AIMLab.

SleepPPG-Net2: Deep learning generalization for sleep staging from photoplethysmography

Shirel Attia[1], MSc candidate

Supervised by Dr. Joachim A. Behar[2]

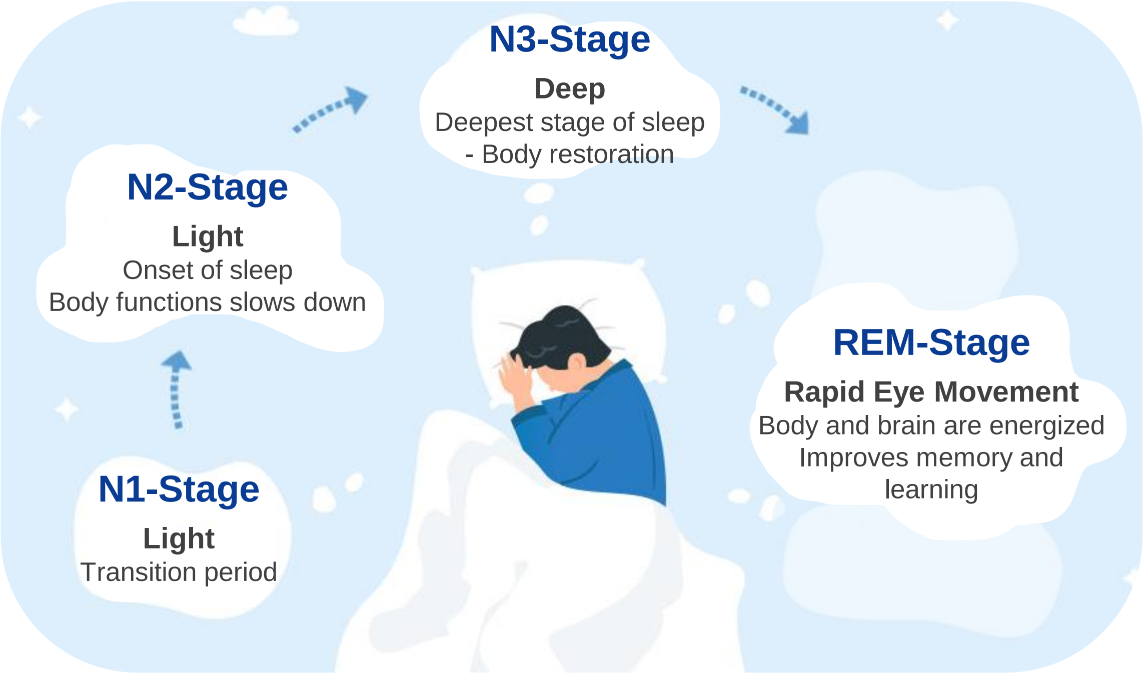
1. Faculty of Data and Decision Sciences, Technion, Israel Institute of Technology, Haifa, Israel
2. Faculty of Biomedical Engineering, Technion, Israel Institute of Technology, Haifa, Israel

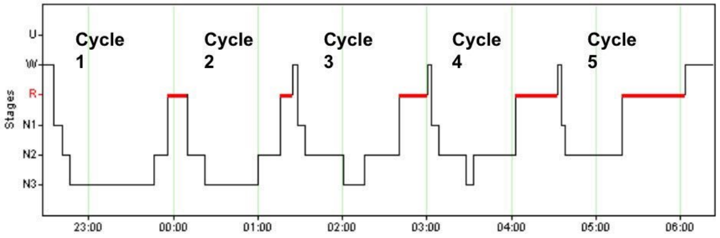
In collaboration with:

Technion: Angeleene Ang and Sharon Haimov.

Ichilov: Revital Shani Hershkovich, Alissa Tabakhov and Riva Tauman.

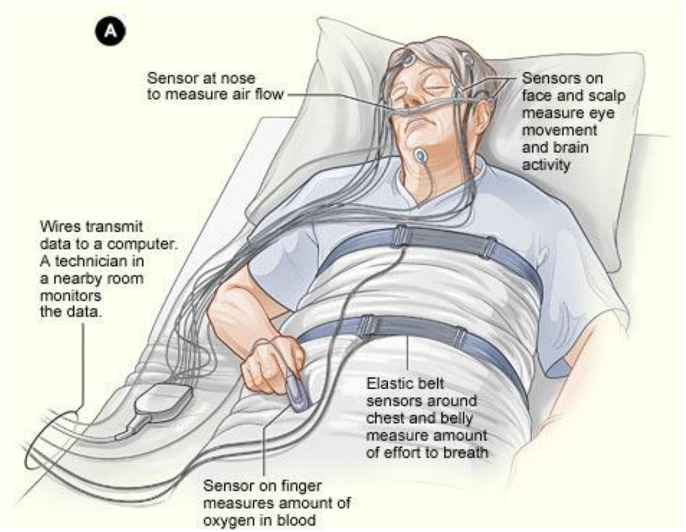
Motivation

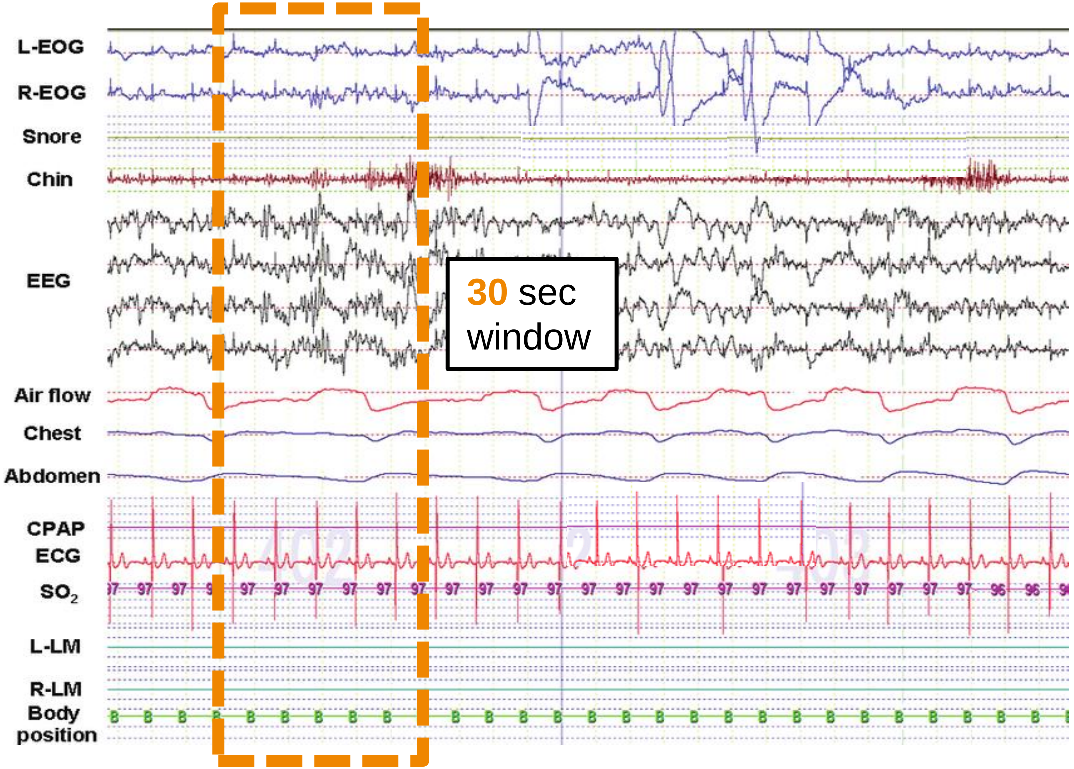




- Classification of these neural patterns into distinct stages facilitates the evaluation and analysis of sleep and allows detection of sleep disorders.

The current gold standard: polysomnography (PSG)





Picture from https://www.aboutmyclinic.com/dranilmadake/terms/Sleep-Studies-POLYSOMNOGRAPHY/4334, https://www.researchgate.net/figure/b-Representative-raw-waveform-on-polysomnogram-during-continuous-positive-airway-pressure_fig3_5256792

The current gold standard: polysomnography (PSG)

Cumbersome

Costly

Widely **unavailable**
(long waiting lists/delays
in receiving a diagnosis).

- There is also a need for the involvement of specialized technicians and doctors for scoring, further adding to the overall resource requirements and time constraints.

Picture from https://balanceepigeneticorthodontics.com/patient-resources/blog/what-s-a-polysomnography

Performance statistic

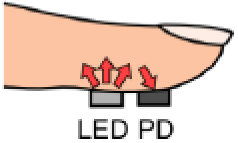
The performance metric used to evaluate sleep staging classification is called Cohen’s Kappa.
It is a statistical measure designed to quantify the level of agreement between two raters.
Cohen’s Kappa for human-labeled polysomnography of sleep staging is 0.76.

κ	Interpretation
<0	Poor agreement
0.01–0.20	Slight agreement
0.21–0.40	Fair agreement
0.41–0.60	Moderate agreement
0.61–0.80	Substantial agreement
0.81–1.00	Almost perfect agreement

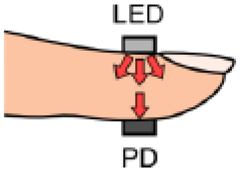
Lee, Yun Ji et al. “Interrater reliability of sleep stage scoring: a meta-analysis.” *Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine*

Photoplethysmography interest and remote health

Reflection



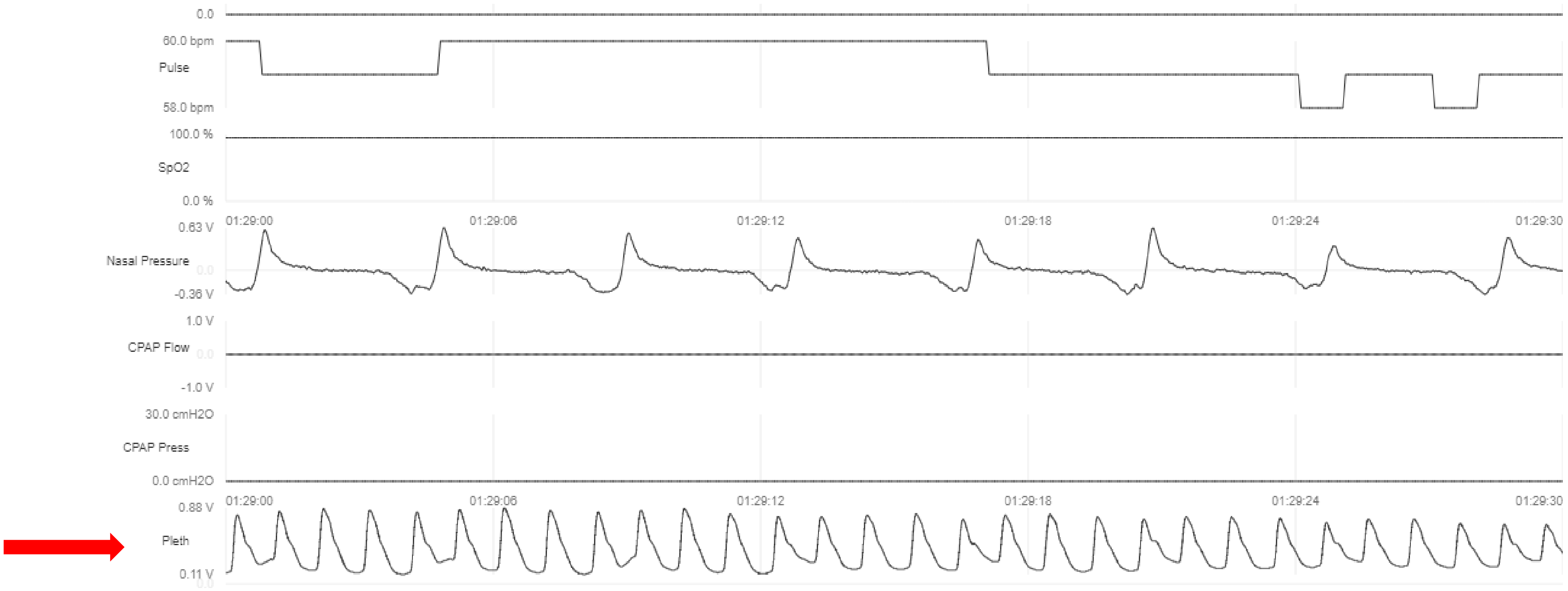
Transmission



- PPG is a non-invasive method that uses light to detect volumetric changes in the microvascular tissue bed.

Pictures from https://www.mdpi.com/2079-9292/3/2/282, https://www.nonin.com/

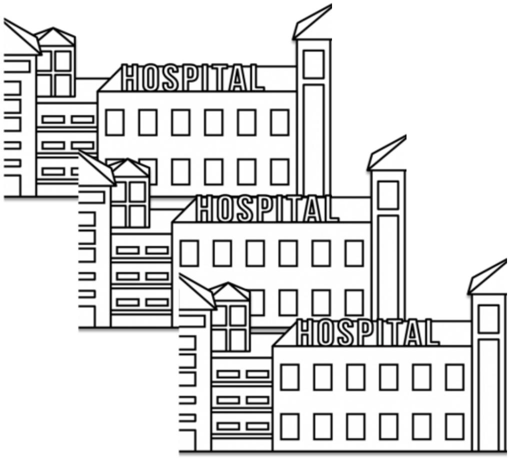
Photoplethysmography interest and remote health



Previous works using PPG

Work	Input	External validation	Acc	k
Korkalainen et al (2020)	raw PPG	No	68.5	0.54
Li et al (2021)	BRV from PPG + accelerometer	No	68.6	0.44
Radha et al (Radha 2021)	BRV from PPG	No	76.36	0.65
Wulterkens et al (2021)	BRV from PPG + accelerometer	No	76.4	0.62
Huttunen et al (2021)	raw PPG	No	74.1	0.64
Kotzen et al (2022)	raw PPG	Yes – on one dataset	84	0.67

Generalization Performance

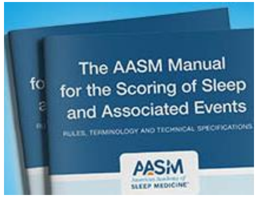


Hospital practice

Age

Sex

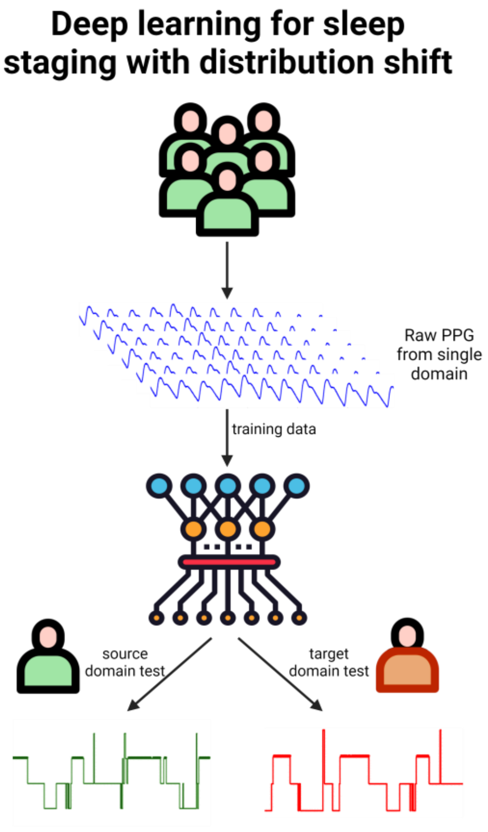
Comorbidity

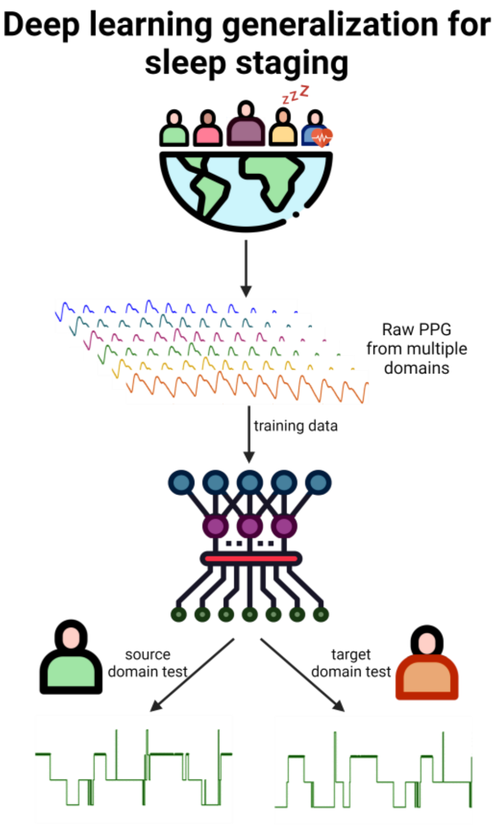


Medical guideline

Ethnicity

- Generalization refers to your model's ability to adapt properly to new, previously unseen data, drawn from the same distribution as the one used to create the model.


Deep learning for sleep staging with distribution shift
Raw PPG from single domain
training data
source domain test
target domain test


Deep learning generalization for sleep staging
zzZ
Raw PPG from multiple domains
training data
source domain test
target domain test

Hypothesis and goal

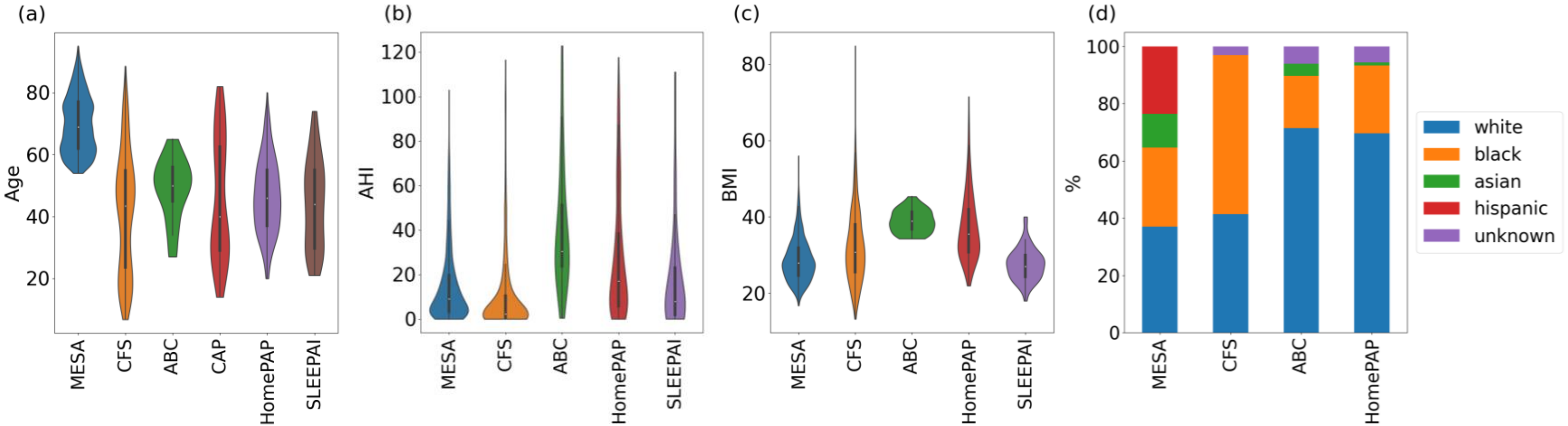
Hypothesis: employing a multi-source domain training approach can improve the model's robustness when applied to diverse datasets.

Goal: demonstrate the new model's consistent improvement over the previous model by utilizing multi-source domain training across all available datasets.

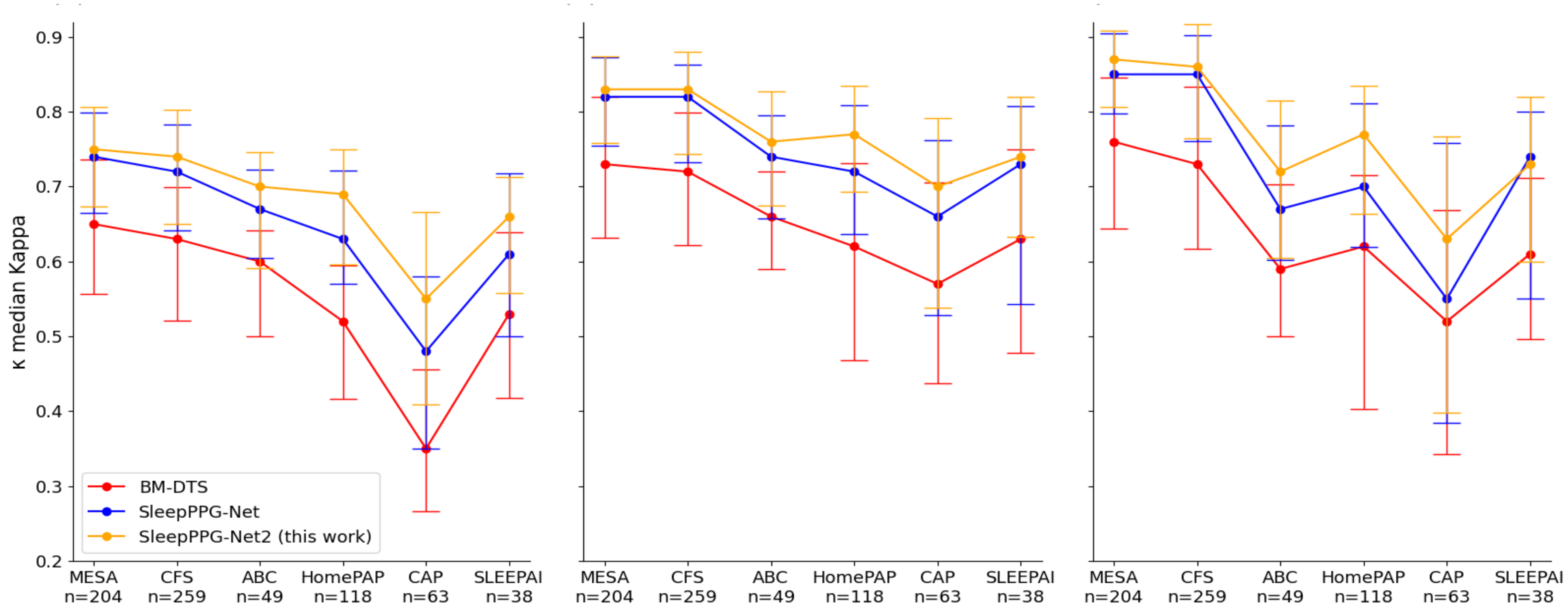
Datasets

We used the raw PPG signal, the metadata and the sleep staging scoring of PSG as reference of 6 datasets that included scoring.

Dataset	Number	Male	fs (Hz)	oximeter	sleep test	Timeframe	Main type of shifts
MESA [1]	2052	46.5%	256	Nonin 8000	Type 1	2000–2002	Ethnicity, high age
CFS [2]	256	40.0%	128	Nonin 8000	Type 2	2001–2006	-
ABC [3]	49	55.3%	256	Nonin 8000	Type 1	2011-2014	Obesity, high AHI
HomePAP [4]	118	54.2%	25-256	-	Type 1	2008-2010	high AHI
CAP [5]	63	63.0%	128	-	Type 1	2001	Comorbidities
SLEEPAI	38	52.2%	75	Nonin 3150	Type 1	2023-2024	-



Results

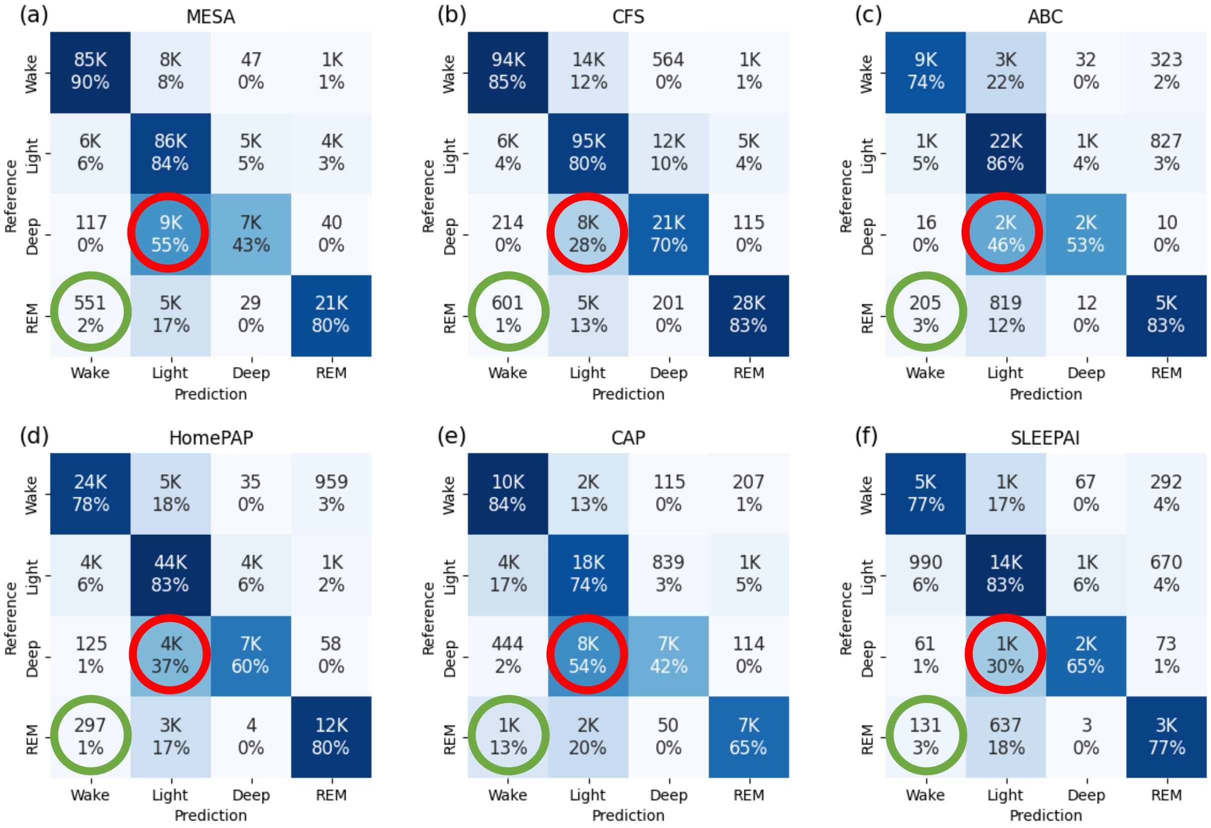


4 classes
(Wake, Light, Deep, REM)

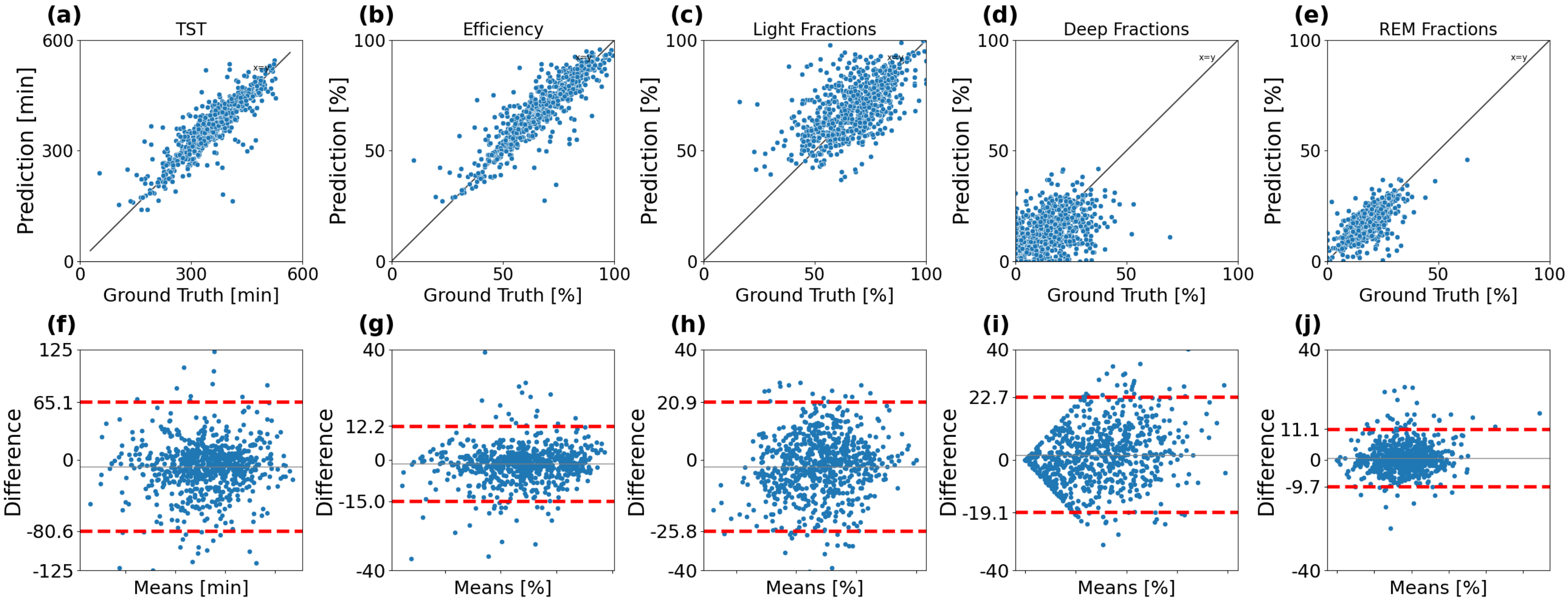
3 classes
(Wake, Non-REM, REM)

2 classes
(Wake, Sleep)

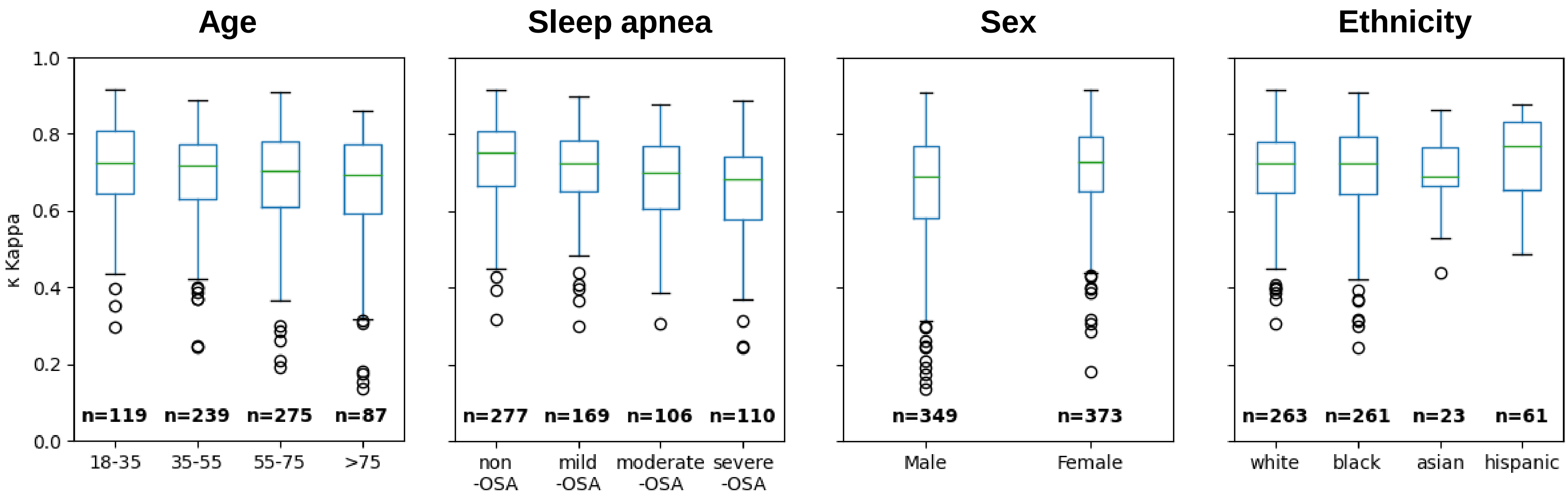
Results


(a) MESA
85K 90%
8K 8%
47 0%
1K 1%
6K 6%
86K 84%
5K 5%
4K 3%
117 0%
9K 55%
7K 43%
40 0%
551 2%
5K 17%
29 0%
21K 80%
(b) CFS
94K 85%
14K 12%
564 0%
1K 1%
6K 4%
95K 80%
12K 10%
5K 4%
214 0%
8K 28%
21K 70%
115 0%
601 1%
5K 13%
201 0%
28K 83%
(c) ABC
9K 74%
3K 22%
32 0%
323 2%
1K 5%
22K 86%
1K 4%
827 3%
16 0%
2K 46%
2K 53%
10 0%
205 3%
819 12%
12 0%
5K 83%
(d) HomePAP
24K 78%
5K 18%
35 0%
959 3%
4K 6%
44K 83%
4K 6%
1K 2%
125 1%
4K 37%
7K 60%
58 0%
297 1%
3K 17%
4 0%
12K 80%
(e) CAP
10K 84%
2K 13%
115 0%
207 1%
4K 17%
18K 74%
839 3%
1K 5%
444 2%
8K 54%
7K 42%
114 0%
1K 13%
2K 20%
50 0%
7K 65%
(f) SLEEPAI
5K 77%
1K 17%
67 0%
292 4%
990 6%
14K 83%
1K 6%
670 4%
61 1%
1K 30%
2K 65%
73 1%
131 3%
637 18%
3 0%
3K 77%
Reference
Wake
Light
Deep
REM
Prediction

Sleep measures



Error Analysis



Conclusion

- Using a multi-source domain training led to consistently higher performance improvements of up to 19% in kappa scores on the target domain.
- The kappa for SleepPPG-Net2 varied between 0.55 and 0.75 for a four-class sleep staging on different datasets.
- The error analysis showed that one source of error is OSA severity – age – gender affects the performances.

- Future directions :
- Overcome the misclassification between Light and Deep sleep.
- Adapt the model to other populations such as children.