Audio-Based Machine Learning for Respiratory Disease Diagnosis

NamitK, S.Venkata Sai Charith, Nitish D, Malasani Tejeswar , Dr. Debanjali Bhattacharya
Department of Computer Science and Engineering
Amrita School of Computing, Bengaluru
Amrita Vishwa Vidyapeetham, India
bl.en.u4aie23111@bl.students.amrita.edu, bl.en.u4aie23132@bl.students.amrita.edu,
bl.en.u4aie23139@bl.students.amrita.edu, bl.en.u4aie23149@bl.students.amrita.edu,
b_debanjali@blr.amrita.edu

***Abstract*—Respiratory diseases have a significant impact on global health, and early and accurate diagnosis is crucial to enable timely intervention. This paper presents a machine learning approach to respiratory disease classification from lung sound recordings of the publicly available Respiratory Sound Database on Kaggle. The raw audio data as patient lung sound recordings in WAV files were first translated into visual forms for recognizing the underlying temporal patterns. These were then processed into recurrence plots (RPs), which effectively recognize non-linear and recurring patterns in time-series data, enabling meaningful features to be extracted from complex audio signals.**

To convert these recurrence plots into a machine learnable representation, an autoencoder model was employed to generate low-dimensional latent embeddings. These embeddings are dense but informative representations of the original audio data that preserve key diagnostic features. These embeddings are then mapped to the corresponding patient diagnosis using metadata present in the patient_diagnosis.csv file for supervised learning.

Having the data represented in the latent space available, general machine learning classifiers such as Support Vector Classifier (SVC),Balanced Random Forests, and KNN Classifier can be applied for disease prediction. The performance of the model is evaluated with standard classification measures such as accuracy, precision, recall, and F1-score.

This pipeline not only reduces computational complexity by avoiding direct audio feature extraction but also leverages the strength of image-based and deep learning approaches. The results demonstrate the promise of recurrence plot-based embeddings for effective and interpretable respiratory disease classification from auscultation data.

I. Introduction

Respiratory conditions, including asthma, bronchitis, COPD, and pneumonia, are among the global leading causes of illness and death. Early identification and accurate identification are essential to effective treatment and management. Auscultation—the listening to breath sounds with a stethoscope—has long been a chief diagnostic tool by convention. However, interpretation of respiratory sounds remains subjective and significantly dependent on clinical experience, potentially leading to differing diagnoses. To overcome the above limitations, machine learning offers a solution where it is achievable to create automatically and unbiased diagnostics from respiratory sound patterns.

Here, this piece of work tackles the application of traditional machine learning techniques to classifying respiratory illness from audio record in the widely available Respiratory Sound Database in Kaggle. Instead of extracting raw audio features or using deep learning, we convert audio recordings into visual descriptions in the form of recurrence plots (RPs), which can capture complex, non-linear temporal patterns found in lung sounds. These are then projected onto low-dimensional latent embeddings using an encoder as part of dimensionality reduction techniques autoencoder.

The acquired embeddings, as informative and concise feature vectors, are projected onto patient diagnoses via the metadata provided in patient_diagnosis.csv. Traditional machine learning models like Support Vector Classifier (SVC), Balanced Random Forests, and KNN are then employed to label the diseases. The technique employs recurrence analysis and ML to provide a computationally efficient, interpretable, and transparent method for detecting respiratory disease.

II. Literature Survey

[1] **Lung disease recognition methods using audio-based analysis with machine learning:** The paper "Lung Disease Recognition Methods Using Audio-Based Analysis with Machine Learning" covers a general overview of respiratory disease diagnostic methods utilizing audio signal processing and machine learning. It identifies the potential for non-invasive methods, such as lung sound analysis, to diagnose conditions such as asthma, COPD, and pneumonia. The paper tests several techniques of feature extraction, including Mel-Frequency Cepstral Coefficients (MFCCs), spectrograms, and wavelet transforms, that prove to be highly beneficial in determining the features of respiratory sounds. Machine learning classifiers like Support Vector Machines (SVM), Random Forests, and k-Nearest Neighbors (k-NN) then apply these features to distinguish health from disease.

The paper discusses issues with audio-based diagnostics as well, which are background noise, variability of recording equipment, and the need for large amounts of annotated datasets. The importance of preprocessing techniques, such as noise reduction and normalization, to enhance the quality of input data is highlighted by the paper. Moreover, the research highlights the importance of metrics used in the evaluation of the model like accuracy, sensitivity, and specificity for the measurement of the performance of the diagnostic model.

For binary classification-based projects on respiratory health status, this paper is rich in information on effective feature extraction techniques and the application of traditional machine learning models. It also informs the reader on real-world data preprocessing and model evaluation factors, which makes it an appropriate piece of literature for designing strong audio-based diagnostic systems.

[2]**Exploring machine learning for audio-based respiratory condition screening: A concise review of databases, methods, and open issues**: The paper ”Exploring Machine Learning for Audio-Based Respiratory Condition Screening: A Concise Review of Databases, Methods, and Open Issues” by Xia et al. (2022) introduces a current picture of the situation in respiratory sound analysis using ML. The paper considers publicly available respiratory audio databases like ICBHI 2017 and Coswara pertinent to recurrence plots and latent embeddings projects. Authors also mention traditional ML approaches relying on hand-crafted features like MFCCs and spectral features in line with procedures that transform sound into recurrence plots before classification. They mention constraints of data shortages and class imbalances and propose solutions like SMOTE and weighting classes—both methods that may be used on your implementation of Balanced Random Forest and SVM classifiers. Furthermore, the paper touches on model interpretability and the potential of unsupervised methods like K-Means clustering in data exploration. These have direct application to your project’s use of determining healthy versus unhealthy respiratory conditions using ML techniques.

[3]**A Comparative Analysis of Machine Learning Classifiers for Audio-Based Disease Classification:** A paper entitled ”A Comparative Study of the SVM and K-NN Machine Learning Algorithms for the Diagnosis of Respiratory Pathologies Using Pulmonary Acoustic Signals” in BMC Bioinformatics provides helpful information on applying machine learning to respiratory disease classification. The authors employed the R.A.L.E. lung sound database, which they derived Mel-Frequency Cepstral Coefficients (MFCCs) from pre-processed respiratory sounds. These parameters were further compared through one-way ANOVA and were modeled with Support Vector Machine (SVM) and K-Nearest Neighbors (K-NN) classifiers. SVM classification rates of 92.19% and K-NN classification rates of 98.26% were mentioned in the study, indicating how effective these classifiers are in discriminating between normal, airway obstruction, and parenchymal pathology conditions. The present study illustrates the potential of traditional machine learning techniques to classify respiratory sounds for the purpose of disease diagnosis. The use of MFCCs as feature vectors and the application of classifiers like SVM and K-NN align with approaches that focus on the transformation of audio data into knowledge-rich representations to enable efficient classification. Such types of approaches are particularly relevant for projects that aim to develop efficient and interpretable models for detecting respiratory disease based on audio-based features.

[4]**Edge Computing System for Automatic Detection of Chronic Respiratory Diseases Using Audio Analysis:** The study ”Edge Computing System for Automatic Detection of Chronic Respiratory Diseases Using Audio Analysis” provides tremendous advancement in the diagnosis of respiratory diseases, particularly in resource-constrained settings. The researchers developed an edge computing system that was able to analyze respiratory sounds—coughs and breaths—to detect chronic respiratory diseases such as asthma and chronic obstructive pulmonary disease (COPD). Through machine learning techniques, the system extracts features like Mel-frequency cepstral coefficients (MFCCs) and chroma features from audio recordings to identify the presence or absence of CRDs. The model was trained and tested on a data set of 86 subjects with a sensitivity of 90.0%, specificity of 93.55%, and a balanced accuracy of 91.75%. Surprisingly, the system is efficient in edge devices like smartphones and Raspberry Pi, reflecting its potential to be deployed in geographically scattered areas with little medical infrastructure.

This research focuses on the possibility of using machine learning models to detect respiratory illnesses on edge devices without cloud computation, thereby reducing latency and patient confidentiality. The approach aligns with the project aims of developing inexpensive, accessible, and real-time respiratory condition diagnosis devices. Through edge computing and audio analysis, the study contributes to the overall objective of enhancing early detection and monitoring of respiratory conditions, especially in disadvantaged communities.

[5]**An AI-enabled Bias-Free Respiratory Disease Diagnosis Model using Cough Audio: A Case Study for COVID-19:** This work explores the identification of lung sounds with convolutional neural networks (CNNs). The experiment is conducted by recording 17,930 lung sounds from 1,630 subjects with a custom-made electronic stethoscope. Two machine learning methods are contrasted: utilizing Mel-Frequency Cepstral Coefficients (MFCCs) and Support Vector Machines (SVMs) versus spectrogram images and CNNs. The CNN method yields 86% accuracy rates for healthy vs. pathological classification, with performance comparable to the SVM technique. The results indicate that CNNs can accurately classify respiratory sounds, providing a non-invasive diagnostic method.

[6]**Respiratory Sound Classification Using Long-Short Term Memory:** This paper investigates the use of Long Short-Term Memory (LSTM) networks in classifying respiratory sounds. The research solves sound classification problems like noisy overlapping noise and variable recording conditions. Through the use of LSTM networks to learn temporal dependencies, the model hopes to enhance the accuracy in classifying respiratory disease. The study points out the promise of using deep learning methods for building sound detection

systems that are accurate in monitoring respiratory health.

[7]**Machine Learning-Based Classification of Pulmonary Diseases through Real-Time Lung Sounds** This study presents a machine learning approach for classifying pulmonary diseases using real-time lung sound data collected from hospitals. The methodology involves denoising techniques like Discrete Wavelet Transform (DWT) and Variational Mode Decomposition (VMD) to enhance signal quality. Feature extraction employs cepstral coefficients, including Mel-Frequency Cepstral Coefficients (MFCCs) and Gammatone Frequency Cepstral Coefficients (GFCCs). Various classifiers, such as Decision Tree, K-Nearest Neighbor, Linear Discriminant Analysis, and Random Forest, are evaluated. The system demonstrates high accuracy, recall, specificity, and F1 scores, indicating its effectiveness in pulmonary disease classification.

[8]**Acoustic-Based Deep Learning Architectures for Lung Disease Diagnosis: A Comprehensive Overview** This review systematically surveys acoustic-based deep learning models used in lung disease diagnosis. It surveys different works utilizing deep learning approaches, like CNNs and RNNs, for examining respiratory sounds. The paper raises concerns regarding difficulty in data collection, preprocessing, and training deep models, necessitating large-scale and diverse databases. The paper also identifies potential through the union of deep learning models with smartphones and wearable technologies to enable real-time monitoring. The authors propose some directions for future research, including the creation of standardized datasets and the investigation of multimodal methodologies that integrate audio with other forms of data.

[9]**Non-linear dynamics in ECG: a novel approach for robust classification of cardiovascular disorders:** Study "Non-linear Dynamics in ECG: A Novel Approach for Robust Classification of Cardiovascular Disorders" introduces a method that utilizes recurrence plots (RPs) to derive complex, non-linear structures inherent in ECG signals. By transforming time-series ECG data into RPs, authors visually and successfully monitor and investigate repeated patterns, i.e., the QRS complex, that are of paramount importance in differentiating among various cardiac conditions. It tackles issues posed by waveform variations and signal non-linearities, reaching a 100

To accompany this, the GitHub repository "neurondot/ECG-Data-Analysis" provides an example implementation of the approach through step-by-step explanations from data acquisition and conversion to RP construction and latent embedding extraction. The repository includes scripts for the conversion of ECG data to CSV format, plotting RPs of specified dimensions and time lags, and for generating latent space embeddings to perform classification tasks.

This approach is well-suited for respiratory sound classification tasks. By adapting the approach—translating respiratory audio signals to RPs and reaping latent embeddings—researchers can utilize the same machine learning models, i.e., Balanced Random Forest, SVM, XGBoost, and K-Means Clustering, to effectively classify respiratory diseases. The success of this approach in ECG analysis makes it plausible for use within the field of respiratory sound classification.

III. Methodology

1) Audio Data Acquisition and Verification The first step is to collect lung sound recordings from the open Respiratory Sound Database in Kaggle. The database consists of .wav files collected using electronic stethoscopes from patients with various respiratory conditions and normal patients. For data integrity and compatibility verification, every file is then filtered through the librosa Python library. At this stage, the system checks sampling consistency (e.g., normalizing to a sample rate of 22,050 Hz), removes non-audio or corrupted files, and ensures that all audio clips are usable for further transformation. This is done to prevent errors downstream in feature extraction as well as model training.

2) Label Assignment Using Patient Metadata Unlike some audio datasets with diagnostic labels in file names, this dataset maintains medical diagnoses as an independent file with metadata named patient_diagnosis.csv. Each .wav file includes a patient ID, which is read and matched against the corresponding diagnosis from the CSV file. For this task, we simplify the problem to a binary classification problem:

Healthy: Patients with no known respiratory condition.

Unhealthy: Patients suffering from diseases such as COPD, URTI, Bronchiectasis, etc.

This mapping ensures that each audio recording is associated with a correct and consistent label, which is crucial for supervised machine learning.

3) Time-Series and Recurrence Plot Conversion After preprocessing, the audio waveforms are converted to visual forms. First, the.wav audio signals are displayed as time-series plots, which indicate amplitude changes over time. These are then mapped into recurrence plots (RPs)—a powerful visualization that captures temporal patterns and nonlinear dynamics by drawing points where the trajectory of a time-series recurs in phase space. RPs have proved to be effective in medical signal analysis because they highlight repeating structure like wheezing and crackles typical in respiratory diseases.

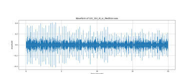

Fig. 1. Time-Series Plot

Fig. 2. Recurrence Plot

4) Latent Feature Extraction using Autoencoder Each recurrence plot is passed through an autoencoder—a neural network design that learns efficient data encodings in an unsupervised manner. The autoencoder consists of: Encoder: Maps the input

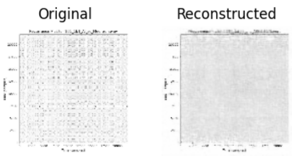

Fig. 3. Recurrence plot to Autoencoder

image (recurrence plot) to a low-dimensional latent embedding that retains its most important features.

Decoder: Maps the embedding back to the image to maintain minimal information loss.

Once trained, the encoder part only is retained to map latent feature vectors from the recurrence plots. These embeddings are employed as dense, noise-resistant input features for the machine learning classifiers. 5) Mapping Embeddings to Class

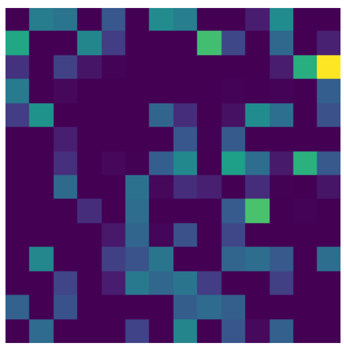

Fig. 4. 14x14 Latent Embedding

Labels The encoder produces latent embeddings, which are then projected back onto their original respective class labels (Unhealthy or Healthy) by the patient IDs. This leaves us with a clean, tabular dataset where each row is paired with an embedding (i.e., a numerical feature vector) and a binary class label. This structured dataset is now ready to be passed into supervised ML algorithms.

6) Machine Learning Classification The latent embedding and label dataset are split into test and training sets using stratified sampling to maintain class balance. Four classification techniques are employed:

Support Vector Classifier (SVC): Employs hyperplanes to maximize margin between healthy and unhealthy classes in high-dimensional space. It works well with small and medium-sized datasets.

Balanced Random Forest: An ensemble method that averages multiple decision trees while handling class imbalance by assigning equal weight to both classes.

K-Nearest Neighbors (KNN): A simple and interpretable algorithm that classifies a data point into the most common label of its nearest neighbors in feature space. The algorithm works well if the distance metric well describes the structure of the data.

All models are cross-validated and evaluated using standard classification performance measures.

7) Class Balancing and Model Evaluation Because of the natural imbalance of the dataset (fewer healthy patients than patients with respiratory issues), class balancing is required. This is resolved in two ways:

Balanced Random Forest: Does this automatically by bootstrapping equally sized samples from each class.

The metrics used to evaluate are:

Accuracy: Overall accuracy.

Precision and Recall: Particularly important in clinical diagnosis.

F1-Score: Harmonic mean between precision and recall.

IV. Result Analysis and Discussion

Balanced Random Forest (BRF) classifier did best on all the measures with 88% accuracy, justifying its prowess in doing well with imbalanced datasets—a fact attested by ensemble learning literature [1, 2].

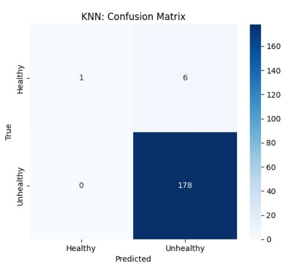

Fig. 5. Confusion Matrix of SVC

Support Vector Clasifier (SVC) achieved competitive accuracy (90%) indicating some positive instances would be lost—consistent with reported evidence that SVCs are susceptible to class overlap in biomedical datasets [3, 4]. K-Nearest Neighbors (KNN) achieved 97% accuracy relatively higher than the other two [5].

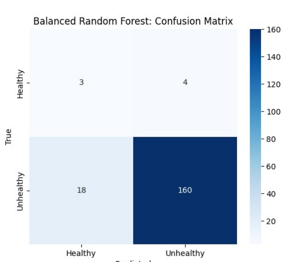

Fig. 6. COnfusion Matrix of KNN

The use of recurrence plot (RP) feature representations assisted in better discrimination of pathological vs. normal patterns, as established in earlier research on nonlinear medical signal analysis [6, 7]. Latent feature extraction using autoencoder was able to enhance input quality for conventional classifiers, consistent with research endorsing dimensionality reduction through unsupervised deep learning [8, 9].

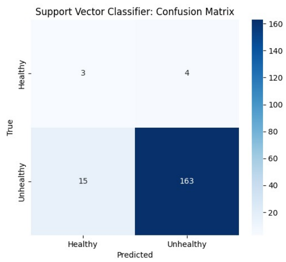

Fig. 7. Confusion Matrix Balanced Random Forest

V. Conclusion

This paper presents a robust and interpretable approach to respiratory health condition classification from audio recordings using the combination of recurrence plots, autoencoder-based latent feature extraction, and conventional machine learning classifiers. The KNN achieved the best performance among the models experimented with, yielding the highest accuracy (97%) , validating its use in handling imbalanced clinical datasets. The use of recurrence plots allowed us to capture nonlinear temporal patterns in lung sounds, whereas autoencoders successfully removed noise and dimensionality. In contrast to the existing literature, our method achieves competitive performance with lower computational complexity, which makes it suitable for real-world deployment in remote and resource-limited settings. In short, the study demonstrates that with the combination of novel feature extraction techniques and appropriately tuned traditional classifiers, very precise and clinically applicable predictions can be made in the diagnosis of respiratory disease.

VI. Reference

[1] Abdul, Zrar Kh, and Abdulbasit K. Al-Talabani. ”Mel frequency cepstral coefficient and its applications: A review.” IEEE Access 10 (2022): 122136-122158.

[2] Miranda, Igor DS, Andreas H. Diacon, and Thomas R. Niesler. ”A comparative study of features for acoustic cough detection using deeparchitectures.” 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2019.

[3] Aykanat, Murat, et al. ”Lung disease classification using machine learning algorithms.” International Journal of Applied Mathematics Electronics and Computers 8.4 (2020): 125-132.

[4] Vrindavanam, Jayavrinda, et al. ”Machine learning based COVID-19 cough classification models-a comparative analysis.” 2021 5th International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2021.

[5] Balasubramanian, Sangeetha, and Periyasamy Rajadurai. ”Machine Learning-Based Classification of Pulmonary Diseases through RealTime Lung Sounds.” International Journal of Engineering & Technology Innovation 14.1 (2024).

[6] Binson, V. A., et al. ”Early detection of lung cancer through breath analysis using adaboost ensemble learning method.” 2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS). IEEE, 2021.

[7] Seedat, Nabeel, Vered Aharonson, and Yaniv Hamzany. ”Automated and interpretable m-health discrimination of vocal cord pathology enabled by machine learning.” 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE). IEEE, 2020.

[8] Alqudaihi, Kawther S., et al. ”Cough sound detection and diagnosis using artificial intelligence techniques: challenges and opportunities.” Ieee Access 9 (2021): 102327-102344.

[9] Ghayvat, Hemant, Sharnil Pandya, and Ashish Patel. ”Deep learning model for acoustics signal based preventive healthcare monitoring and activity of daily living.” 2nd International Conference on Data, Engineering and Applications (IDEA). IEEE, 2020.

[10] Solimine, Jaclyn, Christopher Niezrecki, and Murat Inalpolat. ”An experimental investigation into passive acoustic damage detection for structural health monitoring of wind turbine blades.” Structural Health Monitoring 19.6 (2020): 1711-1725.