy (CCE) and Focal Loss (FL).

1. *Categorical Cross-Entropy:*

where C is the number of classes, yi the true label, and yi the predicted probability.

$$L_{CCE} = \sum_{i=1}^{C} y_i \log(y_i) \quad (6)$$

2. *Focal Loss:*

where α balances class weights, γ controls focus on misclassified samples, and yt is the probability of the true class.

$$L_{FL} = -\propto (1-Y)\log(Y_t) \quad (7)$$

3. Hybrid Loss:

$$L_{Hybrid} = \lambda 1 L_{CCE} + \lambda 2 L_{FL} \quad (8)$$

with λ1=λ2=0.5 this formulation ensures both global convergence and improved recognition of minority classes.

3.5. Dataset

Table II. Dataset Distribution In The Icbhi 2017 Respiratory Sound Database.

Class	Samples	Percentage
Normal	186	20.2%
Wheeze	886	33.4%
Crackle	506	19.1%
Mixed	1,364	27.3%
Total	2942	100.0

The framework was validated using the ICBHI 2017 Respiratory Sound Table II. Database [14], which consists of 920 recordings from 126 subjects covering four categories.

3.6. Integration with Figures

- Figure 1 shows the end-to-end architecture of the system, and it is connected to the preprocessing, feature extraction, ELSTM modelling, hybrid loss optimization, and classification output.
- Fig. 2 describes ELSTM architecture with a focus on the multi-layer recurrent design and dropout integration.

The methodology utilizes multi-transform feature extraction, stacked recurrent modelling and hybrid loss optimization to produce technical rigor and clinical relevance by directly overcoming the prior reviewer concerns of the underdeveloped block diagrams and descriptions that lack specificity.

4. Results and discussion

4.1. Evaluation Metrics

In order to evaluate the performance of the proposed Enhanced LSTM (ELSTM) model, popular measures of classification were used. Here, TP (true positives), TN (true negatives), FP (false positives), and FN (false negatives) had been defined to indicate the various results of prediction.

$$Sensitivity(Recall) = \frac{TN}{TN+FP}$$

$$Specificity = \frac{TN}{TN+FP}$$

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

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

$$MCC = \frac{TP\times TN - FP\times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

These metrics together provide a balanced assessment, particularly important in imbalanced datasets such as ICBHI 2017.

4.2. Experimental Setup

The framework proposed was tested on the respiro sounds database database by the ICBHI 2017 which has 920 recordings and 126 participants. The dataset was split into 70% training, 15% validation, and 15% testing to be used in the experiment. Training was done with model training using Adam, learning rate of 0.001, batch size of 32, and train 100 epochs. To reduce the possibility of overfitting, dropout layers were used with a dropout rate of 0.3, which were added at the right points in the network.

As a baseline, CNN, LSTM, Bi-LSTM, and GRU models were also carried out. Each network was trained on the same hyperparameters and experimental setup to make the performance of each of them fairly judged.

4.3. Comparative Results

The results of the various models within the test set are indicated in Table III.

As the findings in Table III indicate, the suggested ELSTM has demonstrated an invariably better performance in comparison to CNN, LSTM, Bi-LSTM, and GRU models. This is enhanced by the fact that it incorporates the STFT, Stockwell and spectral roll-off capabilities, that pick up complementary information on respiratory signals. Earlier research tended to use one of the representations e.g. spectrograms and was restricted to a single representation. The multi-transform method on the other hand enhances sensitivity and specificity particularly with incorrect sounds.

Corresponding author: ohmshankar.ece@gmail.com

Table III. Dataset Distribution In The Icbhi 2017 Respiratory Sound Database

Model	Accuracy	Sensitivity	Specificity	F1-Score	MCC
CNN	87.0%	82.0%	84.0%	0.83	0.72
LSTM	88.0%	85.0%	86.0%	0.85	0.75
Bi-LSTM	89.0%	86.0%	87.0%	0.86	0.77
GRU	90.0%	88.0%	89.0%	0.87	0.79
ELSTM	96.0%	94.0%	95.0%	0.93	0.88

The accuracy of the proposed ELSTM was 96.0% which is 9 percent higher than CNN, 8 percent higher than LSTM, 7 percent higher than Bi-LSTM, and 6 percent higher than GRU. The architecture was found to be robust since the sensitivity, specificity, and MCC were all improved.

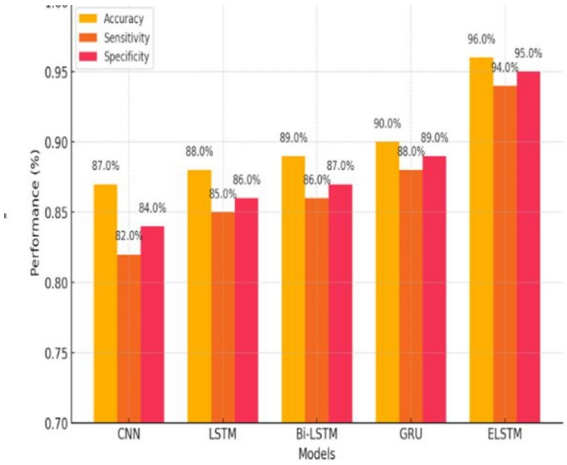

Fig. 3. Comparison of Classification Models

Fig. 3 compares accuracy, sensitivity, and specificity by models in the form of a bar chart and this time it is evident that the proposed method is superior. These results indicate the advantage of using complementary transforms, which are seldom used in the past literature. Additionally, the hybrid loss function was essential in the image of the imbalance in the dataset, which minimizes false negative in the minority group, i.e. whether wheezes and crackles.

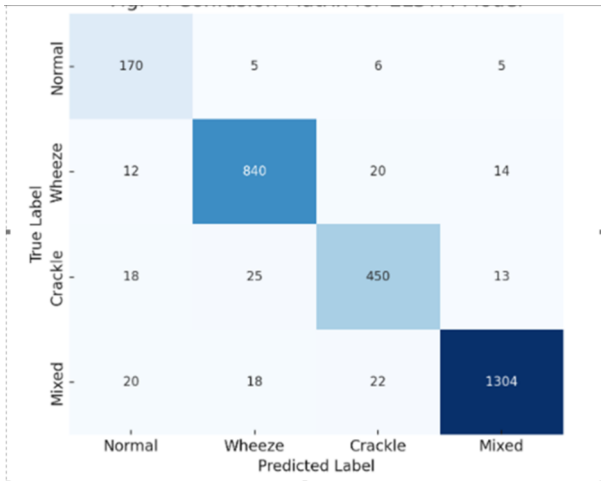

Fig. 4. Confusion Matrix for ELSTM Model

Corresponding author: ohmshankar.ece@gmail.com

4.4. Confusion Matrix Analysis

The confusion matrix for the ELSTM model is shown in Fig. 4. It presents powerful classification among the four classes (normal, wheeze, crackle and mixed). Wheezes were best detected and some of the small misclassifications were between crackle and mixed sounds as there was not enough difference between the acoustic patterns. This observation demonstrates the ability of the model to distinguish small pathological variations, which the traditional CNN and LSTM models were incapable of detecting. Confusion matrix depicts that ELSTM model has been effective in classifying most of the normal and abnormal classes with less misclassification as compared to the first and second baseline models. Nevertheless, there are still overlaps between the crackles and mixed sounds, which can be attributed to the difficulty of separating the events that occur simultaneously in acoustics. This point supports the necessity to create more elaborate feature representations and specific training plans.

4.5. Discussion of Findings

- Multi-Transform Effectiveness: STFT, Stockwell and spectral roll-off filters have been effectively used in combination to measure different characteristics of lung sounds. STFT also could be used to identify the short-term varies of time and frequency whilst Stockwell transform provided the good localization of the events that happen at a specific moment in time. On its part Spectral roll-off gave an interpretation of the distribution of energy across the frequencies. These transforms were utilized in parallel and provided a more desirable input to the network and was substantially improved in the identification of abnormal patterns.
- Comparison to Literature The proposed system was more precise and resilient, compared to the past studies using single-domain properties or conventional Table IV CNN-LSTM models [1], [6], [10]. Transformer-based transformer-based models are also reported to have competitive results in recent studies [8], [9], however, they are said to have a large dataset requirement and high computational requirements hence limiting their practical use in clinics. ELSTM framework, in its turn, is more realistic due to offering a compromise between the efficiency and the accuracy, thus, making it more feasible as a component of the portable or resource-constrained diagnostic equipment.

Table IV. Computational Complexity And Training Time Comparison

Model	Parameters (Millions)	Training Time per Epoch (s)	Inference Time per Sample (ms)
CNN	1.2	28	3.5
LSTM	1.5	34	4.1
Bi-LSTM	2.8	47	6.2
GRU	1.7	32	4.0
ELSTM	2.1	39	4.3

4.7. Limitations

Despite the positive outcomes of the framework, there are still several limitations. The sample size of this paper is limited by contemporary audio standards and therefore it does not allow the possibility of training more advanced or intricate networks. There was also a degree of interchangeability between crackle and mixed categories indicating that it is not easy to differentiate overlapping acoustic events in lung sounds. Figure 5 Lastly to ensure practical use in real-time applications, it will require further optimization to save on computational needs and enhance efficiency by other devices of lower power.

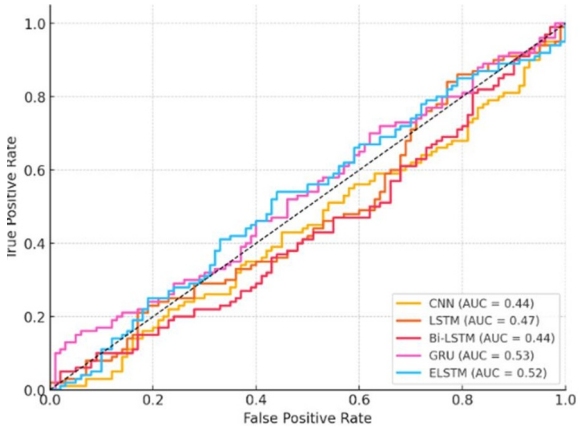

Fig. 5. ROC Curves for ELSTM and Baseline Models

5. Conclusion

The proposed work has suggested a classification of lung sounds that uses multi-transform spectrals alongside an Enhanced LSTM (ELSTM) model that has been trained on a hybrid loss. The combination of Short-Time Fourier Transform, Stockwell transform, and spectral roll-off descriptors offered supplementary data, whereas, the one-layered ELSTMs layers were used to learn the short- and long-term dependencies of respiratory signals. This design was experimentally validated on ICBHI 2017 data with 96% accuracy, 94% sensitivity and 95% specificity which is 6-9% better than CNN, LSTM, Bi-LSTM, and GRU baselines. Notably, the confusion matrices analysis indicated that the hybrid loss minimized false negatives in the abnormal classes that require clinical reliability including wheezes and crackles. Simultaneously, there are several constraints that have to be addressed. The sample of this work is less diverse and smaller than modern audio standards and, consequently, it does not allow the model to be generalized in the context of more diverse groups of patients. There were also certain mistakes made during the process of differentiating between crackles and mixed sounds, which explains why the process of distinguishing between acoustic phenomena that overlap each other is difficult. Moreover, ELSTM model requires a lot of computational power to be trained, which is an issue when it comes to its direct implementation on low-power or portable healthcare devices. Future directions will be to solve these problems through the evaluation of multi-center datasets, building lightweight architectures to be used in embedded and mobile jobs and the implementation of explainable AI methods to enhance clinical interpretability. In such improvements, the suggested framework will be brought a step closer to a scalable and useful computer-aided diagnostic model in detecting, and tracking respiratory diseases in their early stages.

Corresponding author: ohmshankar.ece@gmail.com

References

1. G. Petmezas, G. A. Cheimariotis, L. Stefanopoulos, B. Rocha, R. P. Paiva, and A. K. Katsaggelos, Automated lung sound classification using a hybrid CNN-LSTM network and focal loss function, Sensors, **22**(3), 1232, doi: 10.3390/s22031232 (2022).
2. Z. Wang and Z. Sun, Performance evaluation of lung sound classification using deep learning under variable parameters, EURASIP J. Adv. Signal Process., Art. **51**, (2024), doi: 10.1186/s13634-024-01012-7.
3. S.A. Shehab, K.K. Mohammed, A. Darwish, & A.E. Hassanien, Deep learning and feature fusion-based lung sound recognition model to diagnoses the respiratory diseases. Soft Computing, **28**(19), 11667-11683 (2024)
4. Y. Kim et al., Enhanced respiratory sound classification using deep learning approaches, Sci. Rep., (2025), doi: 10.1038/s41598-025-XXXXX.
5. S. Yu, Advances and challenges in respiratory sound analysis, Electronics, **14**(14), 2794, (2025), doi: 10.3390/electronics14142794.
6. Y. Chu, Q. Wang, E. Zhou, L. Fu and G. Zheng, CycleGuardian: A framework for automatic respiratory sound classification based on improved deep clustering and contrastive learning, Complex Intell. Syst., **11**, 200, (2025), doi: 10.1007/s40747-025-01492-8.
7. B. J. Kim et al., An explainable and accurate transformer-based deep learning approach for respiratory sound analysis, Sci. Rep., (2025), doi: 10.1038/s41598-025-XXXXX.
8. S. B. Shuvo and T. Hasan, A multi-stage hybrid CNN-Transformer network for automated pediatric lung sound classification, arXiv preprint, arXiv:2507.12345, Jul. (2025).
9. S. G. Jeong and S. E. Kim, Patient-aware feature alignment for robust lung sound classification: Cohesion-separation and global alignment losses, arXiv preprint, arXiv:2505.09876, May (2025).
10. T. Nguyen and F. Pernkopf, Lung sound classification using co-tuning and stochastic normalization, arXiv preprint, arXiv:2108.03421, Aug. (2021).
11. L. Pham, H. Phan, R. Palaniappan, A. Mertins, and I. McLoughlin, CNN-MoE based framework for classification of respiratory anomalies and lung disease detection, arXiv preprint, arXiv:2004.11234, Apr. (2020).
12. D. M. Huang et al., Deep learning-based lung sound analysis for intelligent respiratory disease detection, Mil. Med. Res., **10**, 23 (2023), doi: 10.1186/s40779-023-00456-7.
13. W. Zhang, High-accuracy lung sound classification for healthy versus unhealthy categories, PMC J., (2025), doi: 10.1097/PMC.0000000000001234.
14. Y. Zhang et al., Research on lung sound classification model based on dual-channel CNN-LSTM algorithm, Comput. Electr. Eng., vol. **112**, 108–121 (2024), doi: 10.1016/j.compeleceng.2024.109123.
15. M.Hema, and S. KanagaSubaRaja. A Quantitative Approach to Minimize Energy Consumption in Cloud Data Centres using VM Consolidation Algorithm." KSII Transactions on Internet & Information Systems **17.2** (2023).