RespireNet
A Deep Neural Network for Accurately Detecting Abnormal Lung Sounds in Limited Data Setting

Siddhartha Gairola[1], Francis Tom[2], Nipun Kwatra[1], Mohit Jain[1]

[1]Microsoft Research India

[2]Microsoft

https://healthtimes.com.au/hub/medical-technology/65/news/nc1/the-difference-between-stethoscopes-for-nurses-doctors/3518/

Motivation

Lung Auscultation: Listening to sounds from the lung with a stethoscope to diagnose and treat respiratory diseases.

Pros

- Low-cost, non-invasive process and simple to get signal
- Provides valuable information for screening and diagnosing lung diseases

Cons

- Requires medical professionals to analyze the respiratory signal
- Subjectivity in interpretations causing inter-listener variability.

Solution

Automated analysis, combined with digital stethoscopes can help overcome the drawbacks.

Digital Stethoscope | Respiratory Signal | Automated Analysis | Diagnosis Report

Abnormal Lung Sounds

Abnormal respiratory sounds like *crackle* and *wheeze* are useful in identifying specific respiratory diseases.

Wheeze:

- High-pitched continuous sound with frequency 100-2500Hz and Time > 80msec
- Typical symptom of asthma and COPD (chronic obstructive pulmonary disease)

Crackle:

- Discontinuous, non-tonal sound
- With frequency ~650Hz and duration ~5msec (for fine crackles, or frequency of 100-500Hz and duration ~15msec (for coarse crackle)
- Associated with COPD, chronic bronchitis, pneumonia and lung fibrosis

Our Focus

Automated method for detecting abnormal respiratory sounds *crackle* and *wheeze.*

Contributions:

- *RespireNet*, a simple CNN-based model for automatic *classification* of respiratory sounds.
- Detailed analysis of the ICBHI dataset
- Efficient use of limited data by a suite of novel techniques

ICBHI Dataset

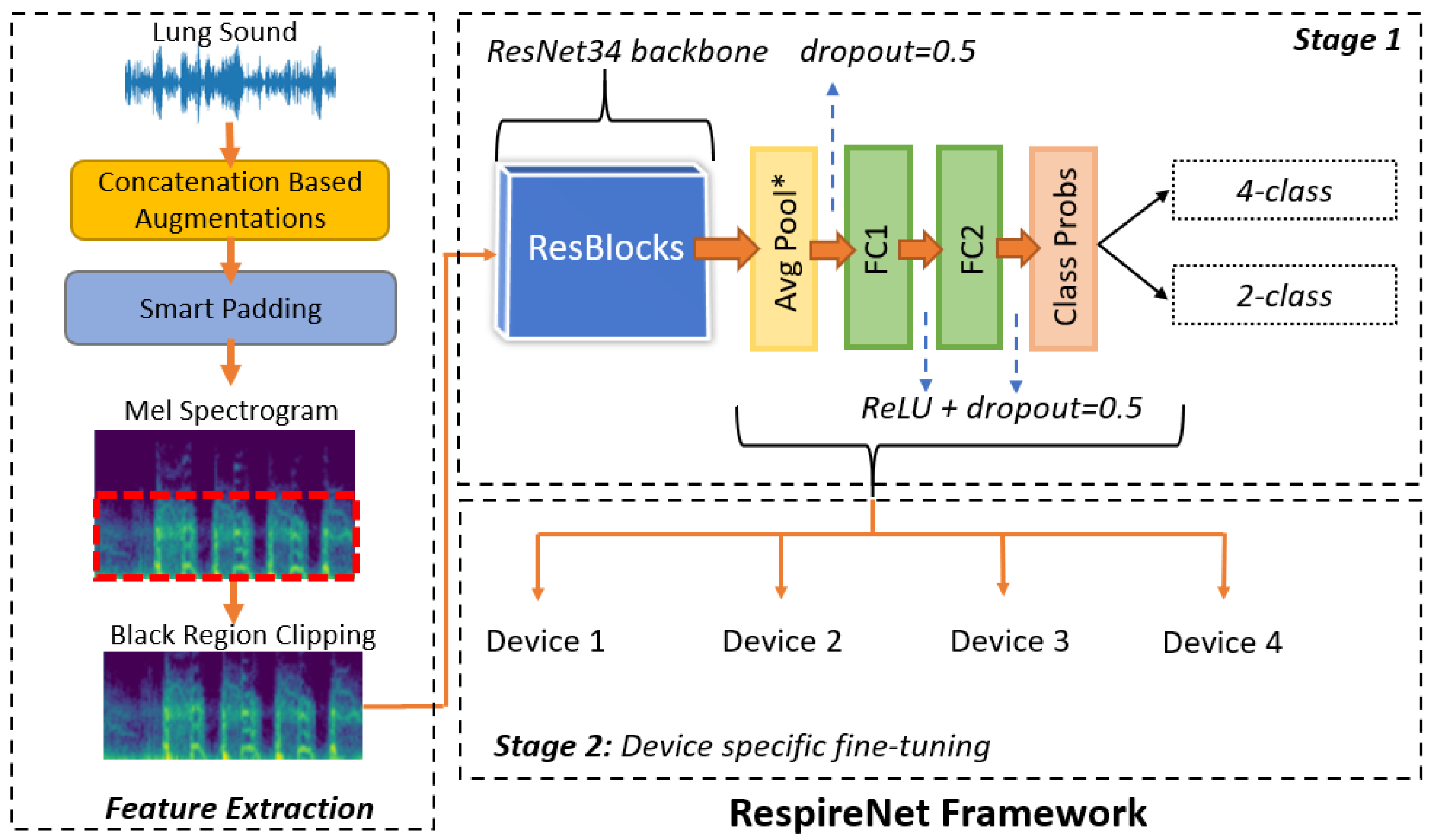
ICBHI Challenge dataset is the largest publicly available respiratory sound dataset.

Dataset Stats:

- 920 recordings containing **6898** respiratory cycles
- Total duration of recordings **5.5 hours**
- Collected from **126 patients**

	Normal	Crackle	Wheeze	Both	Total
Cycles	3642	1864	886	506	**6898**

Proposed Method: Overview



Pre-Processing: Data Standardization

Recordings have varying sampling rates (4kHZ – 44.1kHZ)

- Down-sample recordings to 4kHz

Noise Removal

- Apply 5th order Butterworth band-pass filter to remove noise (heartbeat, background speech, etc)

Normalization

- Normalization to map values between (-1.0, +1.0)

Data Augmentation

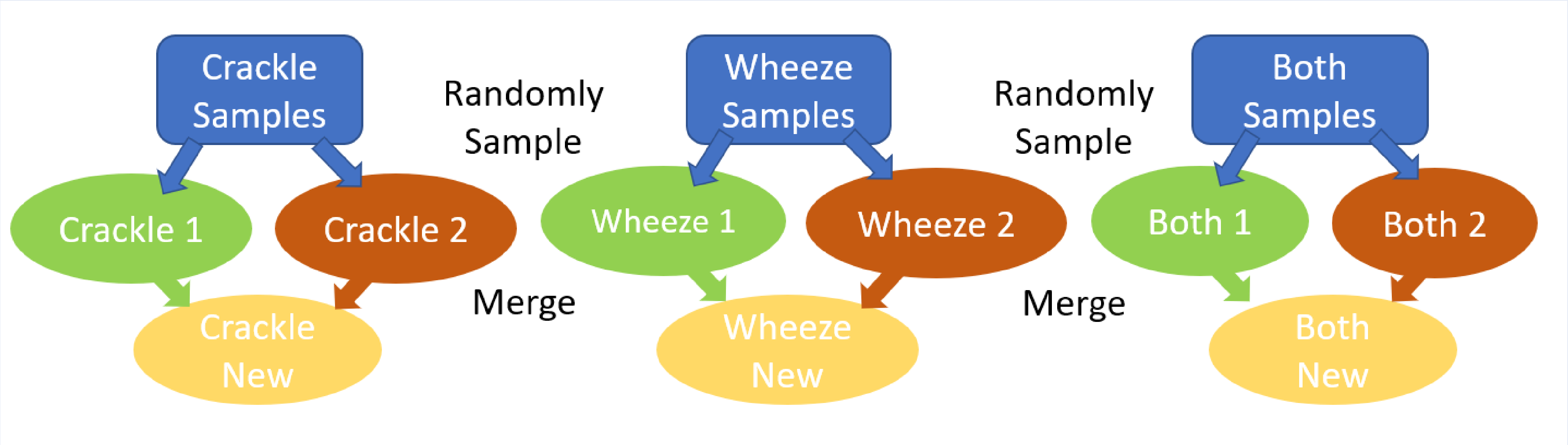
ICBHI dataset has small size and huge class imbalance

- (~53% Normal, ~27% Crackle, ~13% Wheeze, 7% Both)

Standard Augmentations

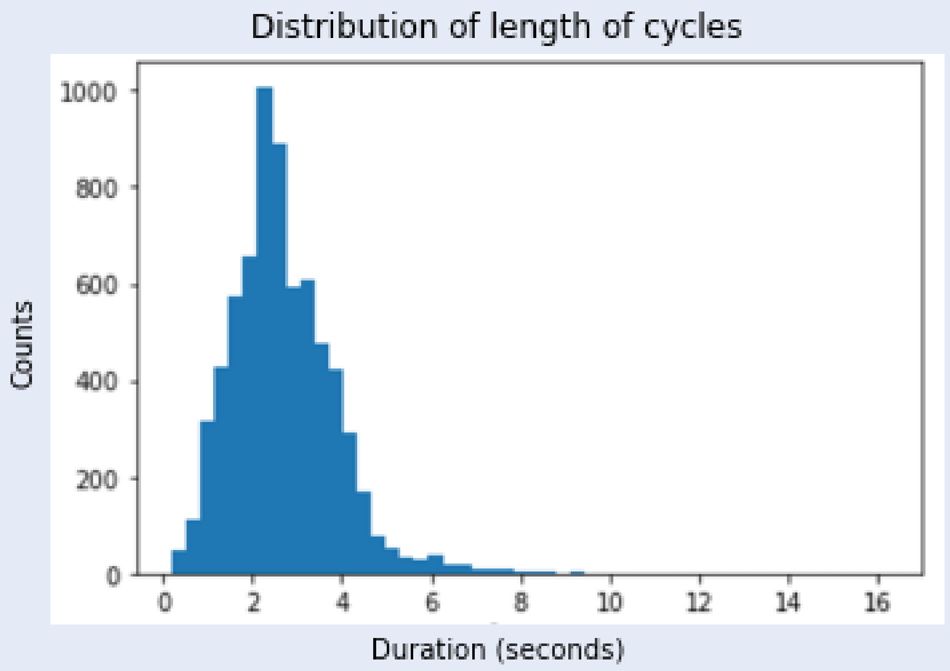
- Noise addition
- Speed variation
- Random Shift
- Pitch Shift

Concatenation Based Augmentation



Smart Padding

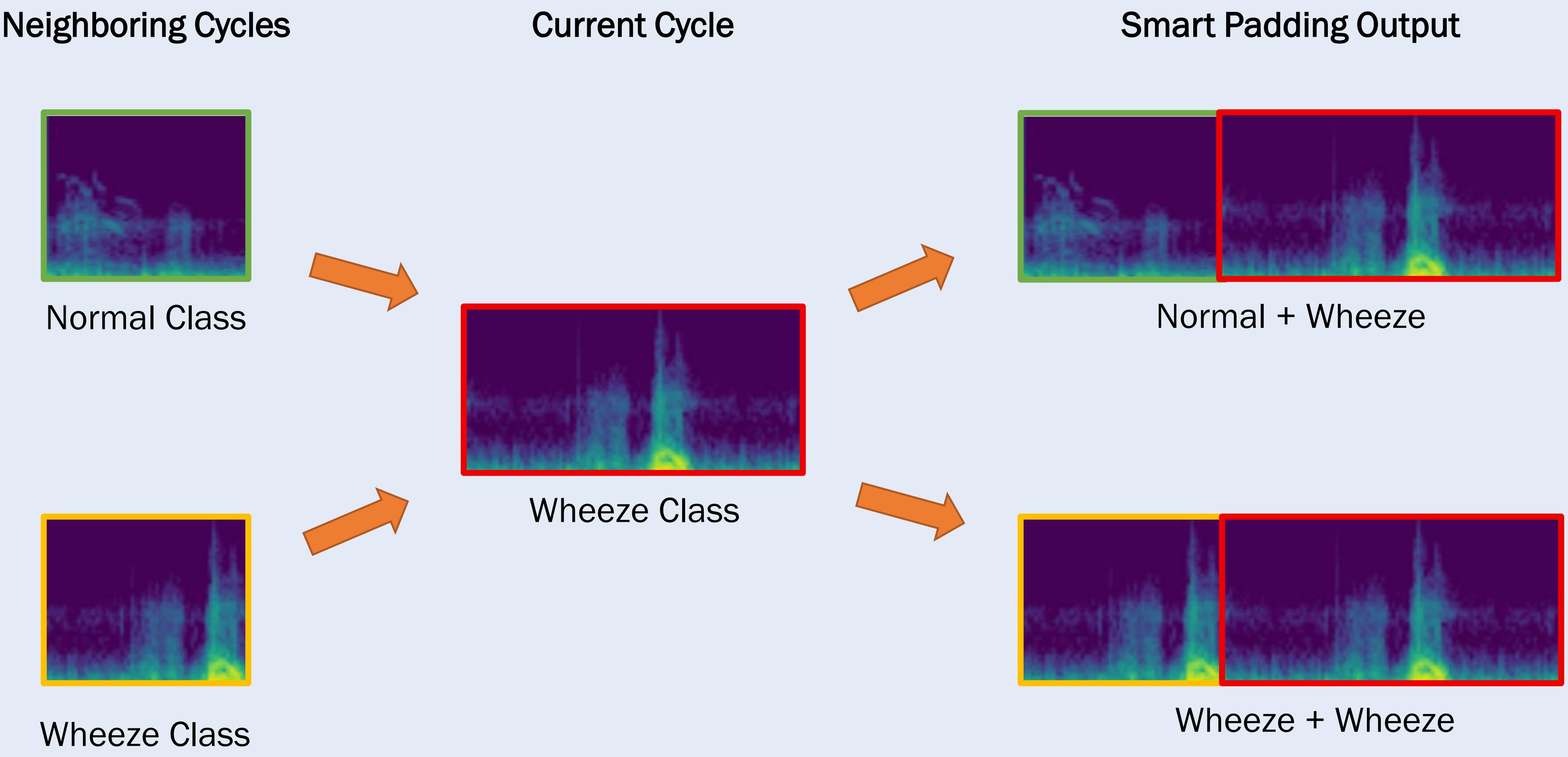
- Breathing cycle length varies within patients as well as across patients
- ICBHI dataset has varying length of breathing cycles ranging from 0.2s to 16.2s (mean cycle length = 2.7s)
- Cycle length must be standardized as CNN model requires fixed size input



Smart Padding

- Standardize cycle length to 7s
- For sample with cycle length < 7s, apply smart padding.
- Experiments demonstrate that a length of 7 second works best for the given dataset

Smart Padding



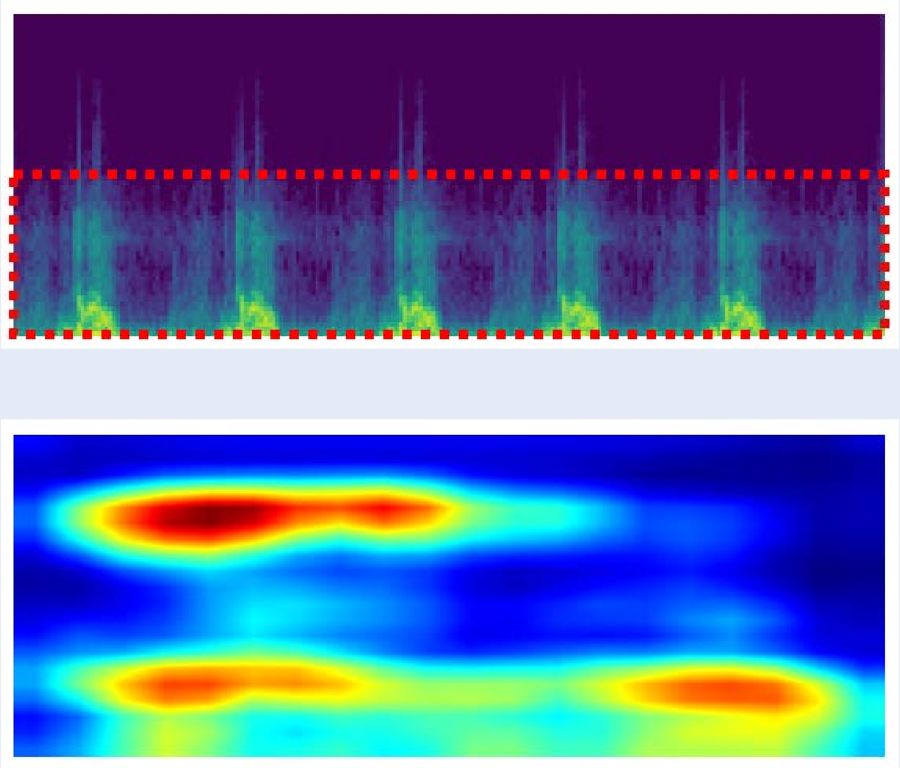
Blank Region Clipping

Many breathing cycles have no information in the higher frequency range

- Eg: 100% of the Litt3200 device samples had no information in the 1500 – 2000 Hz band

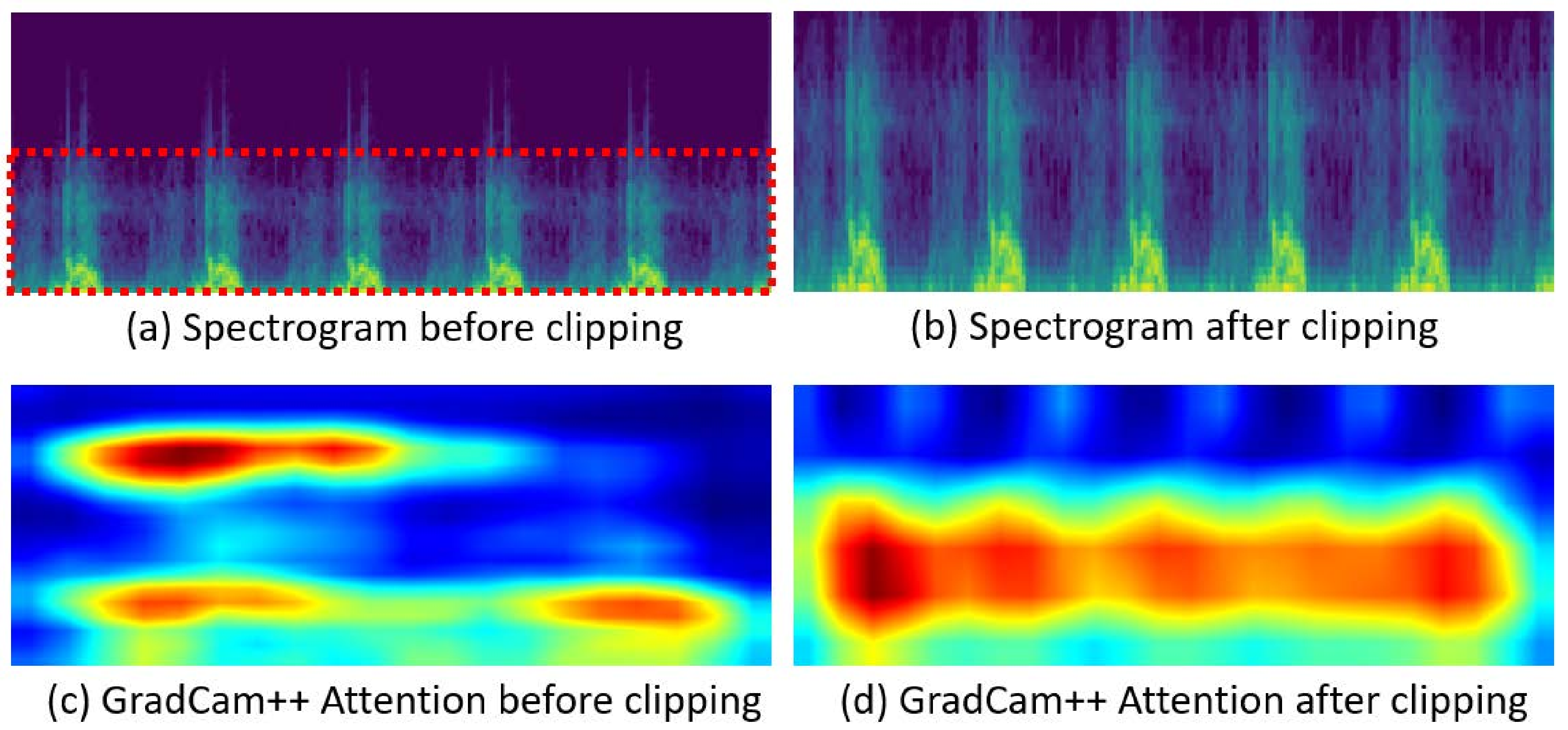
Blank regions in the spectrograms create false edges and hurt network performance

GradCAM++ Visualization



Blank Region Clipping

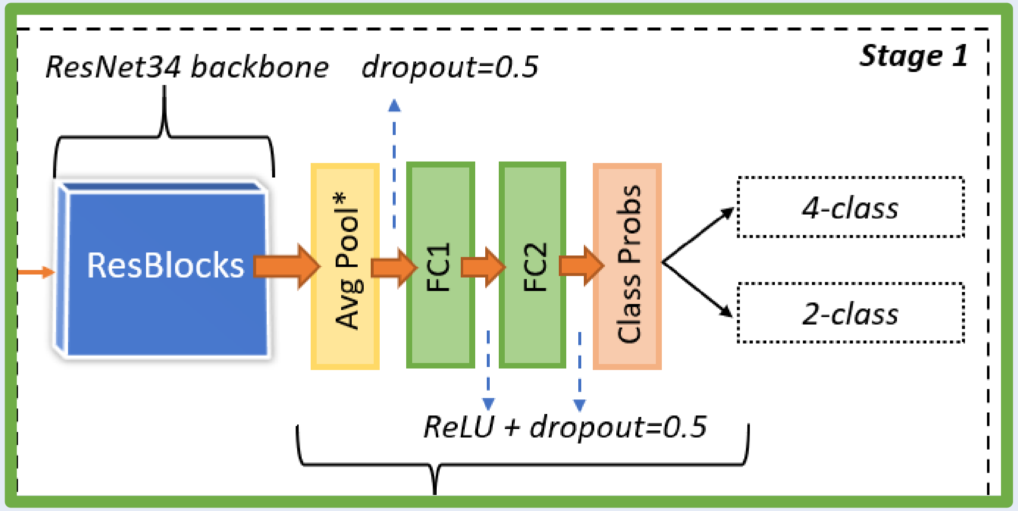
Selectively clip off blank regions



Network Training: Stage 1

ResNet34 Backbone with pre-trained ImageNet weights

- Categorical Cross-entropy loss
- Optimizer: SGD with momentum (=0.9)
- Batch-Size: 64
- Fixed LR: 1e-3
- Epochs: 200



Network Training: Stage 2

- ICBHI Dataset has samples from 4 different recording devices.
- Distribution of samples across devices is heavily skewed
  - Eg: AKGC417L Microphone contributes to 63% of samples
- DNN fails to generalize across devices given the small size of the dataset

Device	Patient Count*	N	C	W	B	Total
AKGC417L	32	1922	1543	500	381	4346
Meditron	64	1037	215	148	56	1456
Litt3200	11	347	77	126	44	594
LittC2SE	23	336	29	112	25	502

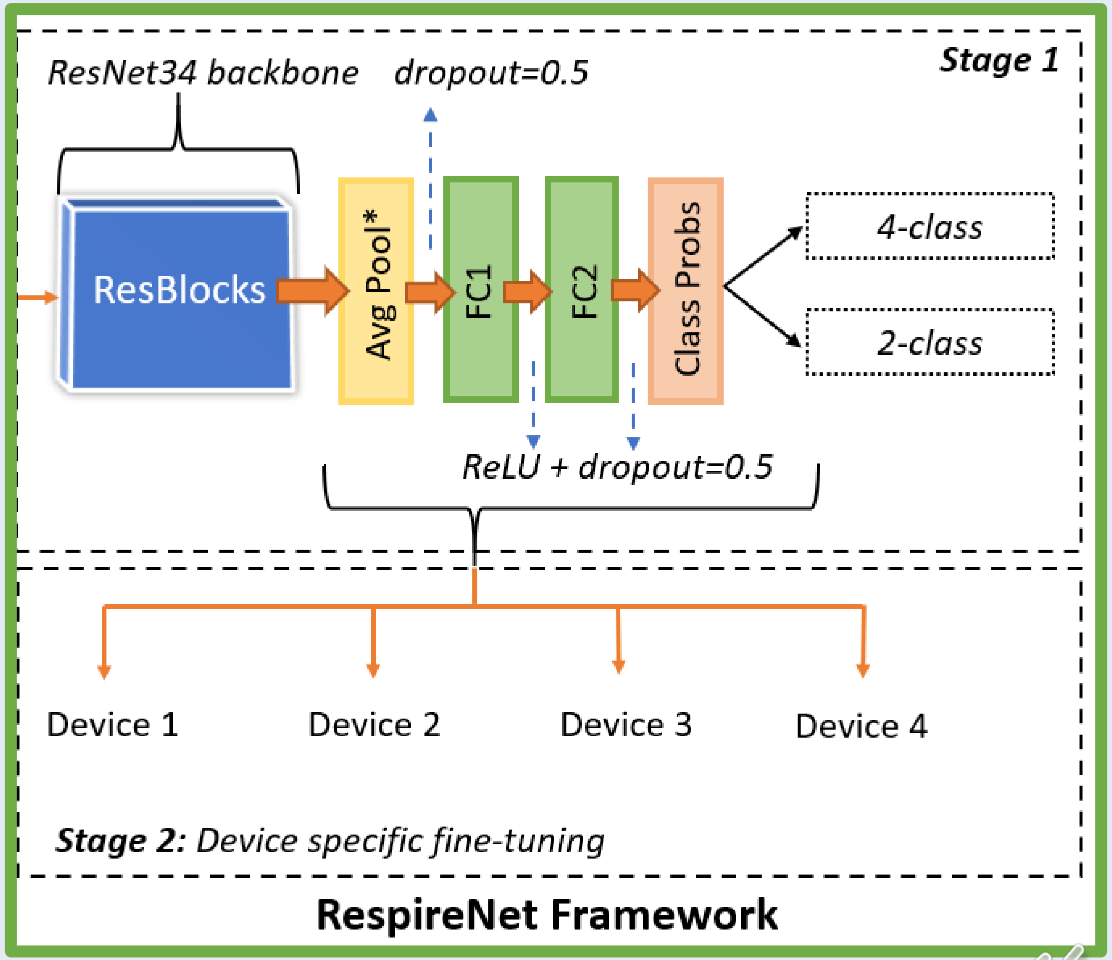
Breathing cycles across classes and devices

Network Training: Stage 2

Device specific fine-tuning

Fine-tune the model from **Stage 1** for each device separately

- LR: 1e-4
- Epochs: 50



Evaluation on ICBHI Dataset

4 Class Classification

- Classify into 4 classes: Normal, Crackle, Wheeze, Both

$$Senstivity = \frac{P_c + Pw + Pb}{N_c + Nw + Nb} \; ; \; Specificity = \frac{P_n}{N_n}$$

- P_i and N_i are the number of correctly classified and total number of samples in class i, respect. (*where i in {normal, crackle, wheeze, both}*)

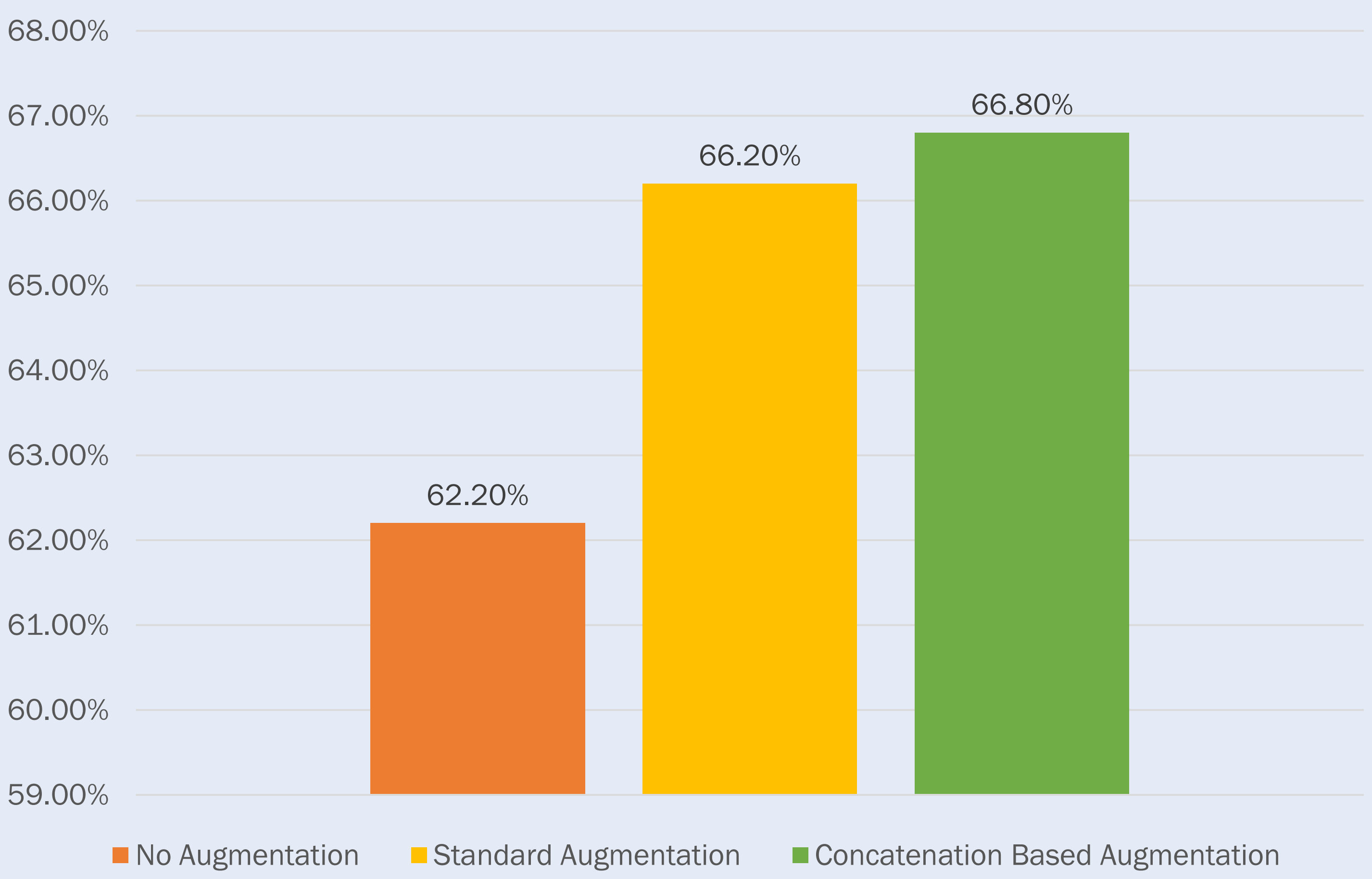
2 Class Classification

- Classify into 2 classes: Normal, Abnormal (Crackle/Wheeze/Both)

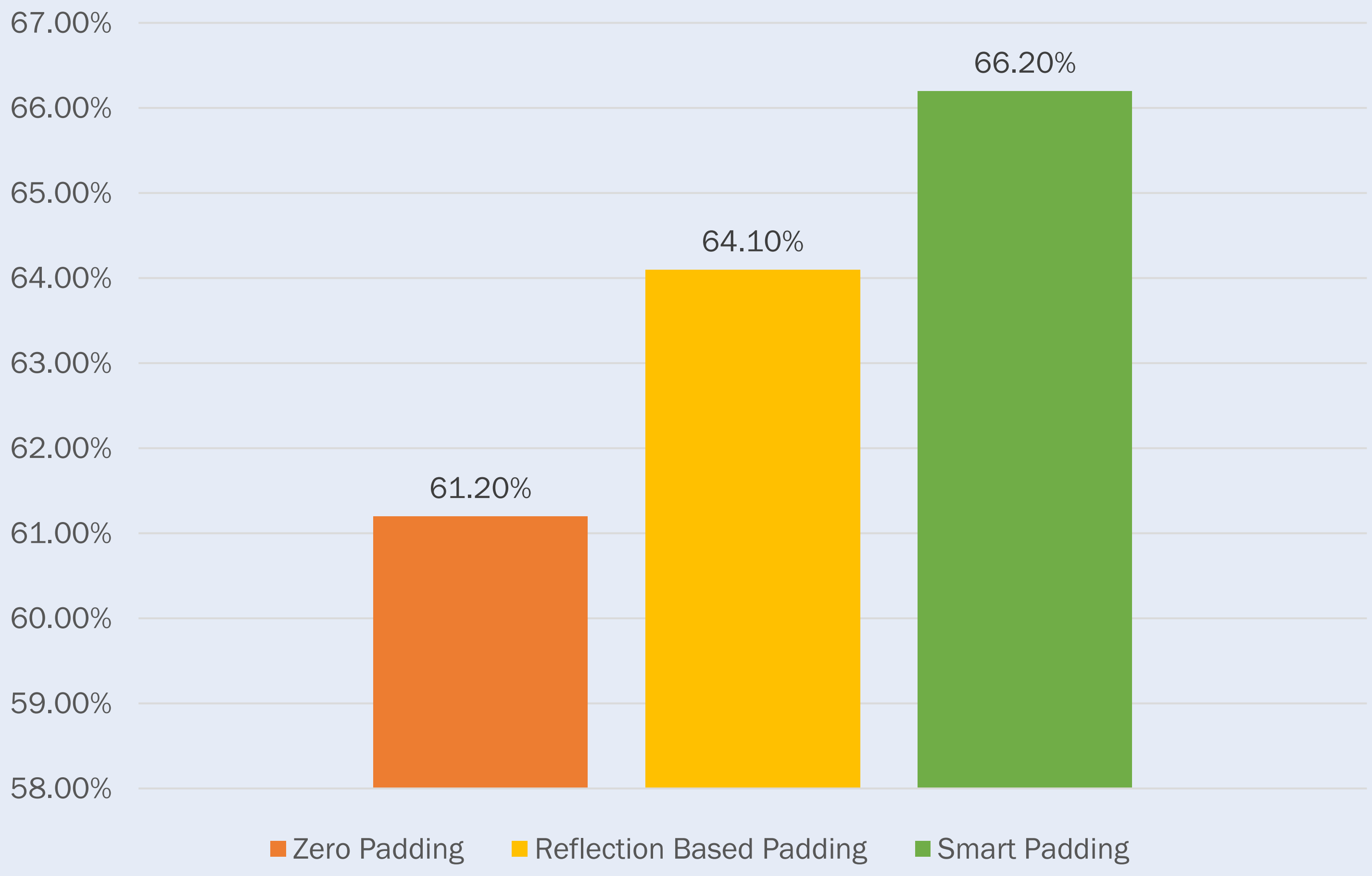
Results

Split & Task	Method	S_p	S_e	Score
60/40 Split & 4-class	Jakovljevic et al. [8]	-	-	39.5%
	Chambres et al. [4]	78.1%	20.8%	49.4%
	Serbes et al. [24]	-	-	49.9%
	Ma et al. [11]	69.2%	31.1%	50.2%
	Ma et al. [12]	63.2%	41.3%	52.3%
	CNN (ours)	71.4%	39.0%	55.2%
	CNN+CBA+BRC (ours)	71.8%	39.6%	55.7%
	CNN+CBA+BRC+FT (ours)	72.3%	40.1%	**56.2%**
80/20 Split & 4-class	Kochetov et al. [9]	73.0%	58.4%	65.7 %
	Acharya et al. [1]	84.1%	48.6%	66.3%
	Ma et al. [12]	64.7%	63.7%	64.2%
	CNN (ours)	78.8%	53.6%	66.2%
	CNN+CBA+BRC (ours)	79.7%	54.4%	67.1%
	CNN+CBA+BRC+FT (ours)	83.3%	53.7%	**68.5%**
80/20 Split & 2-class	CNN (ours)	83.3%	60.5%	71.9%
	CNN+CBA+BRC (ours)	76.4%	71.0%	73.7%
	CNN+CBA+BRC+FT (ours)	80.9%	73.1%	**77.0%**

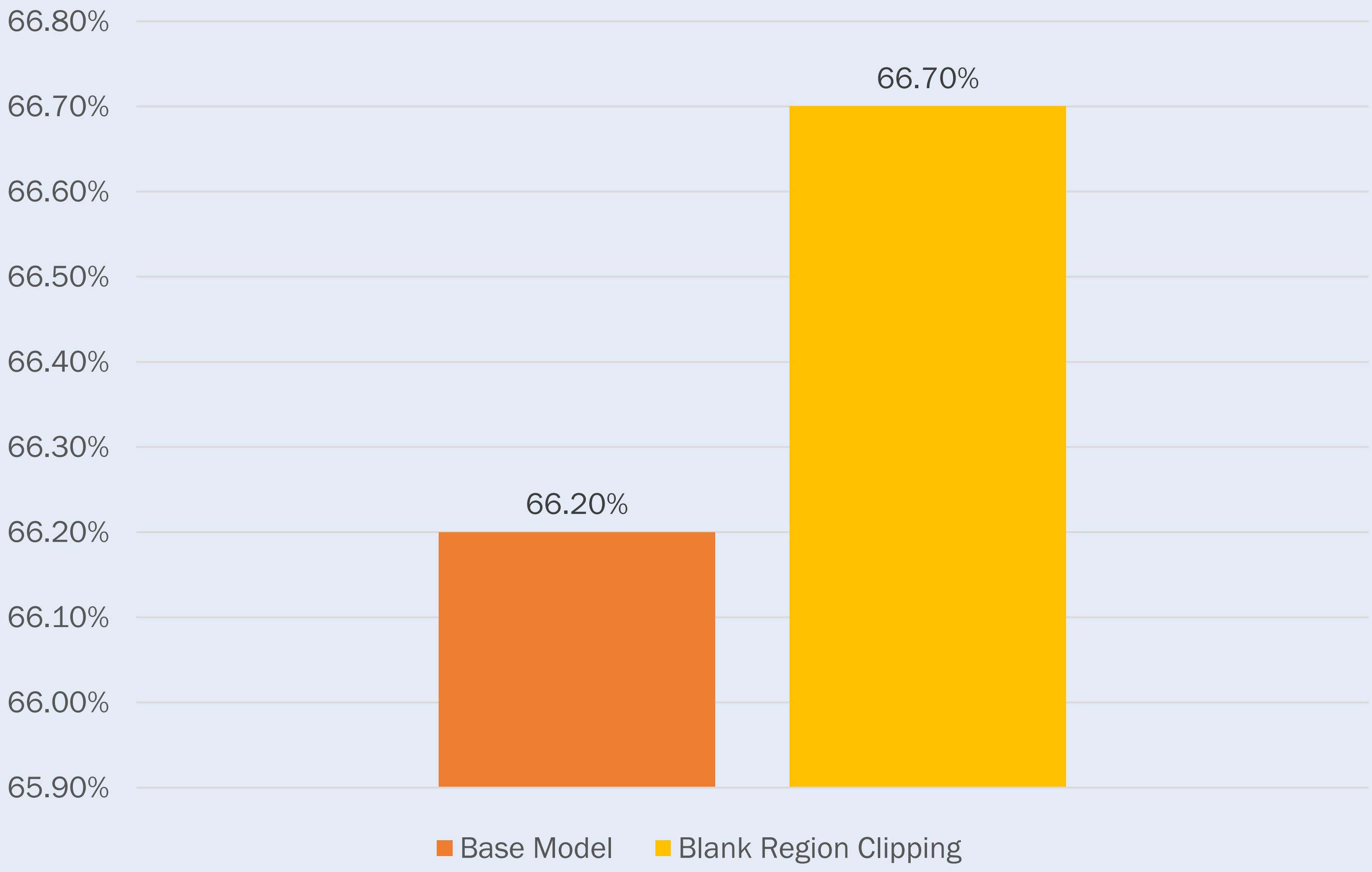
Ablations: Augmentations



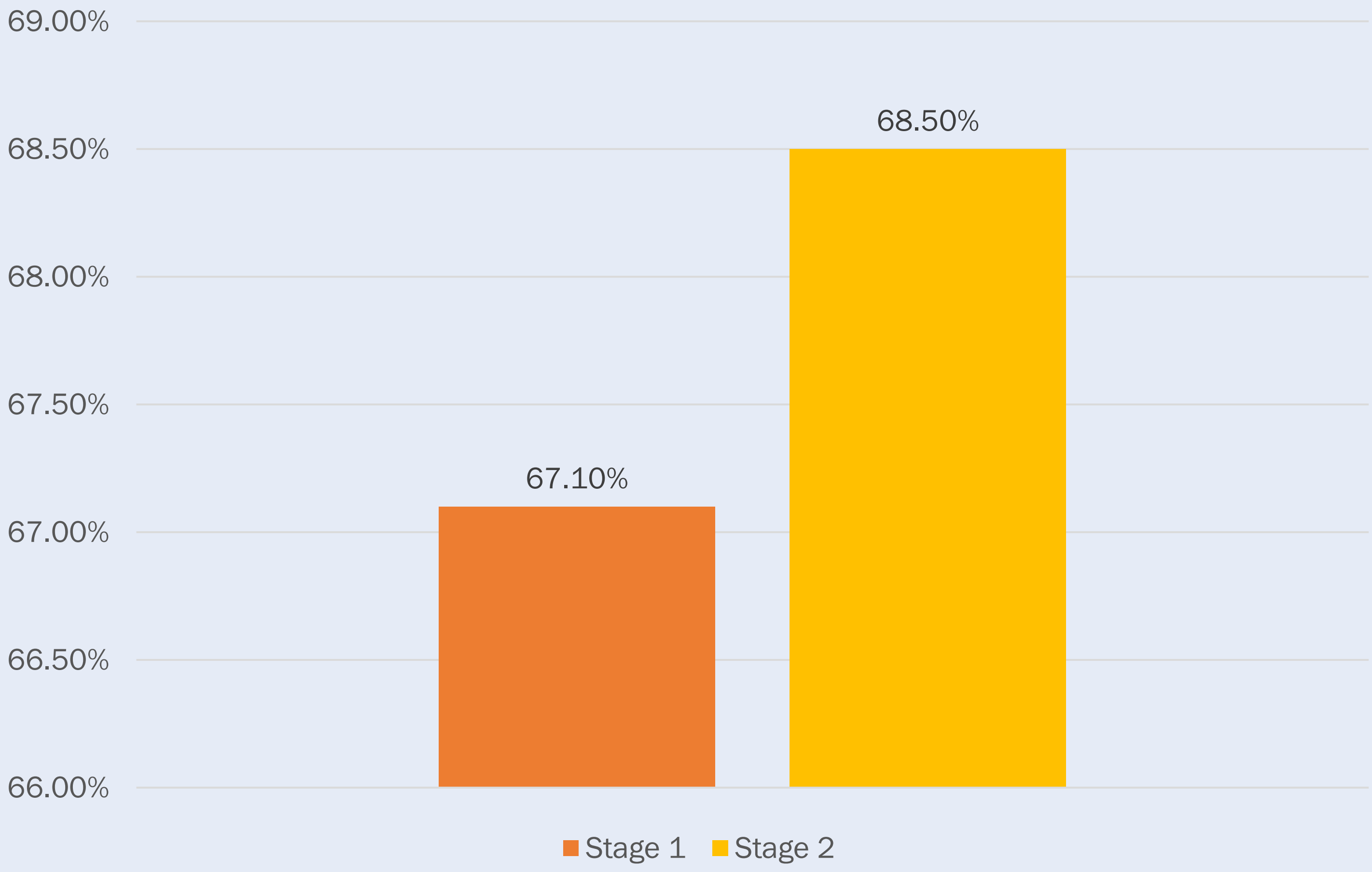
Ablations: Smart Padding



Ablations: Blank Region Clipping



Ablations: Device Specific Fine- tuning



Conclusion

RespireNet a simple CNN-based model, with a suite of novel techniques to utilize small-sized ICBHI dataset.

- Concatenation Based Augmentation
- Smart Padding
- Blank Region Clipping
- Device-Specific Fine Tuning

Thank you ☺