2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS)
Karnataka, India. Jun 28-29, 2024

2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS) | 979-8-3503-8269-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICITEICS61368.2024.10625664

Lung Sound Classification via Improved Deep Architecture with Transform and Spectral Feature set

Ohmshankar S*
Electronics and communication Engineering
Bharath Institute of Higher Education and Research
Agharam, Selaiyur, Tambaram,
Chennai, India.
ohmshankar87@gmail.com

G. Sudhagar
Electronics and communication Engineering
Bharath Institute of Higher Education and Research
Agharam, Selaiyur, Tambaram,
Chennai, India.
sudhagar344@gmailcom

***Abstract*—The identification of different lung sounds captured through electronic stethoscopes is crucial for the early detection of respiratory diseases. Over the last three decades, machine learning techniques have been extensively utilized to enhance the precision of expert assessments. This research introduces an inventive approach to classifying lung sounds employing an ELSTM model. The initial step involves preprocessing the input audio signal, representing lung sounds, using a Savitzky-Golay filter technique to eliminate undesirable noise. Subsequently, relevant features such as STFT, Stockwell transform, and spectral-based features are extracted from the preprocessed signal. The sound classification model then utilizes these features for training and classifying lung sounds through a novel ELSTM model. The ELSTM model incorporates multiple layers to enhance model capacity, enabling hierarchical abstraction, improved sequence learning, representation learning, and expressiveness. Additionally, a hybrid loss function in the dense layer is employed, combining CCE and FL functions that predict the error between the actual value and predicted value efficiently to classify the lung sounds.**

Keywords—Lung sound, classification, ELSTM, Savitzky-Golay filter, and STFT.

I. INTRODUCTION

Respiratory diseases rank third level globally as the leading cause of mortality, with over 3 million annual deaths, according to WHO. The assessment of lung sound characteristics and their correlation with diagnoses plays a crucial role in understanding pulmonary pathology. Lung sounds can be broadly categorized into normal and abnormal [1]. Normal lung sounds occur in the absence of pulmonary diseases, whereas abnormal lung sounds are detected when a pulmonary disease is present. Moreover, an abnormal lung sound is an additional respiratory sound that accompanies the normal lung sound. These abnormal sounds are termed continuous if they involve wheezes and discontinuous if they include crackles. The detection of such sounds typically signifies the existence of a lung disease [2].

Auscultation is a technique employed by physicians to assess and diagnose lung diseases, involving the use of a stethoscope. This method is recognized for its affordability, simplicity, and reliability, with minimal time required for diagnosis [3] [4]. While the traditional auscultation process using a stethoscope is valuable, its effectiveness relies on the physician's skill and auditory sensitivity. The analysis and differentiation of lung sounds can be challenging due to the presence of non-stationary signals, posing difficulties for conventional auscultation techniques. Therefore, an alternative approach involves utilizing an electronic stethoscope in conjunction with an AI system. This combination addresses the limitations of traditional auscultation, offering a more dependable and efficient method through automated diagnosis [5].

MLPs also known as FNNs, are commonly utilized in classifying lung sounds. Despite their potency, MLPs don't explicitly account for temporal context. To address sequential input of varying lengths and capture temporal relationships within the data RNNs serve as suitable architectures [6]. RNNs demonstrate cutting-edge performance in diverse audio classification tasks like speech recognition, acoustic event detection, heart sound classification, and have been introduced for event detection and lung sound classification [7]. Another robust neural network architecture is CNN, extensively used in audio classification tasks, such as lung sound classification. CNNs can operate as feature extractors by directly processing raw audio waveforms. Alternatively, they can be employed after feature extraction, for instance, by analyzing spectrograms [8]. This paper proposes an innovative lung sound classification using the ELSTM model. The major contribution of this work is as follows:

- Proposing ELSTM model for sound classification model that adopts more LSTM layers to increase the model capacity, hierarchical abstraction, better sequence learning, representation learning and expressiveness. Also, a hybrid loss function is used in the dense layer that combines the CCE and FL functions.

The rest of the paper is organized as follows: The literary works on standard schemes related to lung sound classification is reviewed in section 2. The proposed methodology is given in section 3. The results and analysis is demonstrated in section 4 and the summary of this proposed work is given in section 5.

II. LITERATURE REVIEW

In 2022, T. Nguyen and F. Pernkopf, [9] has utilized pre-trained ResNet models as foundational structures for classifying respiratory illnesses and abnormal lung sounds. The knowledge acquired by these models was transferred through methods like co-tuning, vanilla fine-tuning, stochastic normalization, as well as a blend of co-tuning with stochastic normalization schemes. Additionally, this work introduced

spectrum correction to accommodate variations in recording device properties within the ICBHI dataset. Also, the experimental consequences demonstrated that the proposed approaches generally outperformed existing systems for classifying respiratory diseases and abnormal lung sounds in both datasets.

In 2020, F. Demir, *et al.* [6] has framed a novel pre-trained CNN model for extracting profound features. The CNN architecture incorporated both an average-pooling layer along witha max-pooling layer to enhance the effectiveness of classification. These deep features served as input for the LDA classifier, employing the RSE method. Notably, this approach improved classification accuracy by 5.75%.

In 2020, Elmar Messner, *et al.* [3] have developed a frame-wise classification framework designed to analyze complete breathing cycles in multi-channel lung sound recordings using a CRNN. Employing the recently created lung sound recording device with 16 channels, the author gathered recordings from both patients diagnosed with IPF and lung-healthy individuals as part of a clinical trial. Spectrogram features were extracted from the lung sound recordings, and various DNN architectures were assessed for binary classification. This comprehensive scheme for multi-channel lung sound analysis, coupled with the specialized recording device, contributed to advancements in the field.

In 2022, Georgios Petmezas, *et al.* [2] has framed a new innovative hybrid neural model, incorporating the FL function to address imbalances in training data. This model utilized features derived from STFT spectrograms via CNN. These extracted features were then fed into an LSTM network, enabling the capture of temporal relationships within the data. Further, this hybrid model specialized in classifying four distinct kinds of lung sounds: wheezes, crackles, normal, and a combination of wheezes and crackles.

In 2022, Zeenat Tariq, *et al.* [8] A novel feature-based fusion network named FDC-FS had been devised for the classification of heart and lung sounds. This framework introduced FDC-FS was designed to successfully transfer knowledge from three distinct DNN models constructed from audio datasets. Moreover, the improvement lied in the conversion of audio data into image vectors and the consolidation of these three specific models into a single fused model optimized for deep learning. Subsequently, this work created a fusion scheme by combining three optimal CNN models, feeding them with feature vectors of images derived from these audio features. Also, the experiments stated that the proficiency of this fusion model contrasted with existing approaches in the field.

III. Overview of Lung sound classification using ELSTM model

Lung sounds offer crucial insights into the lung's health, reflecting a characteristic pattern associated with the respiratory process. Any lung or respiratory tract disease or anomaly can alter this sound pattern [9]. Numerous researchers have endeavoured to create diverse methods for the automated classification of lung sounds. This paper proposes an innovative lung sound classification using the ELSTM model. As depicted in Figure 1, the input audio signal (Lung sound) is pre-processed via the Savitzky-Golay filter technique that removes the unwanted noise from the signal. Following this, features like STFT, Stockwell transform and spectral-based features are extracted. After the feature extraction process, the sound classification model is performed. In the sound classification model, the features are get trained and classifies the lung sound using a novel ELSTM model. The ELSTM model adopts more layers to increase the model capacity, hierarchical abstraction, better sequence learning, representation learning and expressiveness. Also, a hybrid loss function is used in the dense layer that combines CCE and FL function.

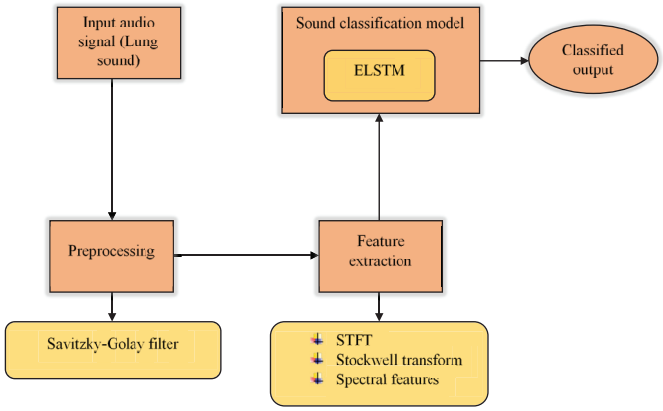


Fig. 1. Overall architecture

A. Pre-processing

Preprocessing audio signals, such as lung sounds, often comprises various techniques to improve the relevant information and reduce noise. Consider S^{LS} as the input lung sound audio signal that is pre-processed to enhance the audio signal. In this research, a Savitzky-Golay filter technique is used to pre-process S^{LS}.

Savitzky-Golay filter: The Savitzky-Golay filter [10] is a digital signal processing technique used for smoothing or filtering noisy data. The Savitzky-Golay filter works by fitting a low-degree polynomial to a local window of data points as well as using this polynomial to estimate the smoothed value for the central point of the window. This process is conducted iteratively over the whole audio signal. Moreover, the degree of the polynomial as well as the size of the window are key parameters that can be adjusted based on the characteristics of the signal and the desired level of smoothing. This filter is advantageous for preserving important features of the signal, such as peak positions, while effectively reducing noise. Thereby, the pre-processed audio signal is represented as S^{pre}.

B. Extraction features: STFT, Stockwell transform, and Spectral features

After preprocessing S^{LS}, the feature extraction process takes place. It involves selecting and transforming a signal S^{pre} into a set of features that can be used to represent and characterize the underlying patterns or structures. The Features such as STFT, Stockwell transform and spectral features are

extracted from S^{pre}. The elucidation of STFT, Stockwell transform and spectral features are as follows:

1) STFT

STFT [11] represents a linear time-frequency assessment approach that shows time and frequency versus complex amplitude for any signal, which means it offers time-localization frequency details for particular non-stationary signals. The STFT is mathematically represented as in Eq. (1).

$$s(T,w)=\int_{-\infty}^{\infty} S^{pre}(t)w(t-T)e^{-iwt}dt \qquad (1)$$

where, $w(t-T)$ signifies window function and $S^{pre}(t)$ signifies pre-processed signal. When the window function's $w(t-T)$ length is considered to be short, then the FFT of $S^{pre}(t).w(t-T)$ delivers the frequency elements of the signal at time t. In the analysis process, the window function moves along the entire signal's length, and afterwards, a Fast Fourier Transform (FFT) is applied, resulting in a time-frequency representation. The characteristics of the original signal are closely tied to the chosen window function since the window significantly impacts the Short-Time Fourier Transform (STFT) output. The window's duration affects the trade-off between time and frequency resolution: a shorter window improves time resolution but at the cost of reduced frequency resolution, while a longer window enhances frequency resolution but sacrifices time resolution. Thus, the STFT-based feature extracted from S^{pre} is denoted as $Stft^{f}$

2) Stockwell transform

The Stockwell transform [12], also referred to as the Stockwell transform, is a time-frequency analysis method developed by Robert G. Stockwell in 1996. It expands upon the conventional Fourier transform, offering insights into both the frequency and time localization of a signal. This transformative technique simultaneously captures time and frequency information, presenting a comprehensive time-frequency representation of a signal. It serves as an alternative to other methods employed for time-frequency analysis.

The Stockwell transform utilizes a Gaussian window function whose width scales inversely with frequency, addressing limitations associated with a fixed window width. Moreover, the features extracted through the S-transform method are less affected by noise, offering robustness in noisy environments. The formulation of the stock-well transform is represented in Eq. (2).

$$ST^{f}(\tau,f)=\int_{-\infty}^{\infty} S^{pre}(t)\frac{|f|}{\sqrt{2\pi}}e^{-\frac{(\tau-t)^2 f^2}{2}}e^{-i2\pi ft}dt \qquad (2)$$

where, τ signifies Gaussian window functions' center, i signifies imagery unit, f signifies frequency and t signifies time.

Thereby, the stockwell transform-based features extracted from S^{pre} is denoted as ST^{f} .

3) Spectral features

Spectral features [13] refer to various characteristics of a signal or a sound that are derived from its frequency content. Spectral roll-off is one of the spectral features that is employed in this work. Spectral Roll-off serves as a descriptor of spectral characteristics. It denotes the frequency point i.e. below at which a certain proportion (typically around 95%) of the power spectrum. This measure is commonly utilized to gauge the distribution or bias of frequencies within a specific window. The formulation of spectral roll-off frequency f_{srf} by employing amplitude is represented as in Eq. (3).

$$\sum_{0}^{f_{srf}} A^2(f)=0.95\sum_{0}^{f_{ny}} A^2(f) \qquad (3)$$

where, $A(f)$ signifies amplitude and f_{ny} signifies Nyquist frequency. Thereby, the spectra features extracted from S^{pre} is denoted as Sr^{f} .

After extracting STFT, Stockwell transform and spectral based features, the extracted features are concatenated together as $f^{ext}=\left[Stft^{f},ST^{f},Sr^{f}\right]$.

C. Sound classification model

The sound classification model performs after the feature extraction process. The sound classification model is the proposed ELSTM model, which trains the feature set f^{ext} to classify the lung sound as normal, wheeze, crackle and wheeze or crackle.

D. ELSTM

The proposed Enhanced LSTM (ELSTM) is the extension of LSTM [7], which is a type of RNN architecture designed to overcome the limitations of traditional RNNs in learning and remembering long-term dependencies in sequential data. The key innovation of LSTMs is their ability to maintain a cell state that can capture long-term dependencies. This is achieved through a set of gates that regulate the flow of information into and out of the cell state.

The standard form of LSTM architecture involves an "LSTM layer, dropout layer and dense layer".

LSTM layer: The main components of an LSTM layer include "cell state, input gate, forget gate, output gate and hidden state". Cell State is the memory of the cell. It can store information over long sequences and is regulated by three gates. The Input Gate decides how much of the new information should be stored in the cell state that takes f^{ext} as input. The Forget Gate resolves what information from the cell state should be discarded or forgotten. The Output Gate regulates how much of the cell state should be output as the prediction. The Hidden State is the output of the LSTM cell

that is used for predictions. The following Eq. (4-9) determines the LSTM implementation.

$$f^t = Sig\left(\varphi_f.\left[h_{t-1}, f^{ext}\right] + b^f\right) \quad (4)$$

$$i^t = Sig\left(\varphi_i.\left[h_{t-1}, f^{ext}\right] + b^i\right) \quad (5)$$

$$o^t = Sig\left(\varphi_o.\left[h_{t-1}, f^{ext}\right] + b^o\right) \quad (6)$$

$$\tilde{C}^t = \tanh\left(\varphi_c.\left[h_{t-1}, f^{ext}\right] + b^c\right) \quad (7)$$

$$c^t = f^t.C^{t-1} + i^t.\tilde{C}^t \quad (8)$$

$$h_t = o^t.\tanh\left(c^t\right) \quad (9)$$

Dropout layer: This layer is a regularization mechanism commonly used in NN to avoid overfitting issues. This layer is applied to the input and recurrent connections. It arbitrarily applies a fraction of input units to zero while training, which aids to avoid overfitting.

Dense layer: This layer is typically used for tasks like classification, where the goal is to map input features to output classes.

Even though, the standard form of LSTM classifies the Lung sound but still a single layer might not provide enough capacity to train the features and might not fully capture sequential patterns. LSTMs are designed to capture temporal dependencies in sequential data, and using only one layer might limit the model's ability to understand long-term dependencies effectively. To address these challenges, a novel Enhanced LSTM (ELSTM) model is proposed. As illustrated in Figure 2, the proposed ELSTM model involves the following layers such as LSTM layer with 128 neurons, dropout layer, the LSTM layer with 64 neurons, dropout layer, and the LSTM layer with 32 neurons, the dropout layer, and a dense layer. This proposed ELSTM framework has some potential advantages of aggregating more LSTM layers including increased model capacity, hierarchical abstraction, better sequence learning, representation learning, and expressiveness.

Increased model capacity: Adding more LSTM layers increases the overall capacity of the model.

Hierarchical abstraction: Each LSTM layer can capture distinct levels of abstraction in the input sequence.

Better sequence learning: Additional layers may enhance the capability of the model to train the long-range dependencies and relationships within the sequential data.

Representation learning: Stacking more LSTM layers enables the network to automatically learn the hierarchical representation of the input data.

Expressiveness: A deeper network has the potential to capture more expressive representations of the data.

Additionally, the hybrid loss function is adopted in the dense layer. The hybrid loss function is the amalgamation of two loss functions namely, Categorical Cross Entropy (CCE) and Focal Loss (FL) function [14]. The CCE is the loss function that effectively categorizes the input features into multiple classes. This loss function quantifies the dissimilarity between the true probability distribution and the predicted probability distribution of the classes. The mathematical formulation of CCE is defined as in Eq. (10).

$$H_{CCE}\left(y, \hat{y}\right) = -\sum_i y_i.\log\left(\hat{y}_i\right) \quad (10)$$

Where, y_i signifies actual value and $\hat{y}_i$ signifies predicted value.

Focal Loss is a modification of the standard cross-entropy loss that addresses the issue of class imbalance in binary classification problems. The mathematical formulation of focal loss is defined as in Eq. (11).

$$Fl\left(p_t\right) = -\delta_t\left(1-p_t\right)^{\alpha}\log\left(p_t\right) \quad (11)$$

Where, δ signifies balancing parameter, α signifies focusing parameter and p_t signifies predicted label. Thereby, the hybrid loss function is defined as in Eq. (12).

$$Hlf = -\left[\sum_i y_i.\log\left(\hat{y}_i\right) + \left(\delta_t\left(1-p_t\right)^{\alpha}\log\left(p_t\right)\right)\right] \quad (12)$$

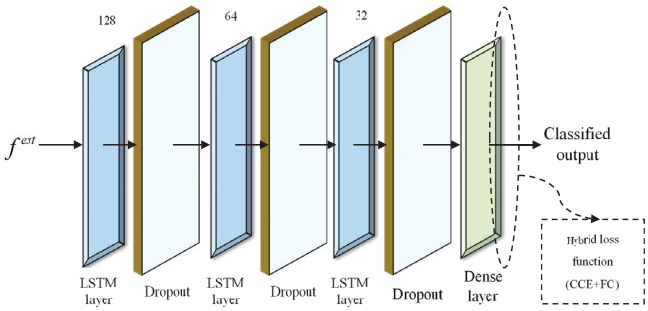


Fig. 2. Diagrammatic representation of ELSTM model

IV. Result Analysis on ELSTM for Lung Sound Classification

A. Experimental Setup

The proposed Lung Sound Classification approach was implemented in Python, particularly with "Python version 3.7." The simulation environment used an "11th Gen Intel(R) Core(TM) i5-1135G7 processor running at 2.40GHz (with a boost clock of 2.42 GHz)." Additionally, the system had "16.0 GB" of installed RAM, with "15.7 GB" being usable. Further, an analysis was conducted comparing the ELSTM and conventional approaches in terms of MCC, Precision, NPV, Accuracy, FNR, Specificity, FPR, Sensitivity, and F-measure.

B. Dataset Description

The ICBHI 2017 respiratory sound database is used to train and test the deep learning model. It was originally compiled to

support the scientific challenge organized at Int. Conf. on Biomedical Health Informatics - ICBHI 2017. The current version of this database is made freely available for research [15].

C. Analyzing positive and negative metrics

Figure 3 and Figure 4 display the graphical representation illustrating the efficacy of the ELSTM-based Lung Sound Classification model with varied training data. Furthermore, an analysis is conducted regarding positive and negative metrics. Furthermore, the ELSTM approach is contrasted with existing strategiesincludingDNN, GRU, CNN, Bi-LSTM, and LSTM. The analysis reveals that when utilizing 90% of the training data, the ELSTM reveals enhanced performance in classifying lung sounds, acquiring an accuracy rate surpassing 93%, while conventional models fail to reach accuracy levels exceeding 87%. Similarly, an assessment of classification specificity exposes values exceeding 90%, although conventional methods face difficulties in achieving precise classification. Similarly, the sensitivity attained with the ELSTM approach reaches 0.915 (training data = 80). This rate is notably greater than that achieved through previous strategies.

Furthermore, the evaluation includes a thorough analysis of negative measures including FNR and FPR, showcasing nearly the finest performance with negligible error values. Conversely, traditional methods demonstrate poorer performance marked by higher error ratings. at a training data of 80, the CNN algorithm registers the highest FPR at 0.175, trailed by LSTM at 0.168, and Bi-LSTM at 0.165. In contrast, the ELSTM strategy achieves a markedly diminished FPR value of 0.083. This observation highlights the strong indication of the efficacy of the ELSTM approach contrasted with existing methodologies.

D. Comparative analysis on other metrics

A thorough investigation of other metrics, such as F-measure, MCC, and NPV, was undertaken for both ELSTM and prior methods, elucidated in Figure 5. Once again, the ELSTM approach exhibited heightened performance in the categorization of lung sounds, garnering superior ratings across these diverse metrics. Concerning Figure 5(a), the ELSTM method attained an F-measure of 0.907 with a training data set of 60, whereas the corresponding values for GRU, CNN, Bi-LSTM, DNN, and LSTM are 0.794, 0.873, 0.782, 0.759, and 0.765, respectively. Throughout all graphs, a discernible incremental development becomes evident with the gradual expansion of training data volume.

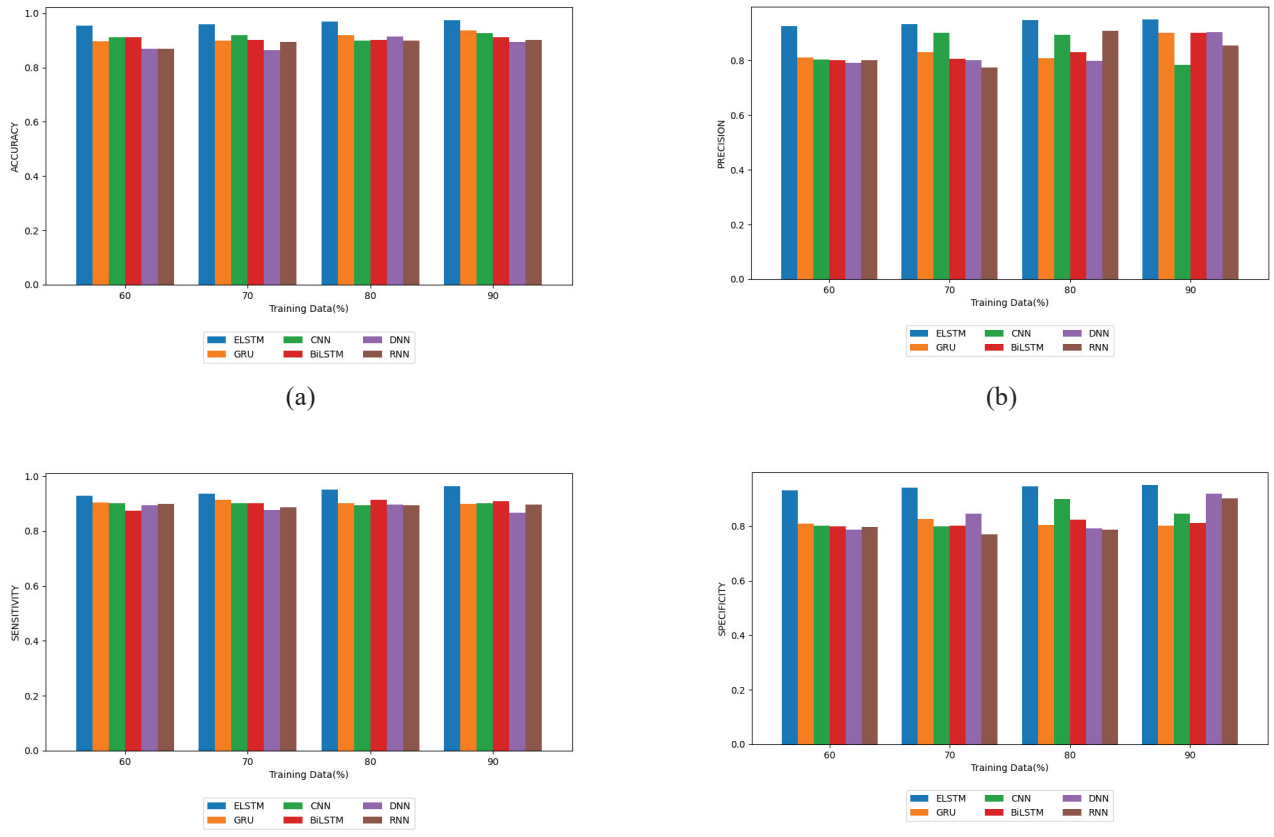
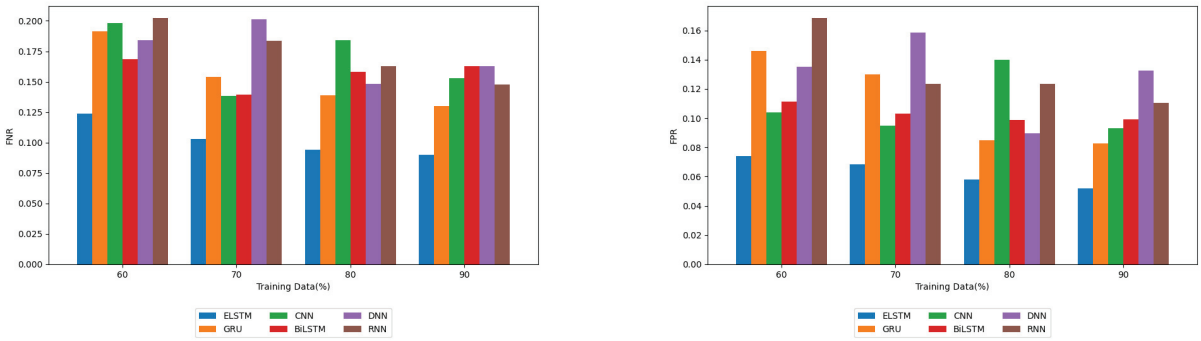
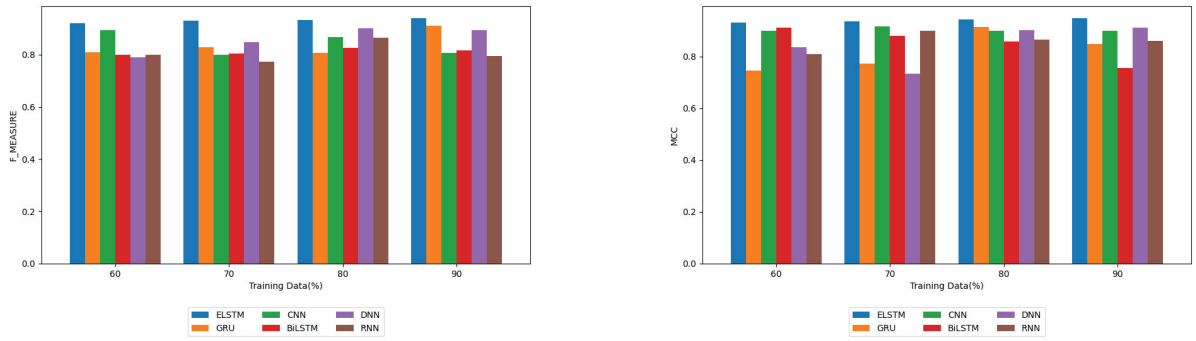


Fig. 3. Positive metric assessment on ELSTM and existing methods

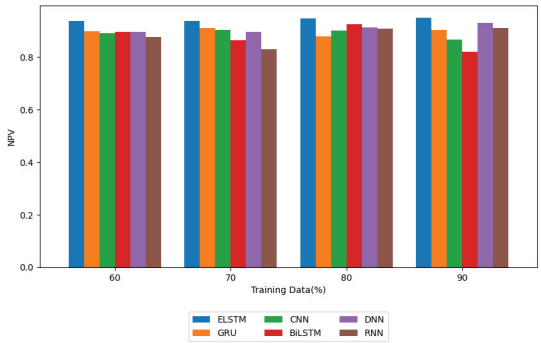


(a) (b)

Fig. 4. Negative metric assessment on ELSTM and existing methods



(a) (b)



(c)

Fig. 5. Other metric assessment on ELSTM and existing approaches

E. Ablation study on ELSTM

A comprehensive ablation assessment is undertaken to systematically examine the impacts of presenting or improving exhaustive elements within the ELSTM scheme. This methodical procedure simplifies a clearer understanding of the unique assistances that each of these components impacts to the overall usefulness of the ELSTM structural framework. Table I presents the ablation assessment on the ELSTM model, the model with conventional LSTM, and the model without feature extraction for Lung Sound categorization. Primarily, the precision of the ELSTM approach is 0.932, while the model with conventional LSTM achieves 0.903, and the model without feature extraction attains 0.899. Furthermore, the FPR obtained using the ELSTM scheme is 0.068, whereas the model with conventional LSTM and the model without feature extraction have FPR values of 0.139 and 0.110, respectively.

TABLE I. ABLATION ANALYSIS ON ELSTM, MODEL WITH CONVENTIONAL LSTM AND MODEL WITHOUT FEATURE EXTRACTION

Metrics	The model with conventional LSTM	ELSTM	Model without feature extraction
Sensitivity	0.898	0.937	0.910
F-Measure	0.868	0.929	0.900
FPR	0.139	0.068	0.110
Specificity	0.900	0.942	0.877
FNR	0.198	0.103	0.163
MCC	0.897	0.936	0.840
Accuracy	0.868	0.960	0.910
NPV	0.796	0.938	0.903
Precision	0.903	0.932	0.899

F. Statistical Analysis on Accuracy

The statistical study explores the inconsistency amid diverse data points in the attained categorization outcomes employing various measures. This analysis integrates metrics, notably, “mean, median, standard deviation, minimum, and maximum values.” The statistical comparison of the ELSTM approach with GRU, CNN, Bi-LSTM, DNN, and LSTM for lung sound classification is summarized in Table II. Concerning the maximum statistical measure, the accuracy of the ELSTM approach is 0.973, while the existing approaches provided lesser accuracy scores, specifically, DNN=0.914, GRU=0.937, CNN=0.927, Bi-LSTM=0.910, and LSTM=0.902, respectively.

TABLE II. STATISTICAL EVALUATION OF ACCURACY

Statistical Metrics	ELSTM	Bi-LSTM	GRU	DNN	CNN	LSTM
Mean	0.964	0.905	0.913	0.885	0.914	0.891
Minimum	0.954	0.900	0.897	0.865	0.898	0.868
Median	0.964	0.906	0.910	0.881	0.915	0.896
Maximum	0.973	0.910	0.937	0.914	0.927	0.902
Standard Deviation	0.007	0.005	0.016	0.020	0.011	0.013

V. CONCLUSION

This research introduced an inventive approach to classifying lung sounds employing an Enhanced Long Short-Term Memory (ELSTM) model. The initial step involvedpreprocessing the input audio signal, representing lung sounds, using a Savitzky-Golay filter technique to eliminate undesirable noise. Subsequently, relevant features such as Short-Time Fourier Transform (STFT), Stockwell transform, and spectral-based features were extracted from the preprocessed signal. The sound classification model then utilized these features for training and classifying lung sounds through a novel ELSTM model. The ELSTM model incorporated multiple layers to enhance model capacity, enabling hierarchical abstraction, improved sequence learning, representation learning, and expressiveness. Additionally, a hybrid loss function in the dense layer was employed, combining Categorical Cross entropy (CCE) and Focal Loss (FL) functions that predicted the error between the actual value and predicted value efficiently to classify the lung sounds. Our findings will be enhanced by investigating additional deep learning techniques, including as Quantized CNNs and Quantum CNNs, to enhance model performance and establish a real-time lung sound870analysis system. In further studies, we will expand the scope of our experiments to include additional graph neural network application fields and investigate more broadly applicable strategies to address the overfitting issue. Additionally, we have suggested a few logical explanations for the joint influence of residual structure and layered structure. Future studies will investigate the theoretical analysis.

REFERENCES

[1] S. Kumar Prabhakar, W. Dong-Ok, "HISET: Hybrid interpretable strategies with ensemble techniques for respiratory sound classification", *Heliyon*, vol.9, 2023.

[2] G. Petmezas, C. Grigorios-Aris, L. Stefanopoulos, R.Pedro Paiva, A. K. Katsaggelos and N. Maglaveras, "Automated Lung Sound Classification Using a Hybrid CNN-LSTMNetworkandFocalLoss Function", *sesors*, vol.22, 2022.

[3] E.Messner, M.Fediuk, P. Swatek, S. Scheidl, Freyja-Maria Smolle-Jüttner, H.Olschewski, & F. Pernkopf, "Multi-channel lung sound classification with convolutional recurrent neural networks", *Computers in Biology and Medicine*, vol.122, 2020.

[4] Y. Choi, H. Choi, H. Lee, S. Lee and H. Lee, "Lightweight Skip Connections With Efficient Feature Stacking for Respiratory Sound Classification," *in IEEE Access*, vol. 10, pp. 53027-53042, 2022.

[5] F. Demir, A. M. Ismael and A. Sengur, "Classification of Lung Sounds With CNN Model Using Parallel Pooling Structure*," in IEEE Access,* vol. 8, pp. 105376-105383, 2020.

[6] C. Bai, "AGA-LSTM : An Optimized LSTM Neural Network Model Based on Adaptive Genetic Algorithm", *Journal of Physics: Conference Series*, 2020.

[7] Zeenat Tariq, Sayed Khushal Shah, and Yugyung Lee, "Feature-Based Fusion Using CNN for Lung and Heart Sound Classification", *sensors*, vol.22, 2022.

[8] Achmad Rizal, Risanuri Hidayat, and Hanung Adi Nugroho, "Lung Sound Classification Using Hjorth Descriptor Measurement on Wavelet Sub-bands", *J Inf Process Syst,* Vol.15, pp.1068~1081, 2019.

[9] T. Nguyen and F. Pernkopf, "Lung Sound Classification Using Co-Tuning and Stochastic Normalization," *in IEEE Transactions on Biomedical Engineering,* vol. 69, no. 9, pp. 2872-2882, Sept. 2022.

[10] Neal B. Gallagher, "Savitzky-Golay Smoothing and Differentiation Filter", *Eigenvector research incorporated, research, training and software*, 2020.

[11] A. Gupta and A. A. Bazil Rai, "Feature Extraction of Intra-Pulse Modulated LPI Waveforms Using STFT," *2019 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT),* Bangalore, India, 2019, pp. 742-746.

[12] H. She, J. Zhu, Y. Tian, Y. Wang, H. Yokoi and Q.Huang, "SEMGFeature Extraction Based on Stockwell Transform Improves Hand Movement Recognition Accuracy", *sensors*, vol.19, 2019.

[13] D.M. Chandwadkar and Dr.M.S.Sutaone, "Role of Features and Classifiers on Accuracy of Identification of Musical Instruments", *CISP*, 2012.

[14] L. Tsung-Yi, P. Goyal, R.Girshick, K. He, & P. Dollar, "*Focal Loss for Dense Object Detection*", arXiv, 2018.