# Wireless Earphone-Based Real-Time Monitoring of Breathing Exercises: A Deep Learning Approach

Hassam Khan Wazir, Zaid Waghoo, and Vikram Kapila

Mechatronics, Controls, and Robotics Laboratory
New York University, Tandon School of Engineering, USA

46TH ANNUAL IEEE International Conference of the Engineering in Medicine and Biology Society
Orlando 2024
July 15-19, 2024

Poster No. 338

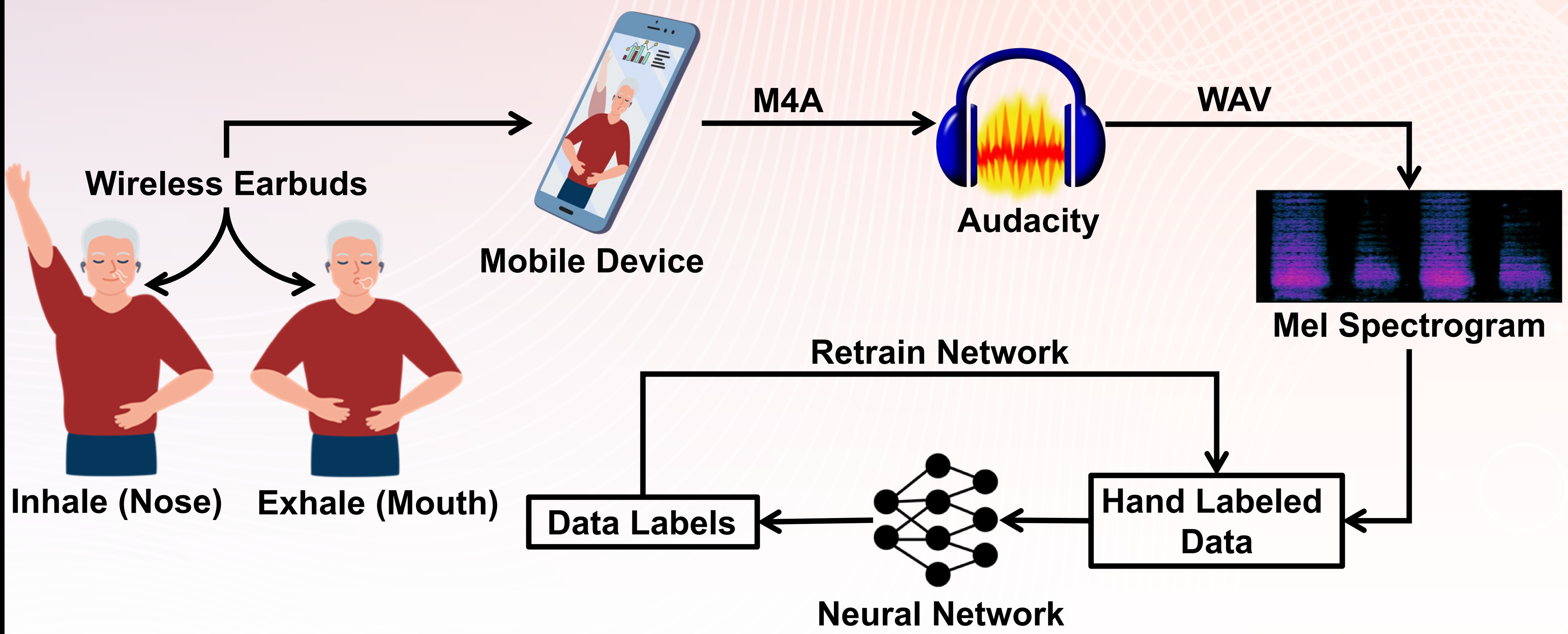
## INTRODUCTION

- **Deep breathing exercises** reduce **hypertension** and **anxiety**
- Used by several medical **therapy interventions**
  - **Lymphedema**, asthma, and chronic obstructive pulmonary disease
- **Tracking breathing** using a monitoring device **can support therapy compliance**
- **Lymphedema is an incurable condition** affecting 1 in 1000 Americans
  - Caused by an **obstruction in the lymphatic vessels**
  - Approximately **40% of breast cancer survivors suffer from lymphedema**
- **The optimal lymph flow (TOLF)**
  - Therapeutic **exercise program** for **post-operative breast cancer patients** to **prevent lymphedema**
  - Includes **easy-to-learn physical and breathing exercises**

**Proposal**

- Develop a system for **detection of breathing phases** (inhale/exhale) and **channels** (nose/mouth)
- Create an **annotated breathing audio dataset**, using wireless earphones

## DATA COLLECTION

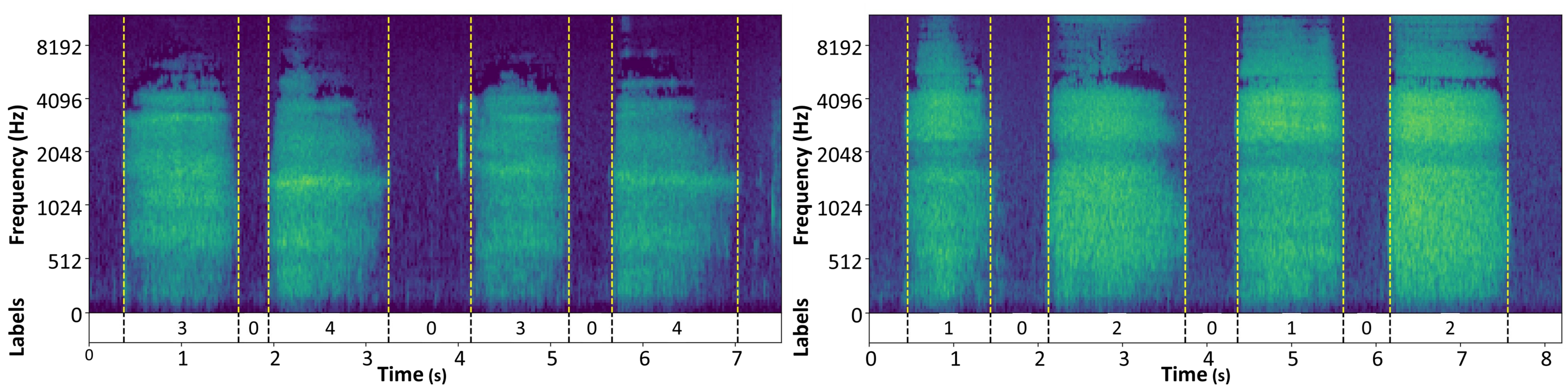



- **Eight healthy individuals performed controlled breathing exercises** (inhaling through the nose and exhaling through the mouth)
- **Breathing audio** was **captured using wireless earbuds** connected to a mobile device and the audio was **converted to WAV format**
- **Mel spectrograms** were generated to visualize the audio frequencies over time
- **Mel spectrograms of two audio files were manually labeled** to identify different parts of the breathing cycle (inhale and exhale phases)
- A binary **convolutional neural network** (CNN) was trained to **distinguish between pauses and breaths** and the pre-labeled files were manually corrected for minor mistakes
- The **CNN was iteratively retrained** with new audio data until it achieved **96% accuracy**

## TRAINING PIPELINE

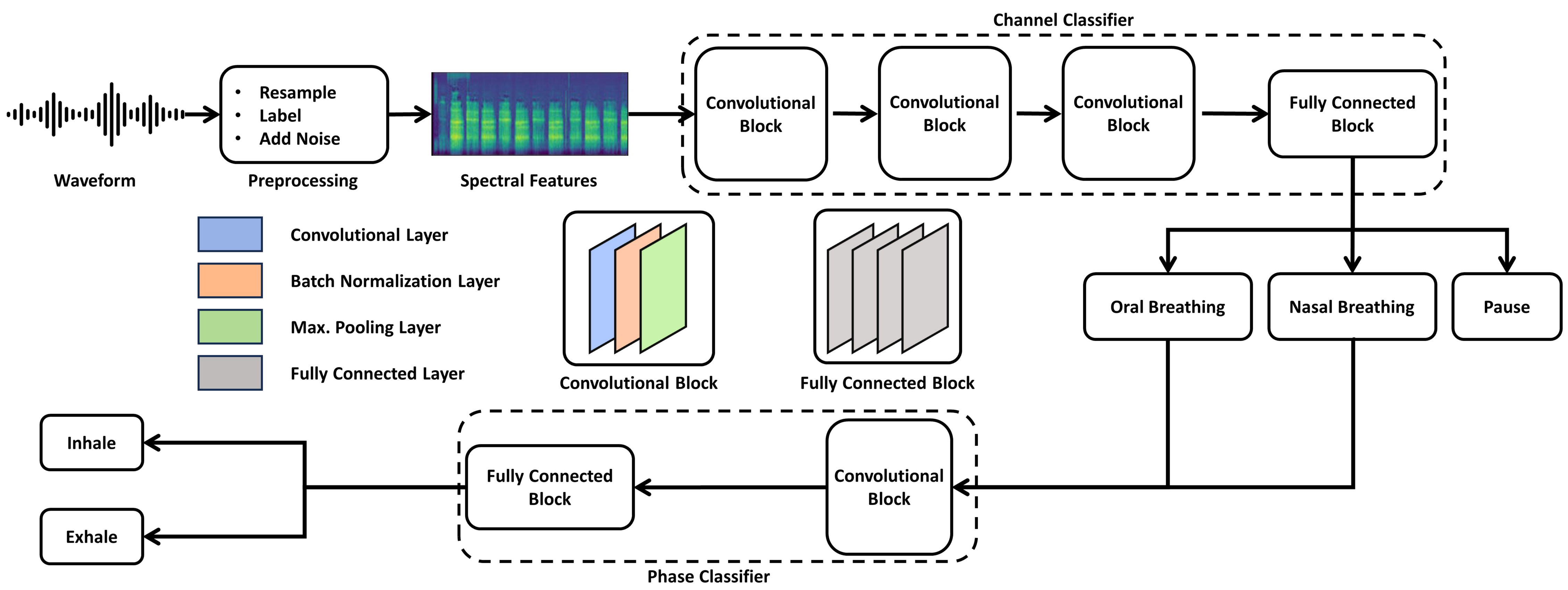
**Audio Preprocessing**

- **Resample audio** clips at 16 kHz
- **Segment and label audio** into 500 ms intervals with a 250 ms stride for overlap
- **Multilabeling** with **one-hot encoding**
- **Augmented training data with noise** (20 dB to 40 dB signal-to-noise ratio)
- **Convert** audio **to mel spectrograms** and **MFCCs**




**Model Training Pipeline**

- **Channel Classifier**
  - Three convolutional blocks each followed by batch normalization and max pooling
  - Four fully connected layers after the last convolutional block
  - **Classifies pause, nasal breathing, and oral breathing**
- **Phase Classifier**
  - Single convolutional block followed by four fully connected layers
  - **Distinguishes inhale from exhale** for nasal/oral segments
- Binary cross-entropy loss function used for training




## RESULTS

- **Leave-one-out cross-validation (LOOCV)** method was used with data from seven subjects for training and the eighth for validation
- **Channel Classifier**: Mel-spectrograms as input features yielded the best results with an average F1 score of 93.98% (SD=5.02%)
- **Phase Classifier**: Mel-spectrograms also outperformed MFCCs, with an average F1 score of 76.20% (SD=8.76%)
- Mel-spectrograms performed better than MFCCs for both channel and phase classification, although the phase classification scores were lower, indicating potential model or dataset limitations

| Classifier | Features | F1 Score (%) | | |
|---|---|---|---|---|
| | | Avg. (SD) | Max. | Min. |
| Channel | Mel spectrogram | 93.98 (5.02) | **97.99** | 82.45 |
| | MFCC | 90.35 (8.49) | 97.50 | 75.93 |
| Phase | Mel spectrogram | 76.20 (8.76) | **89.46** | 63.33 |
| | MFCC | 75.56 (8.55) | 87.35 | 63.71 |