Deep Learning Project Proposal

Team Name: Alpha Neurons

Team Members:

Sonal Shreya (sonalshreya@usf.edu, Role: Model building and explainability implementation)

Sumit Kumar Singh (sumit141@usf.edu,Role: Data preprocessing, Model Building)

Santhoshini Bojanapally (bs441@usf.edu, Role: Data preprocessing, explainability implementation)

Sai Naga Saujanya Gullapaly (sainagasoujanya@usf.edu, Role: Model evaluation and Visualizations)

Title: Enhancing Medical Diagnosis through Spectroscopy-Based Respiratory Sound Analysis.

Abstract

AI-based diagnostic systems have immense potential in transforming healthcare, particularly in the timely detection of respiratory diseases. Nonetheless, lack of interpretability in octopus nets remains one of the chief barriers to their clinical deployment, particularly in children. Models also promise to empower clinicians, and they cannot be simply high-performing systems but rather transparent systems that can explain how and why a given prediction was made so that clinical safety and trust can be built.

This project aims to create an interpretable deep learning pipeline for respiratory sound classification. This paper attempts to classify respiratory diseases from lung sound spectrograms using the ICBHI 2017 Lung Sound Dataset. We utilize Convolutional Neural Networks (CNNs) because of their success in learning spatial patterns from audio spectrogram images. In an effort to address the black-box nature of CNNs, we appended Grad-CAM++, a state-of-the-art explainability technique, to produce visual explanations identifying which parts of spectrograms contribute most to the model's predictions.

With this approach, we seek to achieve not just accurate classification performance, but also robust and clinically relevant local visualizations that render decisions interpretable to healthcare providers. Early results show good accuracy and intuition in visual interpretation. We hope this project will thus fill the first steps gap between AI performance and clinical trust in their plausible use and therefore provide the basis for future real-world validation and deployment of explainable AI in a health-care context.

Introduction

Background and Motivation

Respiratory diseases continue to be one of the most significant causes of morbidity and mortality globally, especially among vulnerable populations, including children and the older adult. Early and accurate diagnosis is key to effective intervention, but access to expert clinical evaluation is often limited, particularly in remote or resource-constrained settings.

AI systems have the potential to change how we detect diseases due to having cheap and scalability systems which can be built. In the context of AI tools, Convolutional Neural Networks (CNNs) have demonstrated potential in automatically classifying the spectrograms of respiratory sounds. But they are "black boxes," meaning that while the model and numbers behind its predictions may be impressive, the way the model comes to its decisions is not intelligible to the clinicians. The absence of transparency restricts their acceptance in clinical practice, where interpretability and trust are critical. So, there is an increasing demand for ai models which are well-performing but also provide interpretable insights into their predictions.

Problem Statement

Convolutional neural networks (CNNs) are increasingly used to classify respiratory sounds from spectrograms, but their lack of interpretability remains a critical barrier to clinical adoption. Without insight into the way CNN arrives at its predictions, physicians are unlikely to trust or act on its output. This opacity is especially problematic in medical contexts where diagnostic decisions carry high stakes. To address this, our project integrates Grad-CAM++ to produce fine-grained visual explanations, enabling clinicians to understand, evaluate, and potentially validate the CNN's focus areas on spectrograms.

Research Question

Which way can we use Grad-CAM++ to generate clinically interpretable visual explanations for CNN-based classification of respiratory sounds from spectrograms?

Sub-questions include:

- What spectrogram features do CNNs focus on when classifying crackles and wheezes, as revealed by Grad-CAM++?
- How consistent are Grad-CAM++ visual explanations across multiple correctly classified samples?
- Can Grad-CAM++ heatmaps be interpreted by clinicians as highlighting diagnostically relevant regions (optional)?
- Does incorporating Grad-CAM++ improve the explainability of CNNs without compromising classification performance?

Objective

Convolutional neural networks (CNNs) are well-suited to analyzing spectrograms of respiratory sounds due to their ability to extract spatial patterns. However, their decision-making processes remain opaque, limiting their usefulness in real-world medical settings. This project focuses on integrating explainability techniques, specifically Grad-CAM++, to make CNN predictions more transparent and clinically actionable.

Through this project we address the lack of interpretability in convolutional neural networks (CNNs) used for classifying respiratory sounds from spectrograms. By applying Grad-CAM++, we aim to generate detailed explanations that highlight the regions of spectrograms contributing to each classification. Using the ICBHI 2017 Respiratory Sound Dataset, our goal is to improve clinical trust in CNN-based models by making their predictions transparent and easier for physicians to validate.- https://bhichallenge.med.auth.gr/ICBHI_2017_Challenge,.

Literature Review/Related Work

In recent years, the capability of Convolutional Neural Networks (CNNs) has been shown through various studies to outperform the traditional ML approaches in both accuracy and robustness in terms of classifying respiratory sounds. The ICBHI 2017 dataset has established itself as a so-called standard benchmark, for example in Demir et al. (2021) and Zhang et al. (2022), which utilized CNNs to analyze spectrograms for wheeze and crackle detection.

One significant challenge, however, persists: CNNs are often considered “black boxes,” and thus their use in clinical settings where interpretability is critical is limited. In order to remedy this, researchers have incorporated Explainable AI (XAI) techniques. Selvaraju et al. which later enhanced by Chattopadhay et al. with Grad-CAM++ for more accurate class specific visual explanations.

Recent works such as Lopes et al. (2023) and Kumar et al., 2023), which applied Grad-CAM variants to biomedical audio and image diagnoses, indicating a trend towards interpretable deep learning in medical applications. Our work builds on this line of research by utilizing CNNs with Grad-CAM++ for the classification of respiratory sounds and providing interpretable visual justification that can guide clinicians' decision-making.

Methodology:

This project focuses exclusively on convolutional neural networks (CNNs) trained on spectrogram representations of respiratory sounds, with Grad-CAM++ used for post hoc visual explanation of predictions.

Dataset Overview

ICBHI 2017 Respiratory Sound Dataset - https://bhichallenge.med.auth.gr/ICBHI_2017_Challenge

The Respiratory Sound Database contains audio samples, collected independently by two research teams in two different countries, over several years. Most of the database consists of audio samples recorded by the School of Health Sciences, University of Aveiro (ESSUA) research team at the Respiratory Research and Rehabilitation Laboratory (Lab3R), ESSUA and at Hospital Infante D. Pedro, Aveiro, Portugal. The second research team, from the Aristotle University of Thessaloniki (AUTH) and the University of Coimbra (UC), acquired respiratory sounds at the Papanikolaou General Hospital, Thessaloniki and at the General Hospital of Imathia (Health Unit of Naousa), Greece.

The database consists of a total of 5.5 hours of recordings containing 6898 respiratory cycles, of which 1864 contain crackles, 886 contain wheezes, and 506 contain both crackles and wheezes, in 920 annotated audio samples from 126 subjects.

The cycles were annotated by respiratory experts as including crackles, wheezes, a combination of them, or no adventitious respiratory sounds. The recordings were collected using heterogeneous equipment and their duration ranged from 10s to 90s. The chest locations from which the recordings were acquired is also provided. Noise levels in some respiration cycles is high, which simulate real life conditions.

Each audio file in the ICBHI 2017 dataset is named using five elements: patient ID, recording index, chest location (e.g., Trachea, Anterior, Posterior), acquisition mode (single or multi-channel), and recording device type.

The corresponding annotation files mark the start/end of respiratory cycles and indicate the presence of crackles or wheezes. Diagnostic labels (like COPD, URTI, LRTI) and demographic data are also provided, making the dataset rich for research in respiratory sound classification.

Data Preprocessing

To support interpretable CNN modeling using Grad-CAM++, the following preprocessing pipeline is applied to respiratory sound recordings:

- **Noise Reduction:** Wavelet denoising and bandpass filtering are used to remove background interference, enhancing the clarity of diagnostically relevant sound features in the spectrograms.
- **Spectrogram Generation:** Raw .wav files are converted into 2D time–frequency spectrograms, which serve as both the input to the CNN and the visual medium for Grad-CAM++ explanations.
- **Data Augmentation:** Controlled variations such as pitch shifting, time-stretching, and noise injection are introduced to improve the CNN's robustness across different recording conditions without compromising interpretability.
- **Normalization:** Spectrogram intensities are normalized across samples to make sure consistent contrast and resolution in both model input and Grad-CAM++ visualizations.

Data Representation – Spectrograms for Visual Interpretability

Respiratory sounds contain rich time–frequency patterns that manifest as visual structures in spectrograms. Wheezes and crackles, for example, appear as distinctive high-frequency bands or short-duration bursts. By converting audio recordings into 2D spectrograms, we create inputs that are not only well-suited for CNN classification but also ideal for applying Grad-CAM++ to generate intuitive heatmap-based explanations.

Base Model – CNN (ResNet-18 or EfficientNet-lite)

We utilize CNN architectures such as ResNet-18 and EfficientNet-lite to classify respiratory spectrograms. These models are capable of learning spatial hierarchies in frequency–time patterns critical for identifying abnormal respiratory events. Their layered structure supports Grad-CAM++ activation mapping, which helps localize the specific regions in the spectrogram that influenced the network's classification.

Explainability – Grad-CAM++ for Transparent Visual Reasoning

Grad-CAM++ is integrated into the pipeline to generate fine-grained visual explanations of CNN decisions. By projecting class-discriminative gradients onto the feature maps of the final convolutional layers, Grad-CAM++ produces heatmaps that reveal which parts of the spectrogram the CNN relied on for its predictions. These heatmaps can help physicians assess whether the model's focus aligns with clinically meaningful sound features, such as wheeze regions or crackle bursts.

Explainability Framework (Innovative XAI Methods):

To address the interpretability gap in CNN-based respiratory sound classification, we incorporate **Grad-CAM++** as our sole explainability method. Grad-CAM++ computes class-specific gradient information at the final convolutional layers and maps it back to the spectrogram, generating a heatmap that visually highlights the regions most influential in the model's decision. These visual explanations allow for clinical inspection and evaluation of whether the model's focus aligns with pathologically relevant sound patterns such as wheezes or crackles.

By using spectrograms as inputs and Grad-CAM++ as the visual explanation tool, we aim to bridge the gap between model performance and clinical trust through transparent, interpretable AI outputs.

The Model Architecture:

The proposed model architecture is a **convolutional neural network (CNN)** trained on 2D spectrogram representations of respiratory sounds. It is composed of the following components:

- **CNN Backbone:** A pre-defined architecture such as ResNet-18 or EfficientNet-lite is used for feature extraction. These networks are well-suited for spatial learning from spectrograms and are compatible with Grad-CAM++ for visualization.
- **Output Layer:** A fully connected (dense) layer followed by a softmax activation is used for multi-class classification (Normal, Crackle, Wheeze, Combined).
- **Explainability Integration:** Grad-CAM++ is applied post hoc to generate class-discriminative heatmaps that highlight the specific regions in the spectrogram that influenced the model's decision.

This architecture is designed not only to deliver accurate classifications but also to provide **transparent, visual insight** into the way those decisions are made, supporting its integration into clinical decision-making pipelines.

Expected Impact

- **Clinical Trust through Explainability:** By integrating Grad-CAM++ into CNN-based respiratory classification, the model offers physicians visual explanations that highlight diagnostically relevant regions of the spectrogram, enabling greater trust and potential use in decision support.
- **Model Transparency for Deployment Readiness:** Transparent decision-making processes help move CNN models closer to real-world clinical adoption by addressing one of the most critical barriers — interpretability.
- **Educational Utility:** The visual outputs generated by Grad-CAM++ may also serve as teaching aids for medical trainees learning to interpret audio-based respiratory anomalies.
- **Focused Advancement in Medical XAI:** This project contributes to a growing field of explainable AI in medical imaging by demonstrating the feasibility of CNN explainability in audio-based diagnostics, without relying on temporal or hybrid models.

Methodology

- **ICBHI 2017 Respiratory Sound Database**
  Contains 920 audio recordings and 6,898 annotated respiratory cycles labeled for crackles, wheezes, and combined events. Each .wav file is paired with a .txt annotation file specifying the time intervals and labels.

Based on the detailed feedback provided, we refined our proposal to focus exclusively on **CNN explainability** using **Grad-CAM++**. We eliminated previously proposed components that broadened the scope, such as temporal modeling with GRUs/LSTMs, attention mechanisms, and comparisons across multiple XAI techniques. Instead, we now center our project around a single, well-defined goal: applying Grad-CAM++ to a CNN trained on spectrograms to generate interpretable visual explanations for respiratory sound classification. Our updated problem statement, architecture, and evaluation metrics now reflect this narrowed, clinically relevant focus.

Audio Processing and Spectrogram Generation:
We have successfully preprocessed the audio files from the ICBHI 2017 Respiratory Sound Database. Each .wav file was segmented using its associated .txt annotation file. These segments were then converted into 2D spectrograms suitable for CNN input and Grad-CAM++ visualization.

Noise Reduction (Wavelet Denoising + Bandpass Filtering):
Contrary to our earlier omission, we have now incorporated noise reduction techniques to improve signal quality. Specifically, we applied wavelet-based denoising and bandpass filtering during preprocessing to reduce background interference and highlight diagnostic sound patterns.

Dataset Compilation:
Each respiratory cycle was labeled (normal, crackle, wheeze, or both) based on annotation metadata and stored along with its corresponding spectrogram and metadata. The dataset has been saved in respiratory_dataset.pkl format and is ready for CNN training.

Milestones and Progress Timeline

Phase	Timeline	Status
Problem Understanding & Literature Review	Week 1	Completed
Project proposal Refinement	Week 2-3	Completed
Dataset Preprocessing	Week 4	Completed
CNN Model Development	Week 5	In Progress
Integration of Grad-CAM++	Week 5	TBD
Model Training & Evaluation	Week 5	TBD
Results Analysis	Week6	TBD

Documentation & Final Report	Week 6	TBD
Presentation Review	Week 6	TBD
Feedback incorporation	Week 7	TBD
Final Presentation	Week 7	TBD

Risks and Mitigation

Potential Risks and Mitigation Strategies

Risk	Description	Mitigation Strategy
Inconsistent Spectrogram Quality	Variations in segment length, volume, or noise could reduce model performance	Apply consistent time-window slicing and normalization; filter low-quality samples
Grad-CAM++ Heatmaps Not Clinically Interpretable	Explanations may not align with known clinical features (e.g., wheeze patterns)	Validate visually; consult with external domain experts if time permits; document any observed patterns and limitations
Training Time or Resource Constraints	Limited compute capacity for CNN training and Grad-CAM++ visualization	Use lightweight models (e.g., ResNet-18, EfficientNet-lite); reduce batch size; perform cloud-based training if necessary

Conclusion:

This proposal presents a focused approach to integrating explainability into convolutional neural networks (CNNs) for respiratory sound classification using spectrograms. By applying Grad-CAM++ to visualize class-discriminative regions in the spectrograms, we aim to make model predictions more transparent and clinically interpretable. Our objective is not only to achieve accurate predictions but to support physician trust by clearly showing how the model arrives at its decisions. Although direct physician feedback is outside the current project scope, the proposed framework lays the groundwork for future collaboration with clinicians to further validate and refine the interpretability of the system.

References:

- Assessment of Performance, Interpretability, and Explainability in Artificial Intelligence–Based Health Technologies: What Healthcare Stakeholders Need to Know - Source
- Application of explainable artificial intelligence in medical health: A systematic review of interpretability methods - Source.
- ICBHI 2017 Respiratory Sound Dataset - https://bhichallenge.med.auth.gr/ICBHI_2017_Challenge
- https://paperswithcode.com/dataset/icbhi-respiratory-sound-database
- https://paperswithcode.com/paper/cycleguardian-a-framework-for-automatic