On-device Machine Learning for Digital Healthcare: The Case of Sleep Medicine

Hyung-Sin Kim

Seoul National University
Graduate School of Data Science

Introduction

Why is Sleep Important?

We spend approximately one-third of our lifespan sleeping

Inadequate sleep

- **Increases the risk of accidents and human errors.**
  (e.g. Motor-vehicle crashes, workplace incidents)
- **Lead to various chronic health issues and mental problems.**
  (e.g. cardiovascular disease, diabetes, stroke, obesity, depression)

Obstructive Sleep Apnea (OSA)

Open airway | Partially blocked airway | Completely blocked airway

Normal | Hypopnea | Apnea

$$\textbf{Apnea Hypopnea Index (AHI)} = \frac{\text{Apneas + Hypopneas}}{\text{Total sleep time (hours)}}$$

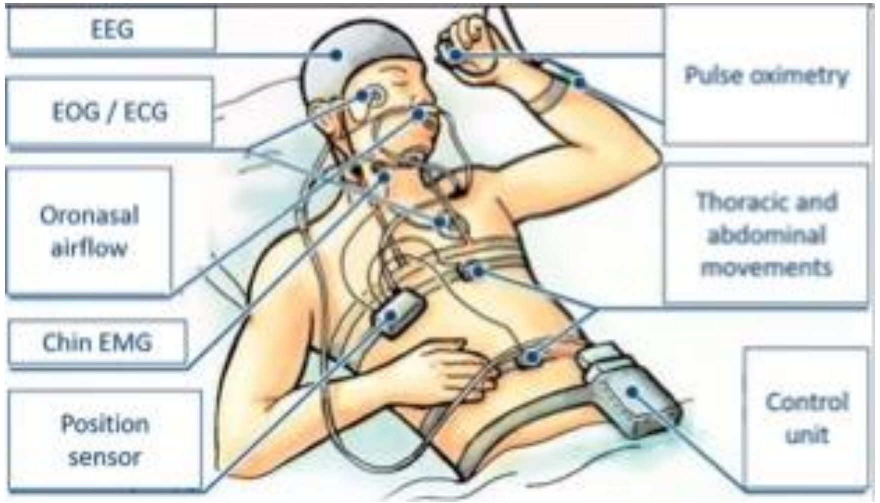
AHI	Rating
< 5	Normal
5 – 15	Mild OSA
15 – 30	Moderate OSA
> 30	Severe OSA

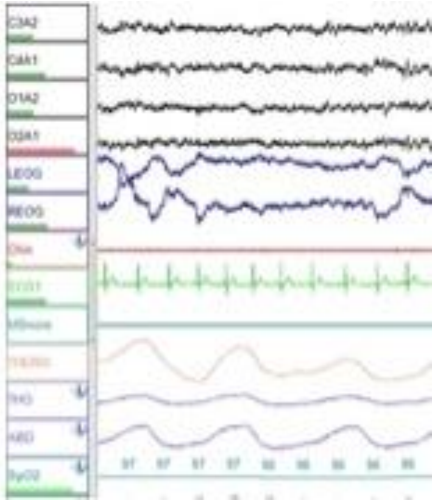
Obstructive Sleep Apnea (OSA)

Early diagnosis of OSA is challenging

- Apnea and Hypopnea often accompany snoring
- Patients with OSA often struggle to **recognize symptoms** such as **snoring** and **breathing cessation on their own.**

OSA Diagnosis: Polysomnography (PSG)





- Limitations of PSG

  1) First night effect (incorrect samples)

  - The discomfort of sleeping with attached sensors in an unfamiliar environment causes the first-night effect

  2) Single-night stay (insufficient samples)

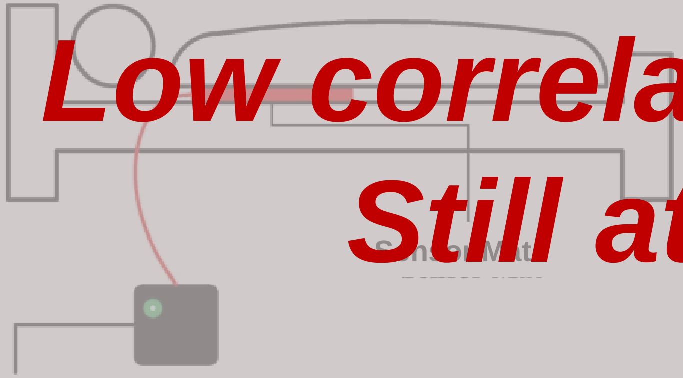
  - The variability of respiratory events results in substantial variation in OSA severity from night to night.

OSA Diagnosis: Alternative Attempts

Monitor daily sleep while minimizing sensor contact for a non-intrusive sleep

Low correlation and accuracy with PSG

Still attaching some sensors



IoT sensors

Wearable device

Snoring sound

OSA Diagnosis: Infrared Video

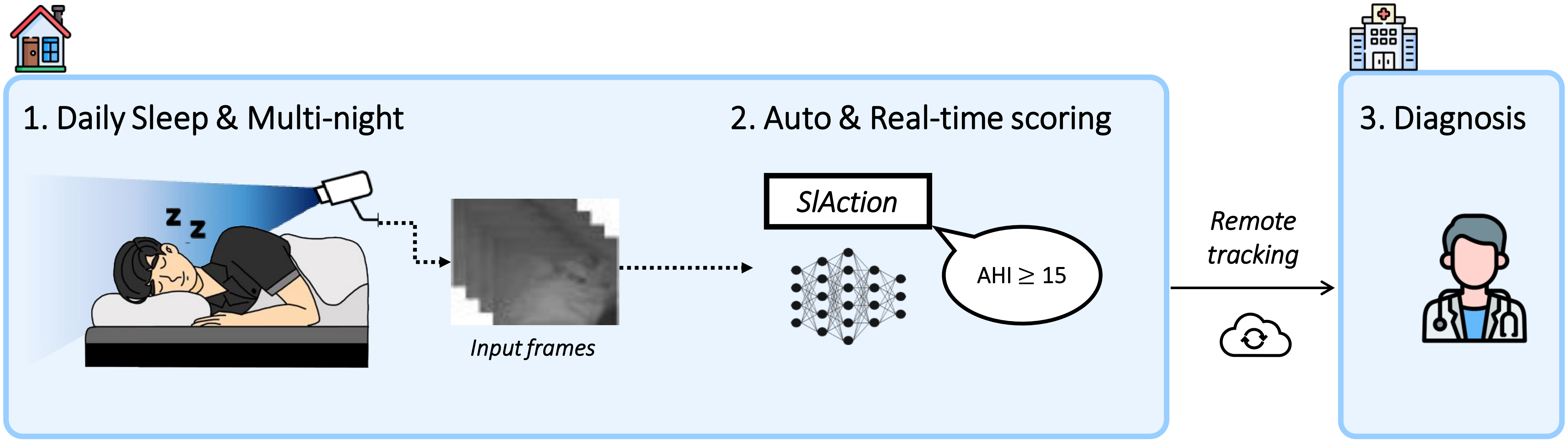
SlAction

Non-contact on-device daily sleep monitoring system for OSA diagnosis

✓ Subject can sleep without sensors

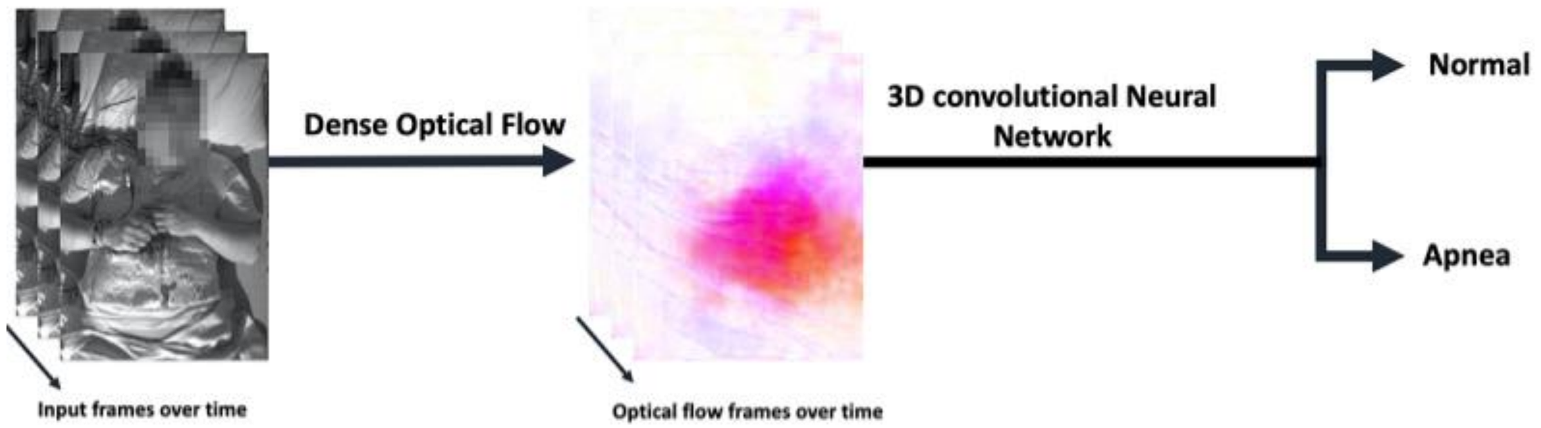
✓ Available anywhere and familiar with users (not expensive)

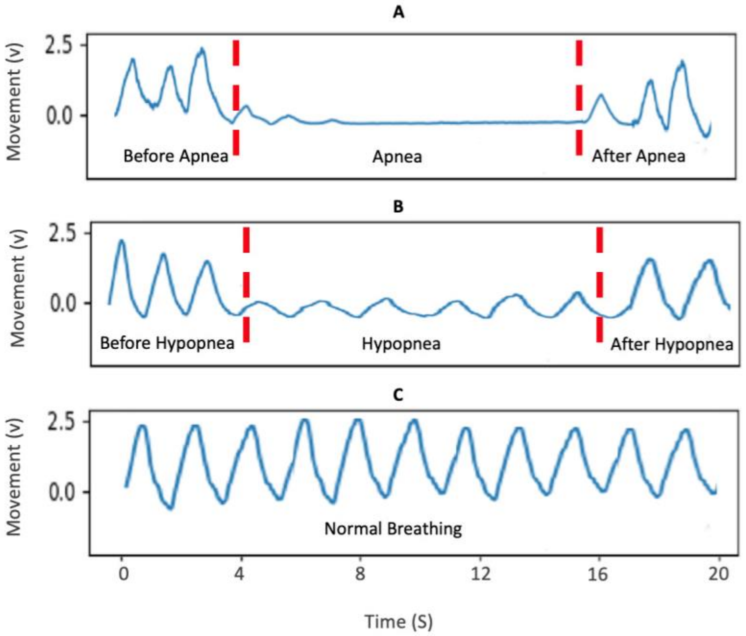
Application Scenario



Previous Work

- Akbarian *et al*., *J. Med. Internet Res*., 2021
  - 3D CNN architecture was used to process movements extracted by optical flow to detect respiratory events



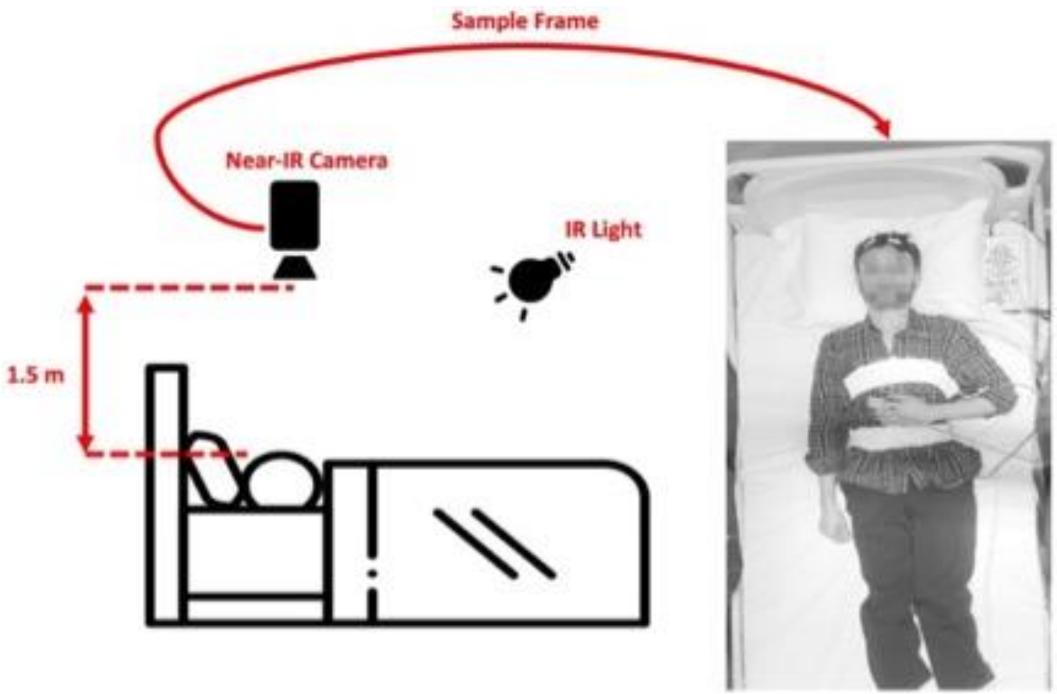


83% accuracy and an F1-score of 86%

Akbarian et al., Noncontact Sleep Monitoring With Infrared Video Data to Estimate Sleep Apnea Severity and Distinguish Between Positional and Nonpositional Sleep Apnea: Model Development and Experimental Validation, *J. Med. Internet Res.*, 23 (11), 2021

Previous Work: Limitations

- ~20 hours to process 5-hours video (**Inference every 0.5s with heavy model**)
- Various environmental constraints
- The algorithm was evaluated only on data from 41 patients
  - including 26 men and 15 women with a mean age of 53 (std 13), BMI of 30 (std 7), AHI of 27 (std 31) events/hour



Sleep Video Dataset

Preliminary Study
(with Clinical Expertise)

Method (*SlAction*)

Sleep Video Dataset

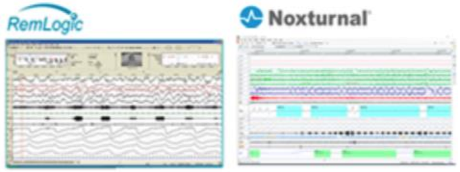
Preliminary Study
(*with Clinical Expertise*)

Method (*SlAction*)

Sleep Video Dataset

SNUH SEOUL NATIONAL UNIVERSITY HOSPITAL

- We collected infrared video of a patient sleeping during PSG
  - Includes corresponding signal data and labels (e.g., Sleep stage, various sleep events)
  - De-identification of personal attributes, including facial features, tattoos, etc
  - 1,000 patients from 4 clinics
  - 5-8 hours per video, 640 x 480 size, 5fps



Case Info
- Basic information about the examination

Report
- A summary of the examination results

Video Info
- Video timestamp synchronization information

Event
- Labeled sleep-related events that have been interpreted

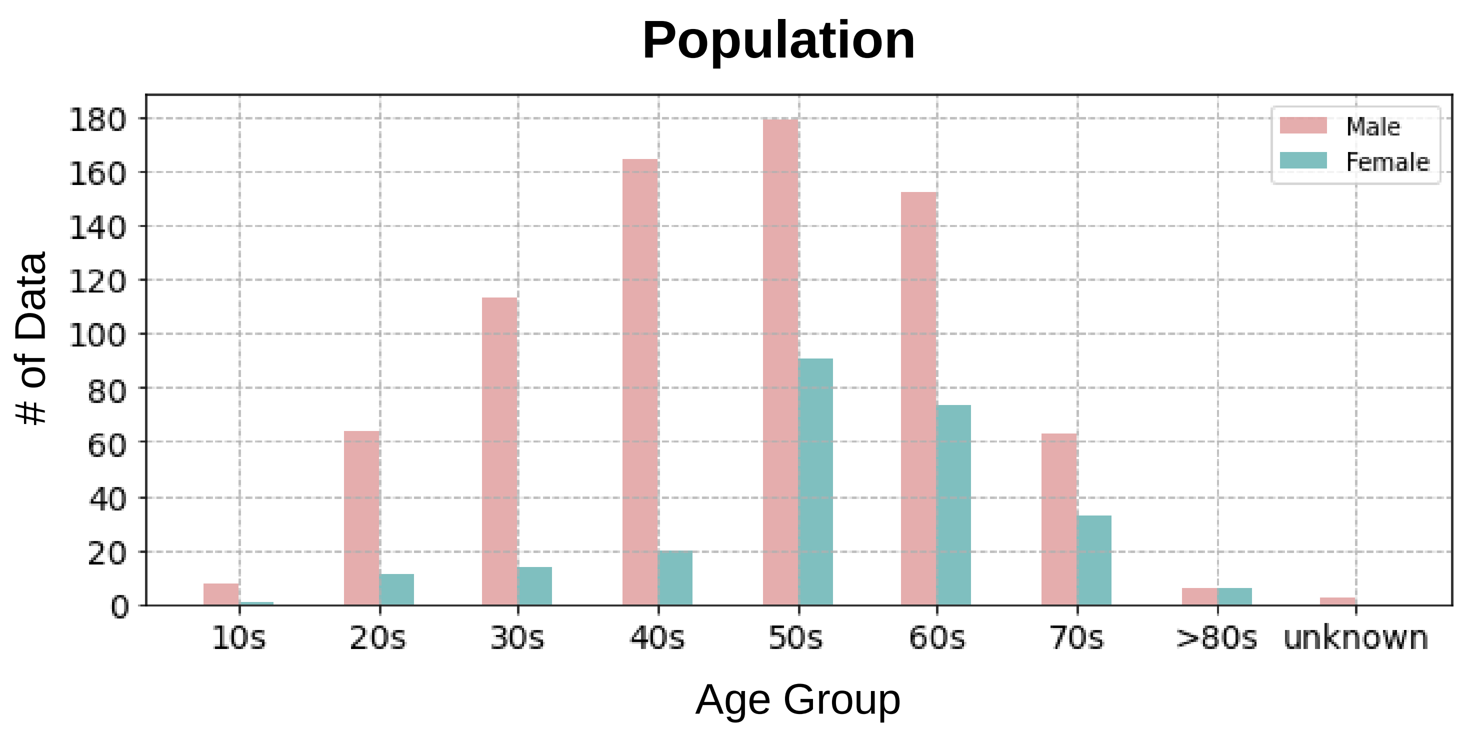

```
{
    "Case_Info": {
        "Case_Number": "A2019-EM-01-0001",
        "Year": 2019,
        "Start_Time": "2019/01/03 21:24:00.000",
        "Analysis_Start": {
            "Start_Time": "2019/01/03 21:24:00.000",
            "Start_Epoch": 1
        }
    },
    "Report": {
        "Sex": "Female",
        "Age": 60,
        "BMI": 26.1,
        "Time_in_Bed(TIB)": 409.5,
        "Total_Sleep_Time(TST)": 347.0,
        "Sleep_Efficiency": 84.7,
        "Sleep_Latency": 4.0,
        "REM_Latency": 80.5,
        ... 중략
        "Total_LM_Arousal_Index": 0.0,
        "Spontaneous_Arousal_Index": 8.6,
        "Total_Arousal_Index": 3.0
    },
    "Video_Info": [
        {
            "File_Name": "A2019-EM-01-0001_video_01.mp4",
            "File_Extension": "mp4",
            "Frame_Rate": 4.995,
            "Frame_Count": 122728.0,
            "Start": "2019/01/03 21:24:00.000",
            "End": "2019/01/04 04:13:30.000",
            "Bit_Rate": 280129.0,
            "Width": 640,
            "Height": 480
        }
    ],
    "Event": [
        {
            "Event_Number": 0,
            "Event_Label": "Wake",
            "Start_Time": "2019/01/03 21:24:00.000",
            "End_Time": "2019/01/03 21:24:30.000",
            "Start_Epoch": 1,
            "End_Epoch": 2,
            "Duration(second)": 30.0
        },
```

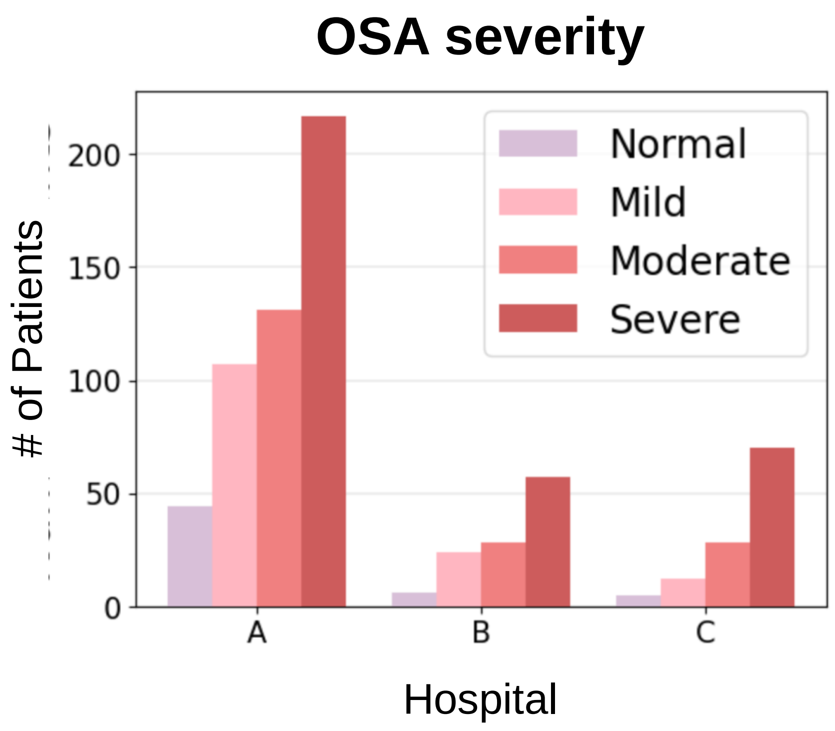

{;} .JSON

Sleep Video Dataset

	Hospital A	Hospital B	Hospital C
Number of patients	499	115	115
Shooting angle	30 degrees	45 degrees	45 degrees
Distance	3 m	3.5 m	3 m
Frame rate	30 FPS	30 FPS	10 FPS

Sleep Video Dataset





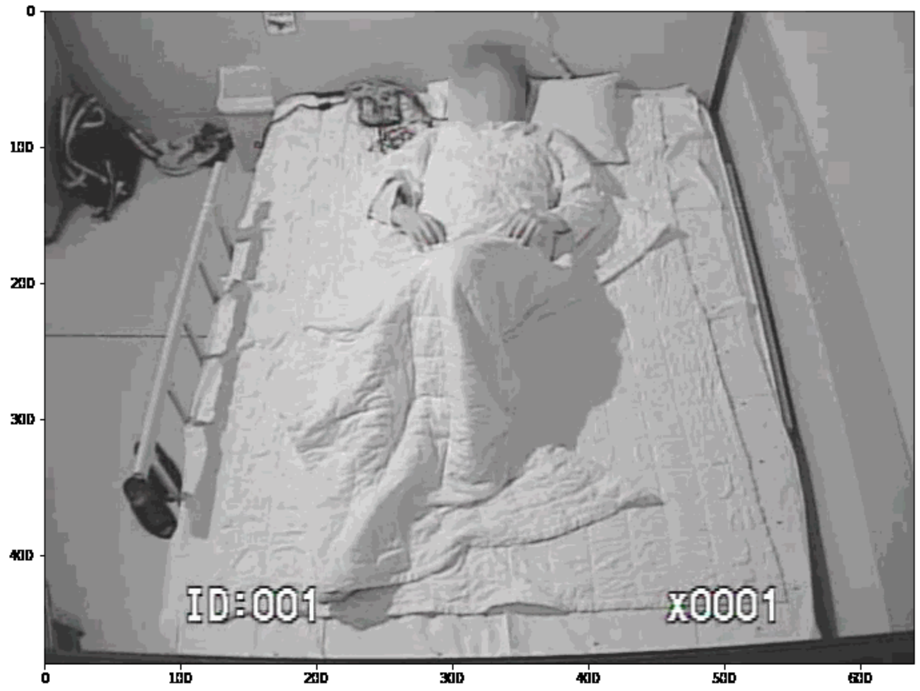
Challenges

- Beyond human perception

  - Sleep videos capture the most <u>inactive moment of human beings</u>

  - Poor quality and various noise that hinders motion differentiation (blurred face, bedding)

Normal



Apnea

Sleep Video Dataset

Preliminary Study

(*with Clinical Expertise*)

Method (*SlAction*)

Key Insight from Clinical Expertise

How about focusing on the movements related to ***respiratory arousal (RA)?***

Respiratory Arousal (RA)

- Arousal event occurring within three seconds (or less) following or overlapping with an apnea/hypopnea event
- Accompanied by more substantial movements compared to the usual state of sleep

- Arousal with no preceding events is labeled as spontaneous arousal (SA)

Correlation between RA and OSA

- A linear correlation exists between AHI (OSA) and RAI (RA)

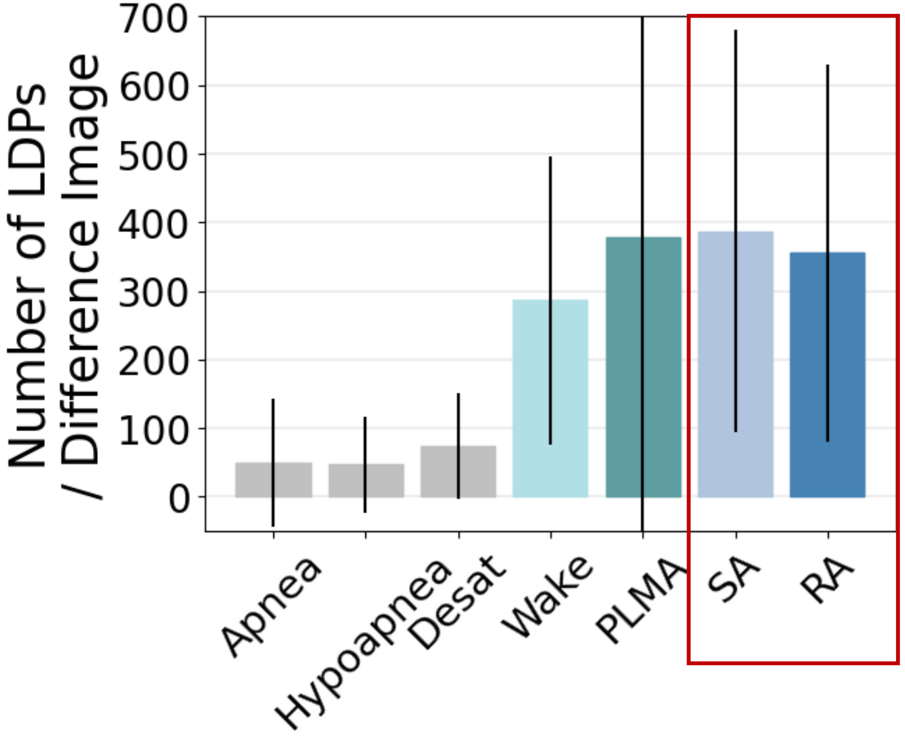
Problem Conversion

Indirectly detect OSA by identifying RA

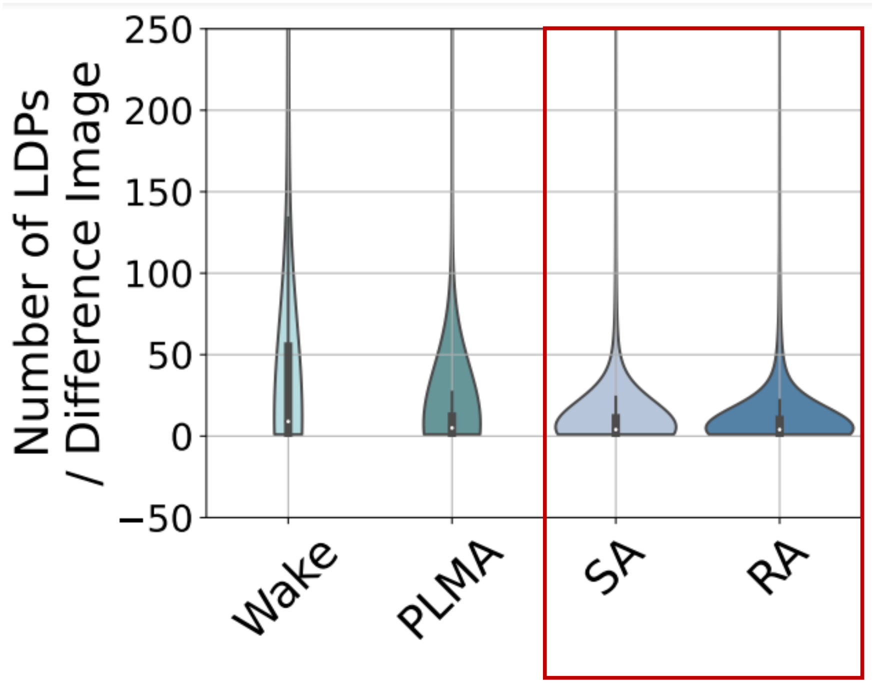
Proof-of-concept Study (1)

- Easily filtering out most sleep events

Analysis on movements during events



Analysis on movements preceding arousals

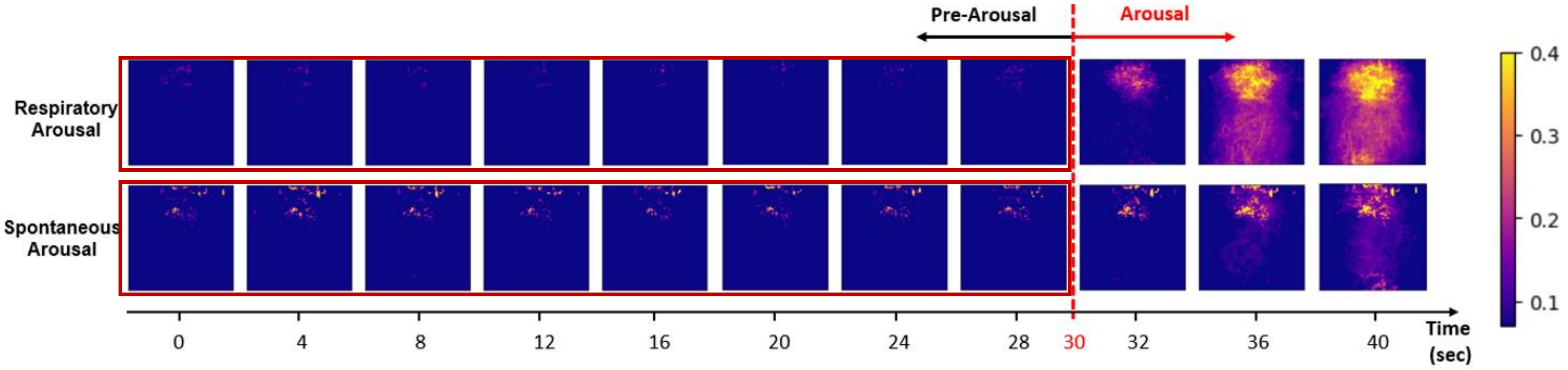


* LDP: Large difference pixels

PLMA: Periodic Limb Movement Arousal
SA: Spontaneous arousal
RA: Respiratory arousal

Proof-of-concept Study (2)

- Detailed analysis on time-series movements (RA vs. SA)

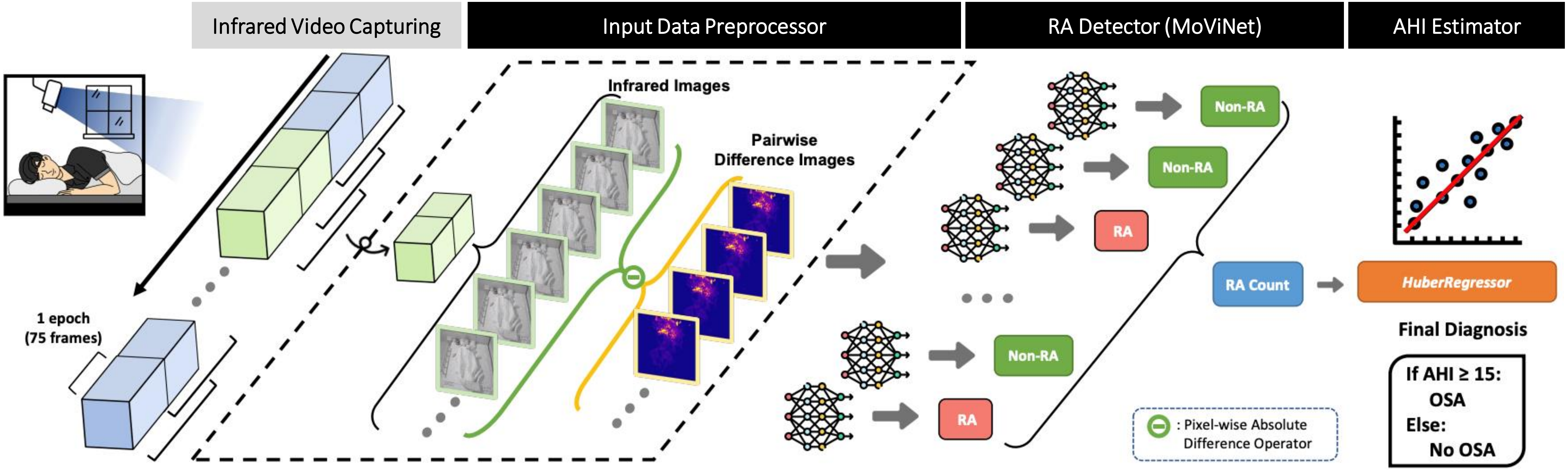


Sleep Video Dataset

Preliminary Study
(*with Clinical Expertise*)

Method (*SlAction*)

Overview



Infrared Video Capturing
Input Data Preprocessor
RA Detector (MoViNet)
AHI Estimator
Infrared Images
Pairwise
Difference Images
1 epoch
(75 frames)
Non-RA
Non-RA
RA
Non-RA
RA
RA Count
HuberRegressor
Final Diagnosis
If AHI ≥ 15:
OSA
Else:
No OSA
: Pixel-wise Absolute
Difference Operator

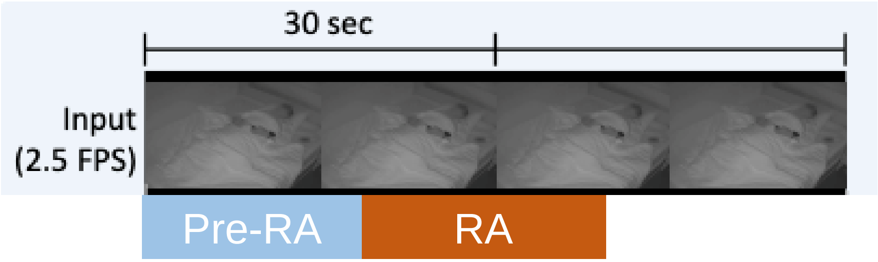
Input Data Preprocessor

Clip (Sliding Window) and Step Size Design

- To effectively differentiate RA, the input data need to include not only the movement patterns of RA but also the **patterns of event preceding RA.**
- **In sleep medicine…**

1) 30 seconds is commonly defined as one epoch, which serves as the fundamental unit for sleep analysis
2) Apnea and hypopnea events are labeled only when lasting for 10 seconds or longer
3) The average duration of RA is 14 seconds in our dataset
4) In each epoch, any arousal lasting for 15 seconds or more is considered as Wake, so RA does not exceed 15 seconds within a single epoch
5) RA events may span across two consecutive epochs

1 epoch



***1 minute** clip (input size)*
***30 seconds** step size*

Reduce the number of inference ~1/60

10 seconds clip
0.5 seconds step size

Input Data Preprocessor

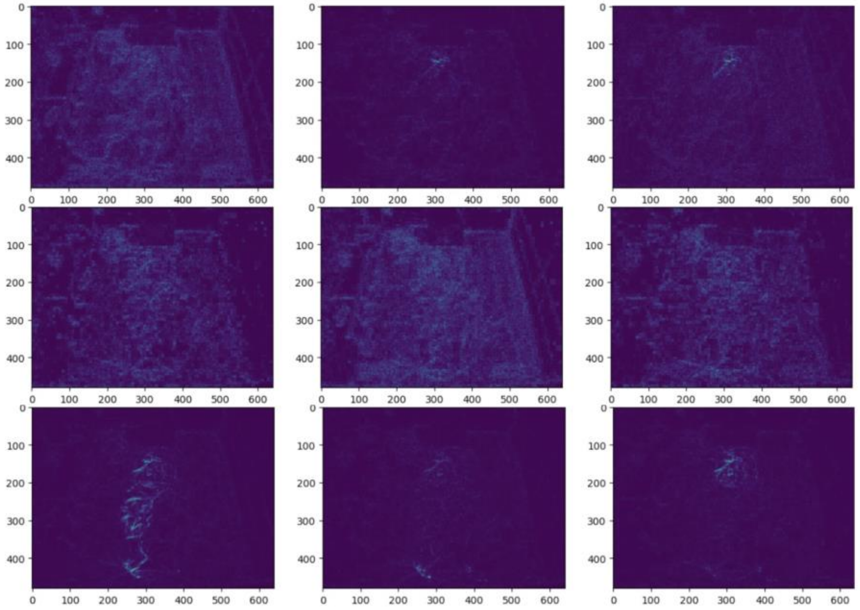
Frame Difference as Motion Input

(n+1)th frame nth frame

Normal Breath

Apnea

Respiratory Arousal



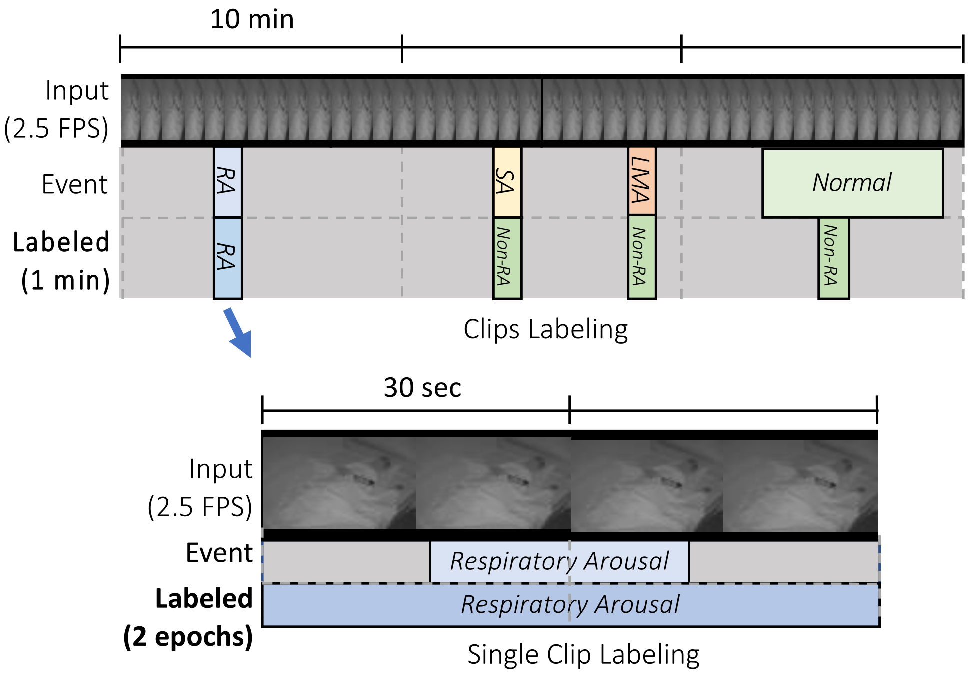
RA Detector

DNN Architecture: MoViNet

- Model: A0 (2+1D Convolution)

Training Dataset Curation

- To ensure that the extracted clips represent distinct temporal segments and enhance diversity of training data.

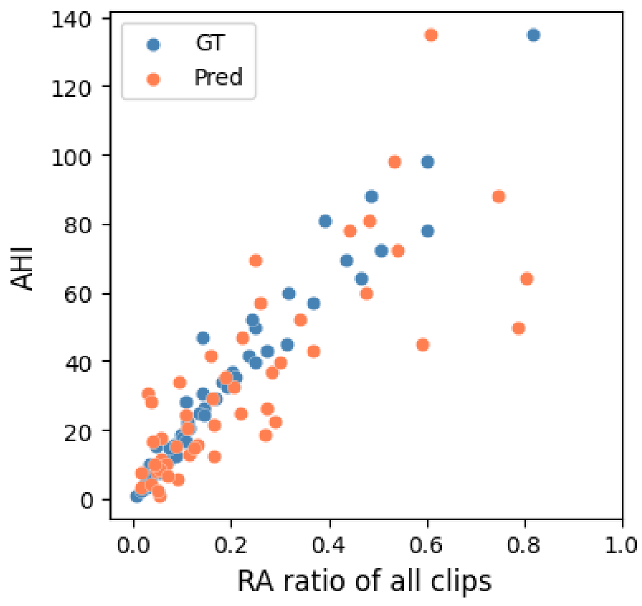


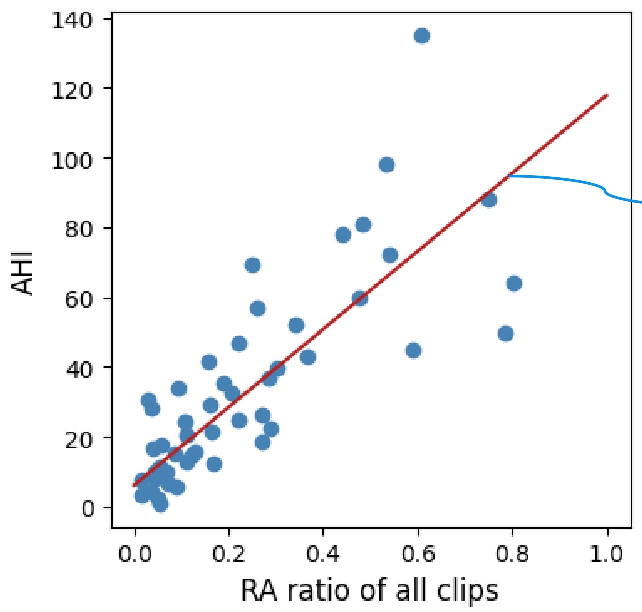
AHI Estimator

$$\text{AHI} = \frac{\text{Apneas + Hypopneas}}{\boxed{\text{Total sleep time (hours)}}}$$

Time in Sleep (TIS) excluding Wake stage

We can only ascertain Time in Bed (TIB) including Wake stage





Huber Regressor

Fitting with validset

Estimated AHI ≥ 15: OSA

(**RA ratio**: RA events during the entire sleep duration, divide it by TIB)

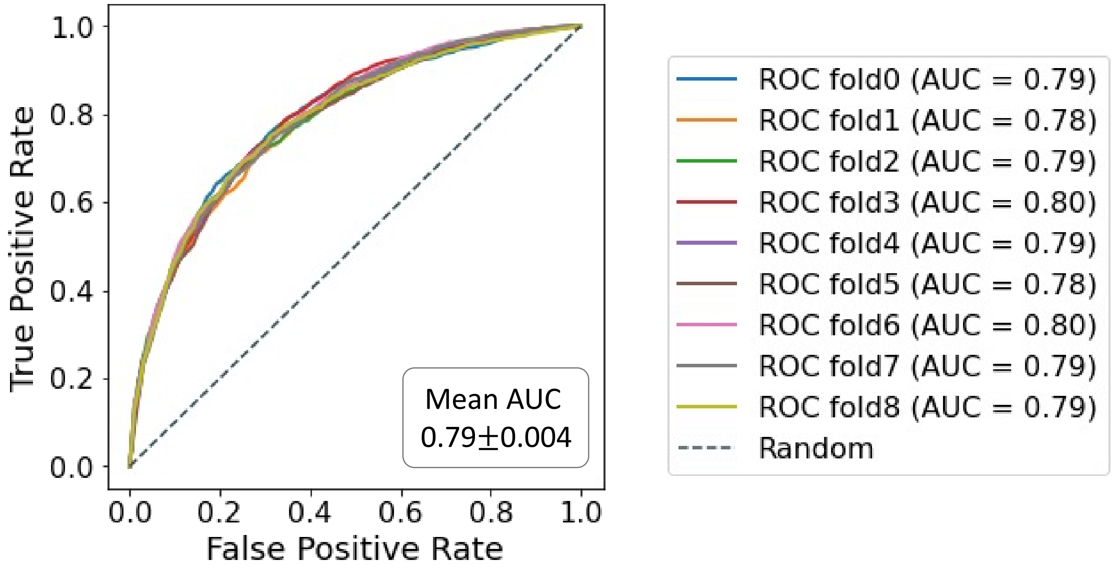
Evaluation

RA Detector

9-fold cross validation

- Dataset: Train 449 (Valid 50) / Test 50
- Metric: Area under the curve (AUC) of the Receiver Operating Characteristic (ROC) curve

AUC on testset



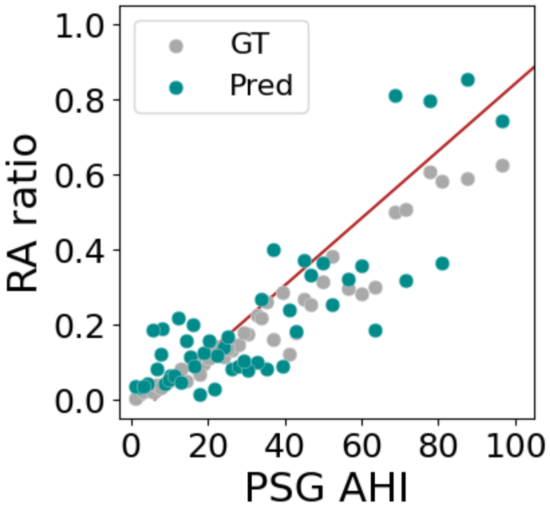
Target Event	Input Type	Input Size	AUC
RA	Frame Difference	60 sec.	**0.79**
RA	Frame Difference	30 sec.	0.62
RA	Original	60 sec.	0.58
Apnea-Hypopnea	Frame Difference	60 sec.	0.57

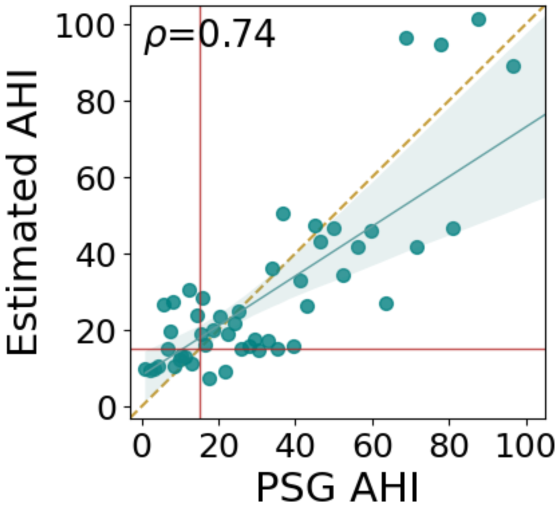
AHI Estimator & OSA Prediction

Estimated AHI vs. PSG AHI

- Metric: Spearman correlation analysis (rank-order correlation)

Dataset	Estimator fitting dataset	AHI Estimation Spearman correlation coefficient (ρ)	AHI Estimation P-value
A valid	A valid	0.827	1.37e-13
A test	A valid	0.744	5.89e-10



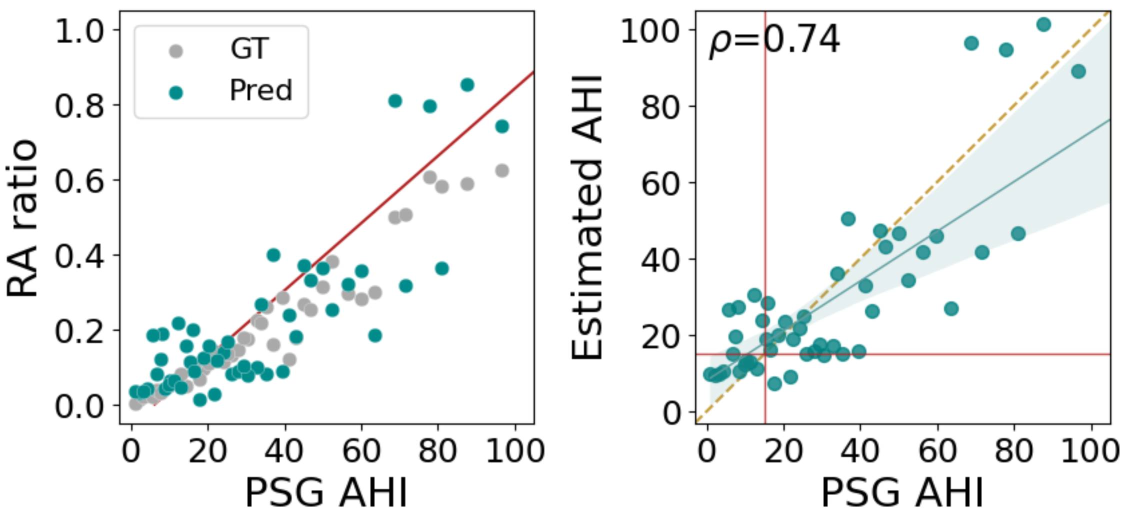


AHI Estimator & OSA Prediction

OSA Prediction

- Metric: Accuracy, Precision, Recall, and F1 Score

Dataset	Estimator fitting dataset	OSA Prediction Accuracy (%)	Precision	Recall	F1 Score
A valid	A valid	84.0	0.886	0.886	0.886
A test	A valid	82.0	0.842	0.914	0.876



AHI Estimator & OSA Prediction

Test Dataset

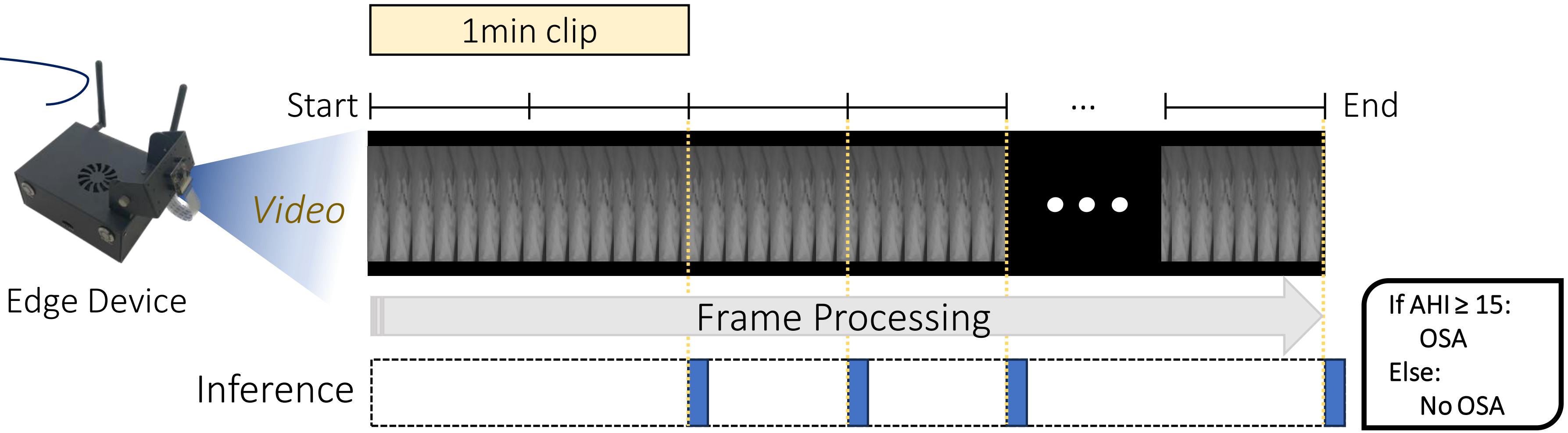
A	B	C	Total
50	115	80	245

Total Results

Dataset	Estimator fitting dataset	AHI Estimation Spearman correlation coefficient (ρ)	P-value	OSA Prediction Accuracy (%)	Precision	Recall	F1 Score
A valid	A valid	0.827	1.37e-13	84.0	0.886	0.886	0.886
A test	A valid	0.744	5.89e-10	82.0	0.842	0.914	0.876
B test	A valid	0.756	8.60e-23	83.4	0.867	0.918	0.891
C test	A & C valid	0.834	8.16e-22	83.7	0.924	0.884	0.903

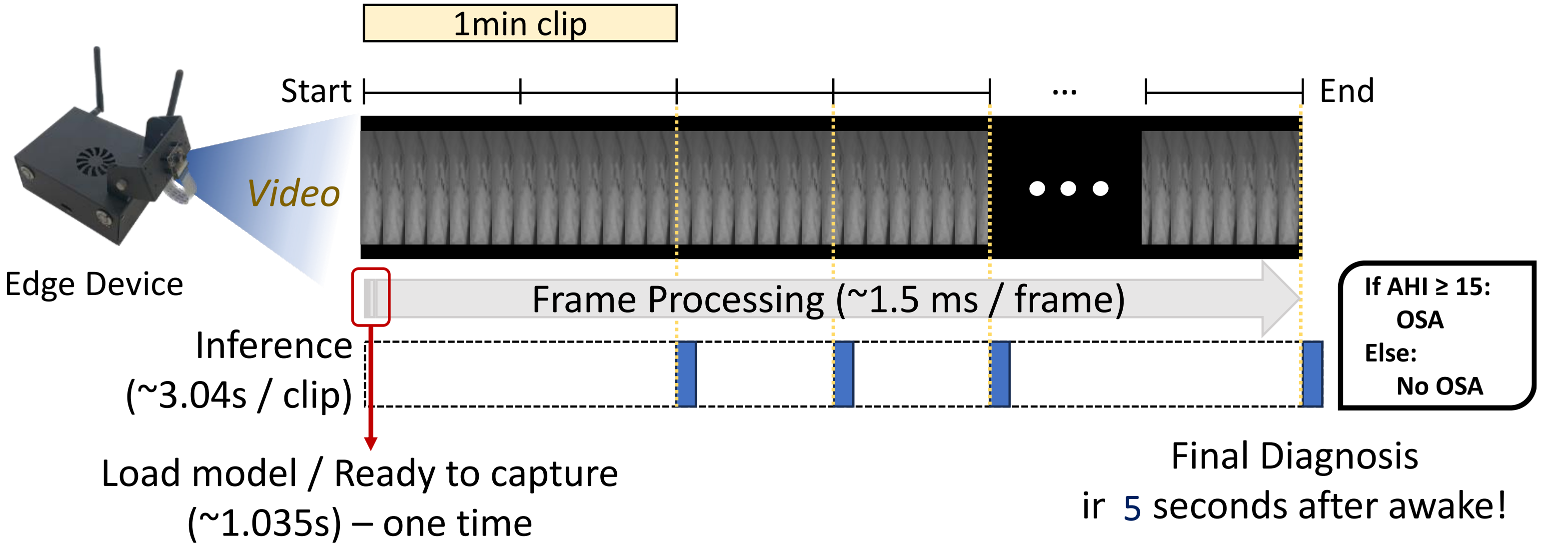
On-Device Inference Operation

NVIDIA Jetson Nano
- quard-core ARM A57 CPU
- 4GB 64-bit LPDDR4 memory



On-Device Performance

Model size (FP16)	Model load Frame capture ready	Frame processing (1 min. clip)	Inference	Total Operation	Peak Memory (RSS)	Peak Memory (Runtime)
5.1 (MB)	1.035±0.007 (s)	0.224±0.042 (s)	3.040±0.046 (s)	3.264±0.088 (s)	839±15.5 (MB)	2.67±0.016 (GB)



~16min. for 5h video
(75x speed up)

- TensorFlow Lite (16FP)
- XNN Pack (highly optimized library for FP NN inference operator, to utilize the CPU for the operation)

Conclusion

Comparison with Previous Work

Dataset

	# of case	# of institutions	Camera distance	Camera angle	fps
Previous (2021)	41	1	1.5 m	90 degree	2
***SIAction* (Ours)**	729 (**Test 245**)	3	3 / 3.5 m	30 / 45 degree	5

- **25 times more cases** from diverse environments and institutions.
- **Evaluate** the system on **a dataset more than 6 times larger.**

Comparison with Previous Work

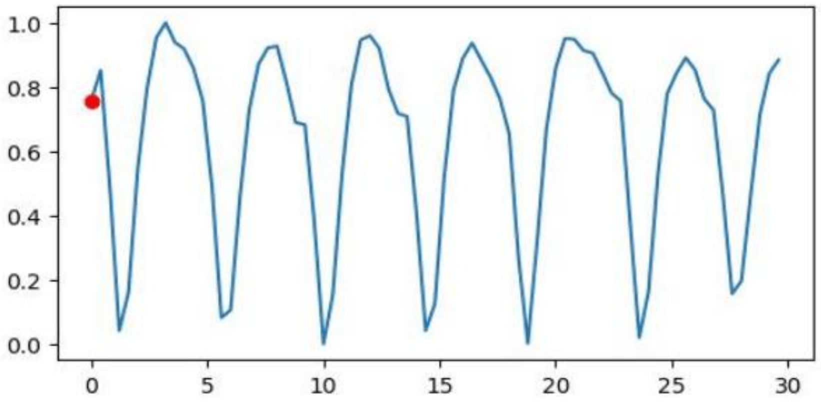
Method

	Target Event	Input				Model	Performance	Speed
		Type	Clip length	Sliding window step	# of inference for 5h. video		F1-score	Analysis time for 5h. Video
Previous (2021)	Apnea/Hypopnea	Optical flow	10 sec.	0.5 sec.	36,000	3D CNN (params: 8.2M)	**0.86**	**20h.** on GPU server
***SlAction* (Ours)**	Respiratory arousal	Frame difference	60 sec.	30 sec.	600	MoViNet (params: 2.5M)	**0.88**	**~16 min.** on CPU only edge device

- By closely collaborating **with sleep domain experts** to reframe the sleep apnea/hypopnea detection problem into a respiratory arousal detection problem.
- Effectively **design the input of the model by integrating knowledge from sleep medicine and data analysis results**, and successfully trained the model.
- **Even on a dataset 25 times larger,** achieving slightly **higher accuracy**.
- **Operating 75 times faster on low-spec CPUs** than previous work, as the model with **3 times fewer parameters** requires only **60 times fewer inference** counts.

Future work

- We are exploring ways to **enhance accuracy**, even if it results in a slightly longer runtime than currently achieved.
- Various learning techniques, such as domain adaptation for personalization, can be applied.
- **Research utilizing collected sleep video**
  - Extraction of respiratory patterns



  - Development of methodologies for diagnosing conditions like periodic limb movement disorder and REM sleep behavior disorder.

Thanks!

hyungkim@snu.ac.kr