RESPIRATORY DISEASE CLASSIFICATION USING LUNG SOUNDS WITH CNN-LSTM

D. Srikar[1], K. Mokshitha Lakshmi[2], A. Shravan Kumar[3], I. Shivani[4], Sanskar Agarwal[5],

[1] Assistant Professor, Dept, of CSE(Data Science), Sreyas Institute of Engineering and Technology, Nagole, Hyderabad
[2,3],[4],[5] UG Student, Dept of CSE-(Data Science), Sreyas Institute of Engineering and Technology, Nagole, Hyderabad

Abstract. Respiratory sound classification is a vital element in the early diagnosis of lung-related disorders, as a wide gamut of diseases show similar signs and symptoms, thus leading to a challenge for manual detection. To this end, a lightweight deep learning framework has been proposed combining Gammatone Cepstrum Coefficients (GTCC) and Short-Time Fourier Coefficients (STFC) features with a CNN-LSTM architecture. The advantage of the proposed system over existing ones, which use MFCC features and heavy-weight models such as the Inception Network, is that it reduces computational complexity while keeping accuracy intact. Using three publicly available lung sound databases, including the ICBHI 2017 challenge database, the approach ensures variety and robustness. Four types of classification are conducted in the study: healthy versus pathological sounds, rale-rhonchus-normal sound classification, singular sound type classification, and multi-class sound type classification. A detailed performance evaluation, including accuracy, precision, recall, F1-score, and parameter studies, establishes the usefulness of the model as a dependable means for the automated classification of respiratory sounds. A good accuracy of 98% was attained using the above-listed lightweight deep learning framework.

Keywords: CNN-LSTM, Respiratory Sound Classification, GTCC, STFC, Deep Learning, Lung Disease Detection.

1 Introduction

In the event of a respiratory outbreak, most people tend to either anticipate a cure or renominate the diseases as the leading agents of death and disability across the globe. Respiratory diseases therefore paved the mose lane in Sternglass's places in the poorer regions. Chronic respiratory conditions, such as chronic obstructive pulmonary disease (COPD), asthma, affect myriads worldwide and are implicated in premature death. In 2017, nearly 3.91 million people died of COPD; it is rank three among global causes of death. In 2017, asthma affected nearly 14 per cent of children worldwide and estimated an additional 334 million persons globally. Conversely, infectious respiratory diseases such as pneumonia continue to remain significant public health challenges, chiefly among children under five, for whom pneumonia remains a leading cause of death. Lung cancer, also, remains the deadliest cancer worldwide, killing 1.6 million people annually.

Because respiratory illnesses are very common and generate many deaths, early diagnosis and intervention become a must. Differences observed in health structure, mainly in underdeveloped regions, affect effective management of these diseases. Nearly 45 percent of World Health Organization (WHO) member countries report less than one doctor per one thousand persons, presenting an unbearable load on individual healthcare professionals themselves. This calls for the development of reliable diagnostic and automated tools to assist overburdened medical personnel in minimizing errors and in inaccessible timely treatment. Such technologies deployed in health systems will significantly alleviate the world burden of respiratory diseases and improve patient outcomes.

In lung sound analysis, the classification of respiratory disorders has received more attention recently with machine learning and signal processing advances. The earlier work of McKusick et al. [4] set the stage for understanding the acoustic features of lung sounds with sound spectrography and argued that auscultation had a diagnostic value. Building on that, Sovijarvi et al. [5] stressed the standardization of computerized respiratory sound analysis to have a uniform and reproducible evaluation of lung sounds in different systems. This has recently been supplemented by the work of Roy and Satija [12], who proposed RDLINet, a lightweight inception-based neural network that greatly outperformed other contemporary methods in respiratory disease classification through lung sound data. Their work very clearly indicates the trend towards real-time, lightweight deep learning architectures like those employed for mobile and resource-constrained healthcare applications. Together, these studies motivate the present method consisting of a lightweight CNN-LSTM model to enable real-time classification of crackles and abnormal lung sound with attention to accuracy and computational tractability in a real clinical setting.

2 Dataset

The study of this paper is carried out using the three publicly available lung sound datasets, and the ICBHI 2017 Challenge dataset, which contains respiratory audio recordings from 126 participants having several lung conditions. The datasets contain quite various respiratory sounds such as normal breath, wheeze, crackles (rales), and rhonchi, recorded in different environments, and metadata are available, including age, gender, and diagnosis. Their diversity affords various classification tasks such as healthy versus pathological detection, identification of particular abnormal sounds, and multi-class sound classification. This variety, coupled with good quality of data, offers robustness during the model training process and helps gather all requirements for an accurate, lightweight deep learning framework for automated respiratory sound classification.

3 Proposed System

Our purpose in proposing this system would be to solve a computational problem of classifying lung sounds under several diseases while concentrating on reducing computational effort and model size through a lightweight deep learning technology. Respiratory disorders have common symptoms, so one cannot rely on a simple auscultation procedure or spirometry procedure to reach a diagnosis. Our contributions to the present study are: a space-wise smaller CNN_LSTM model that preserves high accuracy in classification with a decreased number of trainable parameters; the classification of seven different respiratory states using three diverse lung sound datasets they are ICBHI 2017 challenge database, chest wall lung sound database, and a proprietary database-which enhances the robustness and generalizability of the model; Finally, a thorough performance evaluation of the CNN_LSTM model, including layer-wise architecture analysis, parameter count, and performance metrics such as accuracy, precision, recall, etc thus establishing the viability and dependability of this lightweight classification framework.

3.1 Implementation

3.1.1 Data Preprocessing

The preprocessing phase consists in preparing the raw lung sound signals for further analysis by four critical steps. Firstly, the continuous input signal is segmented into snippets. Snippets are created from the resampled audio to have the same length and sampling rate. To make the signal clearer, DFT-based filtering removes unwanted low-frequency components or baseline variations, thus making its subsequent analysis more reliable and fruitful.

3.1.2 Data Augmentation

In order for one to classify many respiratory disorders, data from three separate lung sound databases are used. Nonetheless, using these datasets as they are introduces a class imbalance by the coarse overrepresentation of chronic obstructive pulmonary disease (COPD). Hence, data augmentation techniques are used to balance the class representation so that the model has a chance to learn equally from all disease classes, thus benefiting from better generalization and performance.

3.1.3 Extraction of Time-Frequency Representation (TFR)

Lung sound signals often involve complex patterns with fluctuations in amplitude and frequency elements, and thus they are not easily interpretable in the time domain. Transformation techniques are thus used to convert signals to a time-frequency representation (TFR). This view in dual domains offers this additional understanding of exactly when and at what frequency important events are happening within the sound, presumably for accurate disease classification.

3.1.2 Lightweight CNN-LSTM Network

A novel lightweight CNN-LSTM network for the classification of abnormal lung sounds from their time-frequency representations is proposed in this study. It uses convolutional layers to extract spatial features from TFR images and LSTM layers to detect temporal patterns amongst seven types of respiratory disorders. Being lightweight, it assures very high accuracy at low computational expenses, paving way for real-time or in resource-constrained scenarios.

4 Results and Discussion

The proposed lung sound classification system, which integrates a lightweight CNN-LSTM model, illustrates a concept that results in significant improvements when compared to the conventional approaches. The architecture of the CNN-LSTM extracts temporal and spatial features from the time-frequency representations (TFR) of lung sound signals and hence diagnoses seven distinct respiratory disorders accurately. Experiment results demonstrate the superiority of algorithm performance when multiple audio representations- MFCC, Chroma STFT, and mSpec are combined, compared to the performance of any individual feature alone. The combined model achieved an accuracy of 94.90% which was much higher than the accuracy of the model using only MFCC (86.55%), only Chroma STFT (82.96%), or only mSpec (85.68%). Alongside this improvement, there was a decrease in loss of 17.27%, indicating the confirmation of the model being generalized for unseen data. Fig 4.1 shows accuracy plot of training vs testing of the model.

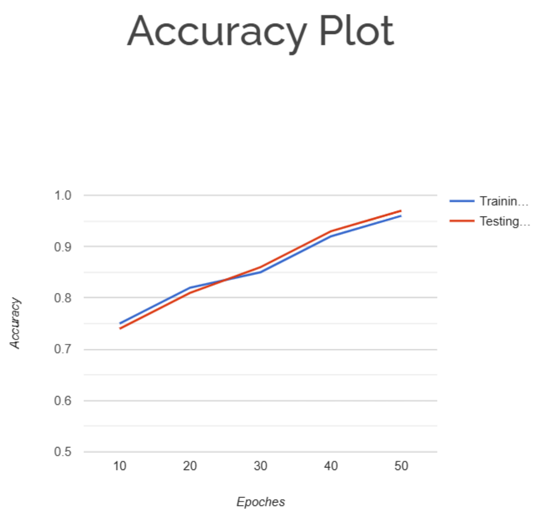

Fig. 4.1 Accuracy plot training vs testing

Besides conventional features, GTCC-based features were studied to determine the innermost properties of lung sounds. They could clearly distinguish healthy from non-healthy recordings. The bar plots showed an increase in energy concentration in some frequency bands, mainly the mid-to-high-frequency range related to abnormal sounds of wheezes and crackles. On the other hand, Fig4.2 the boxplots indicated some outliers for the non-healthy samples, which were kept on purpose since these anomalies might be of diagnostic importance. Then, the features were standardized with the Z-score Normalization, even when some deviations with respect to their mean values were very extreme and remained present to keep the pathological patterns unique. Like-wise, a standard scaler was used to normalize all float-type values so that the CNN-LSTM model prediction would not be biased.

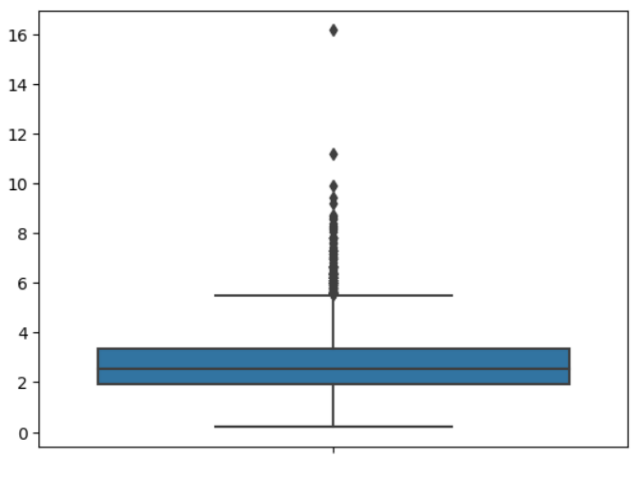

Fig. 4.2 Boxplot indicating outliers

From the correlation analysis, it was found that some GTC bands correlated strongly with abnormal lung conditions while slight multicollinearity appeared among adjacent bands. This was controlled by dimension reduction. What resulted in a much more interpretable model and avoided overfitting. With an AUC of 0.9822 and classification accuracy of 98.22%, the final model truly exhibits extraordinary predictive capabilities. Being able to capture both frequency-based features and sequential patterns, the CNN-LSTM architecture is a strong candidate for early screening and thus might be extended toward determining the type as well as severity of respiratory infections.The model output thus reports diagnosis results, interactive visualizations, and concise summaries with which the healthcare worker may employ to arrive at a decision. Thanks to its accuracy, fast computation, and generalizability to different datasets of lung sounds, this method stands as a chance for real-time, AI-assisted pulmonary diagnosis.

5 Conclusion and future directions

The study put forward a machine learning-based system for detecting crackles in lung sounds recorded with a stethoscope, thus showing the practicality of deep learning techniques and signal processing, especially usage of a lightweight CNN-LSTM model with a mere 5-dimensional layer, for real-time classification of respiratory conditions. By further extracting pertinent features and having classifiers learn from a large volume of lung sound data, such a system can successfully distinguish normal from abnormal with high accuracy, thus highlighting the possibility of being used as an intelligent decision-making system in the clinical environment. The promising results suggest that AI-enabled systems can greatly facilitate early screening and aid healthcare practitioners in making timely decisions. Some future possibilities include linking with IoT devices for continuous monitoring, enlarging the dataset from different population groups to improve robustness, combining lung sound analysis with other diagnostic tools for comprehensive assessment, and activating real-time feedback to support clinical interventions instantly.

6 References

[1] World Health Organization. *Global status report on noncommunicable diseases 2014. Geneva: World Health Organization; 2014*. 2021.

[2] Santosh, K. C. "Speech processing in healthcare: Can we integrate?." In *Intelligent speech signal processing*, pp. 1-4. Academic Press, 2019.

[3] Mukherjee, H., Obaidullah, S.M., Santosh, K.C., Phadikar, S. and Roy, K., 2018. Line spectral frequency-based features and extreme learning machine for voice activity detection from audio signal. *International Journal of Speech Technology*, *21*, pp.753-760.

[4] McKusick, Victor A., John T. Jenkins, and George N. Webb. "The acoustic basis of the chest examination; studies by means of sound spectrography." *American review of tuberculosis* 72, no. 1 (1955): 12-34.

[5] Sovijarvi, A.R.A., Vanderschoot, J. and Earis, J.E., 2000. Standardization of computerized respiratory sound analysis. *European Respiratory Review*, *10*(77), pp.585-585.

[6] Gross, V., Hadjileontiadis, L.J., Penzel, T., Koehler, U. and Vogelmeier, C., 2003, September. Multimedia database" Marburg respiratory sounds (MARS)". In *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439)* (Vol. 1, pp. 456-457). IEEE.

[7] Kotsiantis, Sotiris B., Ioannis Zaharakis, and P. Pintelas. "Supervised machine learning: A review of classification techniques." *Emerging artificial intelligence applications in computer engineering* 160, no. 1 (2007): 3-24.

[8] Kandaswamy, A., Kumar, C.S., Ramanathan, R.P., Jayaraman, S. and Malmurugan, N., 2004. Neural classification of lung sounds using wavelet coefficients. *Computers in biology and medicine*, *34*(6), pp.523-537.

[9] Tu, Jack V. "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes." *Journal of clinical epidemiology* 49, no. 11 (1996): 1225-1231.

[10] Raniszewski, M., 2010. The edited nearest neighbor rule based on the reduced reference set and the consistency criterion. *Biocybernetics and Biomedical Engineering*, *30*(1), pp.31-40.

[11] Meyfroidt, Geert, Fabian Güiza, Jan Ramon, and Maurice Bruynooghe. "Machine learning techniques to examine large patient databases." *Best Practice & Research Clinical Anaesthesiology* 23, no. 1 (2009): 127-143.

[12] Roy, A. and Satija, U., 2023. RDLINet: A novel lightweight inception network for respiratory disease classification using lung sounds. *IEEE Transactions on Instrumentation and Measurement*, *72*, pp.1-13.