Supplemental information

A. SOMNIA AVAILABILITY

The SOMNIA data [22] and HealthBed data [23] used in this study are available from the Sleep Medicine Centre Kempenhaeghe upon reasonable request. The data can be requested by presenting a scientific research question and by fulfilling all the regulations concerning the sharing of the human data. The details of the agreement will depend on the purpose of the data request and the entity that is requesting the data (e.g. research institute or corporate). Each request will be evaluated by the Kempenhaeghe Research Board and, depending on the request, approval from independent medical ethical committee might be required. Access to data from outside the European Union will further depend on the expected duration of the activity; due to the work required from a regulatory point of view, the data is less suitable for activities that are time critical, or require access in short notice. For inquiries regarding availability, please contact Merel van Gilst (M.M.v.Gilst@tue.nl).

B. Additional Quantitative Results

TABLE S.I
HOLD-OUT TEST SET RESULTS FOR ALL MODELS. '#' REFERS TO THE NUMBER OF RECORDINGS IN THE TEST SET WHERE THE SIGNAL WAS AVAILABLE. AVERAGE ACCURACY IN TERMS OF 5-,4-,3-, AND 2-CLASS SLEEP STAGING IS SHOWN HERE.

		Accuracy [%]			
Signal(s)	**#**	W/N1/N2/N3/REM	W/N1-N2/N3/REM	W/NREM/REM	Wake/Sleep
All PSG electrodes	497	85.9	89.8	93.4	96.4
All odd EEG electrodes	500	85.5	89.5	93.3	96.4
F3-M2 (EEG)	500	85.6	89.5	93.1	96.2
C3-M2 (EEG)	500	85.3	89.4	93.2	96.2
O1-M2 (EEG)	500	83.4	87.9	92.6	96.1
Recommended PSG electrodes	497	86.3	90.2	93.6	96.5
All even EEG electrodes	500	85.8	89.8	93.3	96.5
F4-M1 (EEG)	500	85.5	89.4	93.0	96.1
C4-M1 (EEG)	500	85.5	89.6	93.2	96.3
O2-M1 (EEG)	500	83.8	88.2	92.7	96.2
E2-M2 (EOG)	497	85.0	89.2	93.2	96.2
E1-M2 (EOG)	500	84.3	88.8	92.9	95.9
Chin1-ChinZ (EMG)	500	74.9	80.9	88.1	92.7
Chin2-ChinZ (EMG)	500	74.9	80.9	88.0	92.6
Chin1-Chin2 (EMG)	500	74.5	80.6	87.8	92.5
HSAT expanded	434	79.0	84.3	90.8	94.4
HSAT reduced	434	78.3	83.6	90.2	94.1
Nasal cannula	434	76.5	81.9	89.0	93.4
Finger PPG	500	75.1	80.8	88.0	92.4
Thoracic belt	500	76.3	82.7	89.7	93.6
HSAT expanded	434	78.5	84.0	90.6	94.1
HSAT reduced	434	77.7	83.1	89.9	93.5
thermistor	434	72.9	79.2	87.0	91.6
ECG	500	76.9	82.2	89.1	93.2
Thoracic belt	500	76.3	82.7	89.7	93.6
HSAT expanded	66	78.2	83.5	90.5	93.7
HSAT reduced	66	76.4	81.2	88.5	92.1
PAP flow	66	69.5	74.4	83.1	87.8
Finger PPG	500	75.1	80.8	88.0	92.4
Thoracic belt	500	76.3	82.7	89.7	93.6
HSAT reduced	434	77.6	82.9	89.6	93.8
Nasal cannula	434	76.5	81.9	89.0	93.4
IHR from finger PPG	500	71.8	77.6	84.9	90.0
HSAT reduced	65	74.9	80.1	87.0	92.0
IBR from PAP flow	65	69.2	74.8	83.0	89.3
IHR from finger PPG	500	71.8	77.6	84.9	90.0
Left Leg and SCM	33	71.0	77.0	85.1	90.7
Left Leg (EMG)	500	66.9	72.2	81.2	88.5
Left SCM (EMG)	33	66.2	72.4	80.5	89.5
Right Leg and SCM	33	70.3	76.3	84.4	90.2
Right Leg (EMG)	500	66.7	72.0	80.8	88.2
Right SCM (EMG)	33	67.0	73.1	81.7	89.7
Left Leg and FDS	60	67.4	72.8	81.8	88.7
Left Leg (EMG)	500	66.9	72.2	81.2	88.5
Left FDS (EMG)	60	63.7	69.5	78.8	88.2
Right Leg and FDS	60	68.0	73.1	82.0	88.6
Right Leg (EMG)	500	66.7	72.0	80.8	88.2
Right FDS (EMG)	60	63.2	69.1	78.2	87.9
Abdominal belt	500	76.4	83.0	89.9	93.6
Snore microphone	500	72.1	78.0	85.9	91.9
IHR from ECG	500	70.1	75.8	83.8	89.1
IBR from RIP thorax	500	70.0	76.1	83.9	89.7
IBR from RIP abdomen	500	69.9	76.1	84.0	89.7
SpO2	500	68.7	74.7	82.8	89.1
IBR from nasal cannula	434	66.9	73.3	81.5	87.5
IBR from Thermistor	434	65.4	72.0	80.8	87.4
Suprasternal notch	72	63.1	69.5	78.4	86.0
IBR from esophageal pressure	24	59.0	67.9	76.7	83.2
IBR from Suprasternal notch	72	58.5	65.3	74.5	82.2
Esophageal pressure	24	55.5	62.1	72.5	80.7

TABLE S.II
HOLD-OUT TEST SET RESULTS FOR ALL MODELS. '#' REFERS TO THE NUMBER OF RECORDINGS IN THE HOLD-OUT TEST SET WHERE THE SIGNAL WAS AVAILABLE. AVERAGE COHEN'S KAPPA IN TERMS OF 5-,4-,3-, AND 2-CLASS SLEEP STAGING IS SHOWN HERE.

		Cohen's Kappa			
Signal(s)	**#**	W/N1/N2/N3/REM	W/N1-N2/N3/REM	W/NREM/REM	Wake/Sleep
All PSG electrodes	497	0.793	0.826	0.853	0.858
All odd EEG electrodes	500	0.789	0.822	0.850	0.863
F3-M2 (EEG)	500	0.791	0.823	0.846	0.855
C3-M2 (EEG)	500	0.787	0.821	0.848	0.857
O1-M2 (EEG)	500	0.760	0.795	0.835	0.850
Recommended PSG electrodes	497	0.799	0.833	0.857	0.864
All even EEG electrodes	500	0.794	0.827	0.851	0.863
F4-M1 (EEG)	500	0.790	0.822	0.844	0.850
C4-M1 (EEG)	500	0.791	0.824	0.848	0.857
O2-M1 (EEG)	500	0.764	0.800	0.837	0.851
E2-M2 (EOG)	497	0.784	0.820	0.850	0.858
E1-M2 (EOG)	500	0.776	0.815	0.845	0.852
Chin1-ChinZ (EMG)	500	0.630	0.677	0.737	0.716
Chin2-ChinZ (EMG)	500	0.631	0.678	0.735	0.716
Chin1-Chin2 (EMG)	500	0.624	0.672	0.731	0.709
HSAT expanded	434	0.697	0.740	0.801	0.793
HSAT reduced	434	0.686	0.731	0.791	0.783
Nasal cannula	434	0.661	0.705	0.767	0.762
Finger PPG	500	0.640	0.685	0.746	0.735
Thoracic belt	500	0.657	0.708	0.775	0.765
HSAT expanded	434	0.687	0.733	0.797	0.787
HSAT reduced	434	0.674	0.720	0.785	0.774
thermistor	434	0.603	0.650	0.717	0.702
ECG	500	0.669	0.711	0.773	0.772
Thoracic belt	500	0.657	0.708	0.775	0.765
HSAT expanded	66	0.678	0.722	0.797	0.778
HSAT reduced	66	0.652	0.690	0.759	0.735
PAP flow	66	0.562	0.594	0.661	0.634
Finger PPG	500	0.640	0.685	0.746	0.735
Thoracic belt	500	0.657	0.708	0.775	0.765
HSAT reduced	434	0.676	0.719	0.777	0.774
Nasal cannula	434	0.661	0.705	0.767	0.762
IHR from finger PPG	500	0.597	0.637	0.693	0.688
HSAT reduced	65	0.626	0.666	0.719	0.699
IBR from PAP flow	65	0.538	0.580	0.634	0.579
IHR from finger PPG	500	0.597	0.637	0.693	0.688
Left Leg and SCM	33	0.575	0.624	0.699	0.716
Left Leg (EMG)	500	0.526	0.558	0.621	0.632
Left SCM (EMG)	33	0.502	0.546	0.594	0.681
Right Leg and SCM	33	0.565	0.613	0.689	0.700
Right Leg (EMG)	500	0.523	0.556	0.616	0.629
Right SCM (EMG)	33	0.517	0.561	0.627	0.679
Left Leg and FDS	60	0.532	0.566	0.633	0.647
Left Leg (EMG)	500	0.526	0.558	0.621	0.632
Left FDS (EMG)	60	0.482	0.510	0.555	0.615
Right Leg and FDS	60	0.540	0.569	0.639	0.647
Right Leg (EMG)	500	0.523	0.556	0.616	0.629
Right FDS (EMG)	60	0.476	0.503	0.544	0.609
Abdominal belt	500	0.661	0.715	0.779	0.770
Snore microphone	500	0.597	0.639	0.692	0.714
IHR from ECG	500	0.571	0.609	0.670	0.659
IBR from RIP thorax	500	0.564	0.611	0.662	0.626
IBR from RIP abdomen	500	0.563	0.612	0.664	0.625
SpO2	500	0.542	0.589	0.642	0.624
IBR from nasal cannula	434	0.521	0.568	0.616	0.563
IBR from Thermistor	434	0.497	0.545	0.597	0.536
Suprasternal notch	72	0.464	0.499	0.553	0.573
IBR from esophageal pressure	24	0.419	0.473	0.523	0.474
IBR from Suprasternal notch	72	0.394	0.433	0.473	0.431
Esophageal pressure	24	0.373	0.410	0.466	0.483

TABLE S.III
HOLD-OUT TEST SET RESULTS FOR ALL MODELS. '#' REFERS TO THE NUMBER OF RECORDINGS IN THE HOLD-OUT TEST SET WHERE THE SIGNAL WAS AVAILABLE. AVERAGE F1-SCORES PER SLEEP STAGE ARE SHOWN HERE.

		per class F1 scores					
Signal(s)	#	Wake	N1	N2	N3	REM	**MF1**
All PSG electrodes	497	0.885	0.588	0.881	0.834	0.882	0.813
All odd EEG electrodes	500	0.888	0.599	0.876	0.821	0.871	0.810
F3-M2 (EEG)	500	0.882	0.593	0.878	0.833	0.869	0.810
C3-M2 (EEG)	500	0.883	0.596	0.874	0.826	0.872	0.809
O1-M2 (EEG)	500	0.878	0.574	0.857	0.767	0.857	0.785
Recommended PSG electrodes	497	0.891	0.598	0.883	0.845	0.882	0.819
All even EEG electrodes	500	0.886	0.599	0.879	0.833	0.871	0.813
F4-M1 (EEG)	500	0.876	0.591	0.879	0.834	0.868	0.808
C4-M1 (EEG)	500	0.882	0.598	0.877	0.828	0.872	0.810
O2-M1 (EEG)	500	0.878	0.572	0.859	0.787	0.858	0.790
E2-M2 (EOG)	497	0.887	0.577	0.865	0.828	0.876	0.806
E1-M2 (EOG)	500	0.881	0.582	0.855	0.821	0.874	0.801
Chin1-ChinZ (EMG)	500	0.766	0.352	0.778	0.677	0.813	0.676
Chin2-ChinZ (EMG)	500	0.767	0.353	0.778	0.683	0.810	0.677
Chin1-Chin2 (EMG)	500	0.760	0.342	0.775	0.677	0.807	0.672
HSAT expanded	434	0.833	0.434	0.811	0.724	0.845	0.729
HSAT reduced	434	0.824	0.393	0.804	0.718	0.832	0.715
Nasal cannula	434	0.806	0.375	0.787	0.702	0.811	0.695
Finger PPG	500	0.787	0.366	0.775	0.689	0.798	0.681
Thoracic belt	500	0.813	0.433	0.783	0.689	0.828	0.708
HSAT expanded	434	0.832	0.429	0.807	0.700	0.844	0.722
HSAT reduced	434	0.821	0.386	0.800	0.694	0.833	0.707
thermistor	434	0.759	0.331	0.762	0.640	0.774	0.652
ECG	500	0.821	0.399	0.786	0.699	0.829	0.706
Thoracic belt	500	0.813	0.433	0.783	0.689	0.828	0.708
HSAT expanded	66	0.820	0.373	0.814	0.669	0.853	0.704
HSAT reduced	66	0.788	0.320	0.798	0.670	0.834	0.684
PAP flow	66	0.709	0.239	0.733	0.591	0.784	0.610
Finger PPG	500	0.787	0.366	0.775	0.689	0.798	0.681
Thoracic belt	500	0.813	0.433	0.783	0.689	0.828	0.708
HSAT reduced	434	0.818	0.398	0.799	0.712	0.819	0.710
Nasal cannula	434	0.806	0.375	0.787	0.702	0.811	0.695
IHR from finger PPG	500	0.755	0.351	0.744	0.665	0.757	0.651
HSAT reduced	65	0.748	0.267	0.793	0.662	0.798	0.653
IBR from PAP flow	65	0.641	0.155	0.747	0.626	0.742	0.581
IHR from finger PPG	500	0.755	0.351	0.744	0.665	0.757	0.651
Left Leg and SCM	33	0.784	0.239	0.721	0.595	0.727	0.621
Left Leg (EMG)	500	0.709	0.226	0.698	0.598	0.689	0.586
Left SCM (EMG)	33	0.757	0.237	0.668	0.574	0.614	0.575
Right Leg and SCM	33	0.774	0.234	0.713	0.585	0.723	0.616
Right Leg (EMG)	500	0.705	0.220	0.696	0.597	0.680	0.582
Right SCM (EMG)	33	0.754	0.232	0.673	0.543	0.644	0.572
Left Leg and FDS	60	0.727	0.235	0.698	0.614	0.712	0.599
Left Leg (EMG)	500	0.709	0.226	0.698	0.598	0.689	0.586
Left FDS (EMG)	60	0.696	0.257	0.672	0.620	0.578	0.566
Right Leg and FDS	60	0.727	0.242	0.706	0.604	0.714	0.604
Right Leg (EMG)	500	0.705	0.220	0.696	0.597	0.680	0.582
Right FDS (EMG)	60	0.693	0.248	0.663	0.617	0.580	0.560
Abdominal belt	500	0.816	0.442	0.781	0.694	0.831	0.712
Snore microphone	500	0.771	0.362	0.749	0.668	0.729	0.654
IHR from ECG	500	0.732	0.346	0.729	0.637	0.740	0.634
IBR from RIP thorax	500	0.691	0.199	0.737	0.666	0.736	0.606
IBR from RIP abdomen	500	0.688	0.199	0.737	0.664	0.738	0.605
SpO2	500	0.697	0.181	0.722	0.634	0.712	0.592
IBR from nasal cannula	434	0.639	0.179	0.712	0.651	0.709	0.577
IBR from Thermistor	434	0.610	0.157	0.701	0.610	0.698	0.557
Suprasternal notch	72	0.671	0.232	0.662	0.561	0.596	0.546
IBR from esophageal pressure	24	0.568	0.129	0.657	0.579	0.607	0.504
IBR from Suprasternal notch	72	0.541	0.112	0.639	0.566	0.587	0.489
Esophageal pressure	24	0.614	0.117	0.569	0.522	0.529	0.485

TABLE S.IV
HOLD-OUT TEST SET RESULTS FOR ALL MODELS. '#' REFERS TO THE NUMBER OF RECORDINGS IN THE TEST SET WHERE THE SIGNAL WAS AVAILABLE. UNWEIGHTED AVERAGE RECALL IN TERMS OF 5-,4-,3-, AND 2-CLASS SLEEP STAGING IS SHOWN HERE.

		Unweighted average recall			
Signal(s)	**#**	W/N1/N2/N3/REM	W/N1-N2/N3/REM	W/NREM/REM	Wake/Sleep
All PSG electrodes	497	0.820	0.883	0.918	0.922
All odd EEG electrodes	500	0.818	0.874	0.915	0.925
F3-M2 (EEG)	500	0.819	0.878	0.910	0.920
C3-M2 (EEG)	500	0.818	0.874	0.912	0.924
O1-M2 (EEG)	500	0.793	0.848	0.905	0.918
Recommended PSG electrodes	497	0.830	0.892	0.922	0.927
All even EEG electrodes	500	0.821	0.878	0.914	0.922
F4-M1 (EEG)	500	0.817	0.876	0.906	0.915
C4-M1 (EEG)	500	0.822	0.877	0.911	0.921
O2-M1 (EEG)	500	0.798	0.856	0.906	0.918
E2-M2 (EOG)	497	0.828	0.898	0.922	0.936
E1-M2 (EOG)	500	0.825	0.891	0.917	0.930
Chin1-ChinZ (EMG)	500	0.683	0.774	0.835	0.853
Chin2-ChinZ (EMG)	500	0.684	0.776	0.835	0.854
Chin1-Chin2 (EMG)	500	0.677	0.771	0.830	0.845
HSAT expanded	434	0.746	0.831	0.888	0.901
HSAT reduced	434	0.732	0.828	0.882	0.897
Nasal cannula	434	0.714	0.810	0.863	0.882
Finger PPG	500	0.704	0.799	0.855	0.881
Thoracic belt	500	0.727	0.804	0.868	0.892
HSAT expanded	434	0.736	0.821	0.887	0.907
HSAT reduced	434	0.719	0.814	0.883	0.905
thermistor	434	0.658	0.750	0.821	0.865
ECG	500	0.732	0.828	0.890	0.911
Thoracic belt	500	0.727	0.804	0.868	0.892
HSAT expanded	66	0.729	0.823	0.894	0.917
HSAT reduced	66	0.705	0.812	0.880	0.914
PAP flow	66	0.645	0.754	0.832	0.883
Finger PPG	500	0.704	0.799	0.855	0.881
Thoracic belt	500	0.727	0.804	0.868	0.892
HSAT reduced	434	0.727	0.819	0.876	0.893
Nasal cannula	434	0.714	0.810	0.863	0.882
IHR from finger PPG	500	0.674	0.763	0.823	0.864
HSAT reduced	65	0.675	0.788	0.843	0.854
IBR from PAP flow	65	0.603	0.723	0.769	0.781
IHR from finger PPG	500	0.674	0.763	0.823	0.864
Left Leg and SCM	33	0.645	0.760	0.840	0.884
Left Leg (EMG)	500	0.615	0.723	0.801	0.856
Left SCM (EMG)	33	0.593	0.701	0.763	0.862
Right Leg and SCM	33	0.644	0.758	0.831	0.867
Right Leg (EMG)	500	0.611	0.719	0.795	0.855
Right SCM (EMG)	33	0.608	0.718	0.783	0.859
Left Leg and FDS	60	0.625	0.729	0.802	0.846
Left Leg (EMG)	500	0.615	0.723	0.801	0.856
Left FDS (EMG)	60	0.592	0.684	0.719	0.810
Right Leg and FDS	60	0.625	0.729	0.803	0.853
Right Leg (EMG)	500	0.611	0.719	0.795	0.855
Right FDS (EMG)	60	0.586	0.679	0.716	0.812
Abdominal belt	500	0.730	0.806	0.865	0.892
Snore microphone	500	0.675	0.764	0.810	0.872
IHR from ECG	500	0.654	0.741	0.803	0.845
IBR from RIP thorax	500	0.630	0.747	0.793	0.811
IBR from RIP abdomen	500	0.627	0.744	0.792	0.809
SpO2	500	0.614	0.732	0.790	0.830
IBR from nasal cannula	434	0.598	0.711	0.758	0.780
IBR from Thermistor	434	0.576	0.689	0.743	0.759
Suprasternal notch	72	0.573	0.672	0.737	0.824
IBR from esophageal pressure	24	0.531	0.643	0.694	0.728
IBR from Suprasternal notch	72	0.506	0.612	0.669	0.709
Esophageal pressure	24	0.497	0.608	0.663	0.781

C. Additional qualitative results

We here show a qualitative example for each of the signal(s) as shown in tables S.III and S.I. We show the most typical example for each, defined as the recording where it achieved median performance in terms of accuracy.

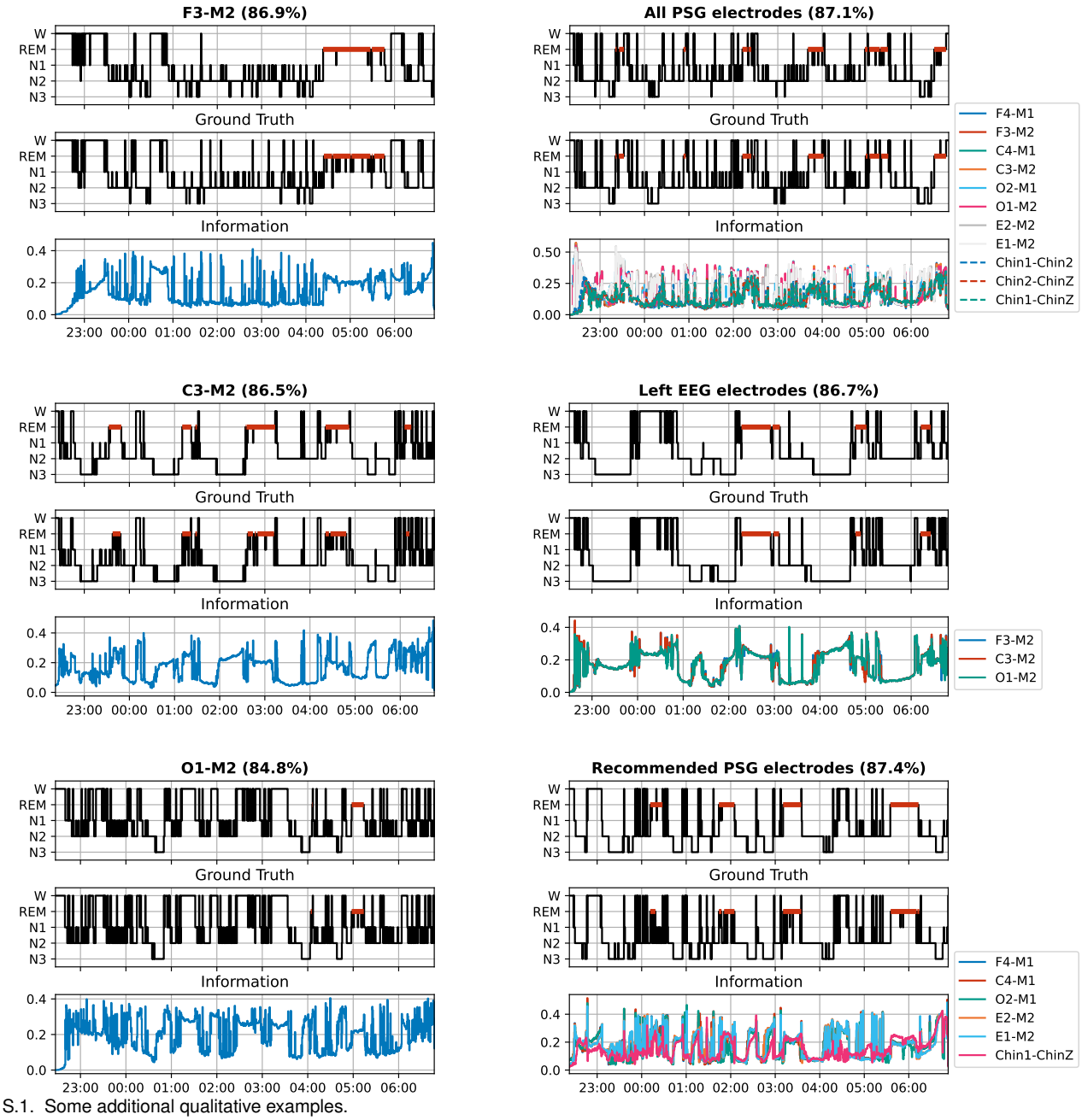


Fig. S.1. Some additional qualitative examples.

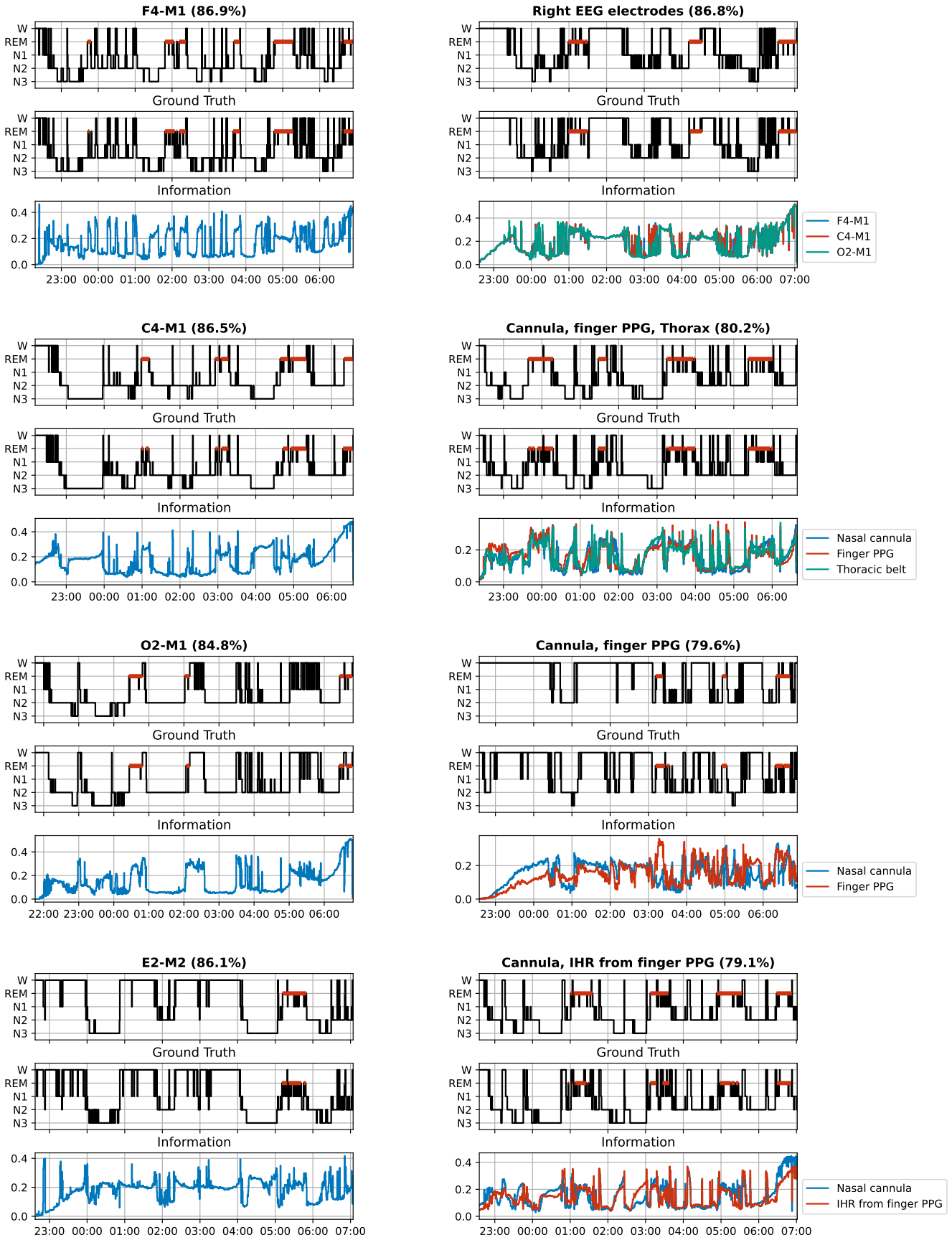


Fig. S.2. Some additional qualitative examples.

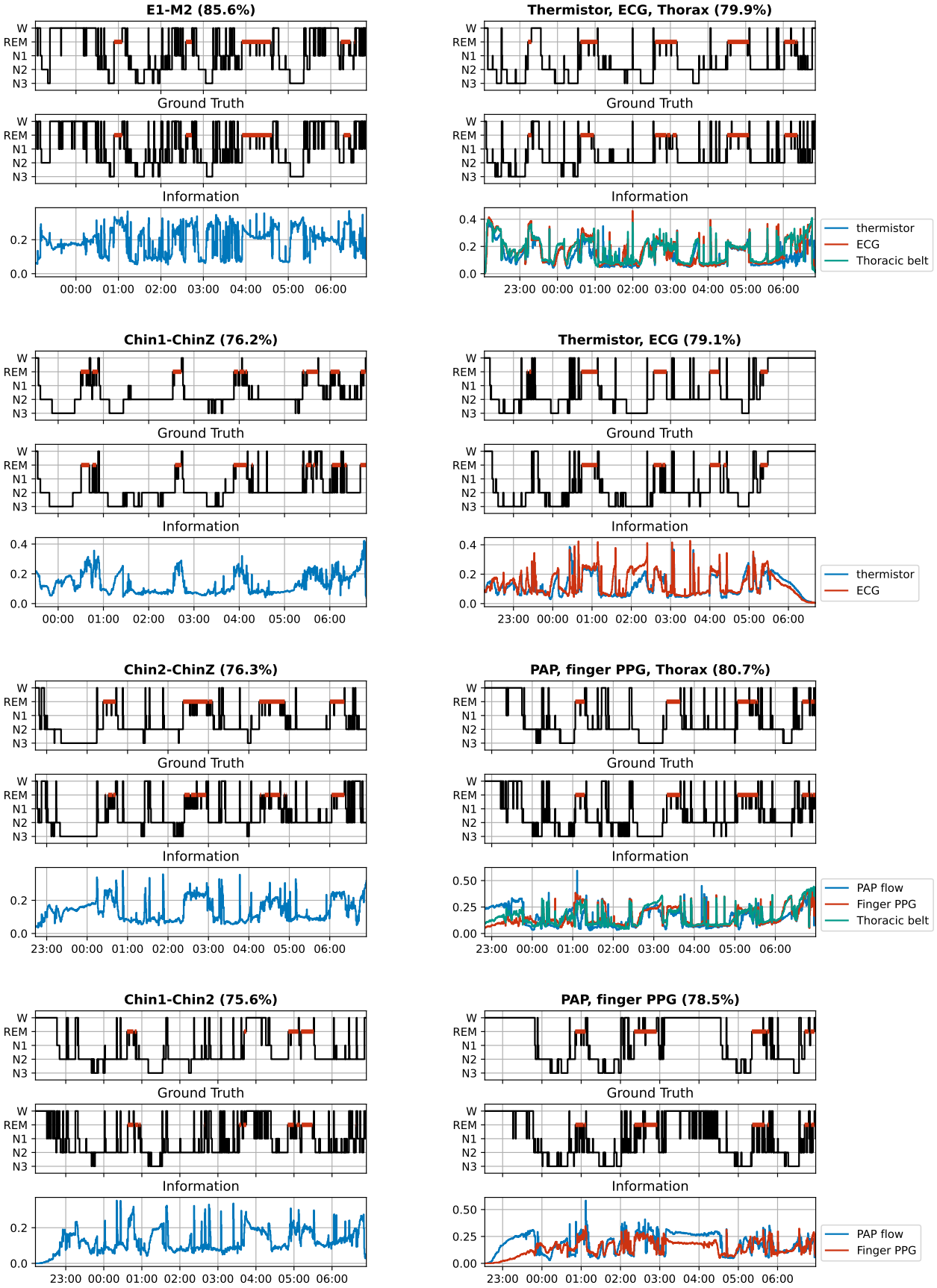


Fig. S.3. Some additional qualitative examples.

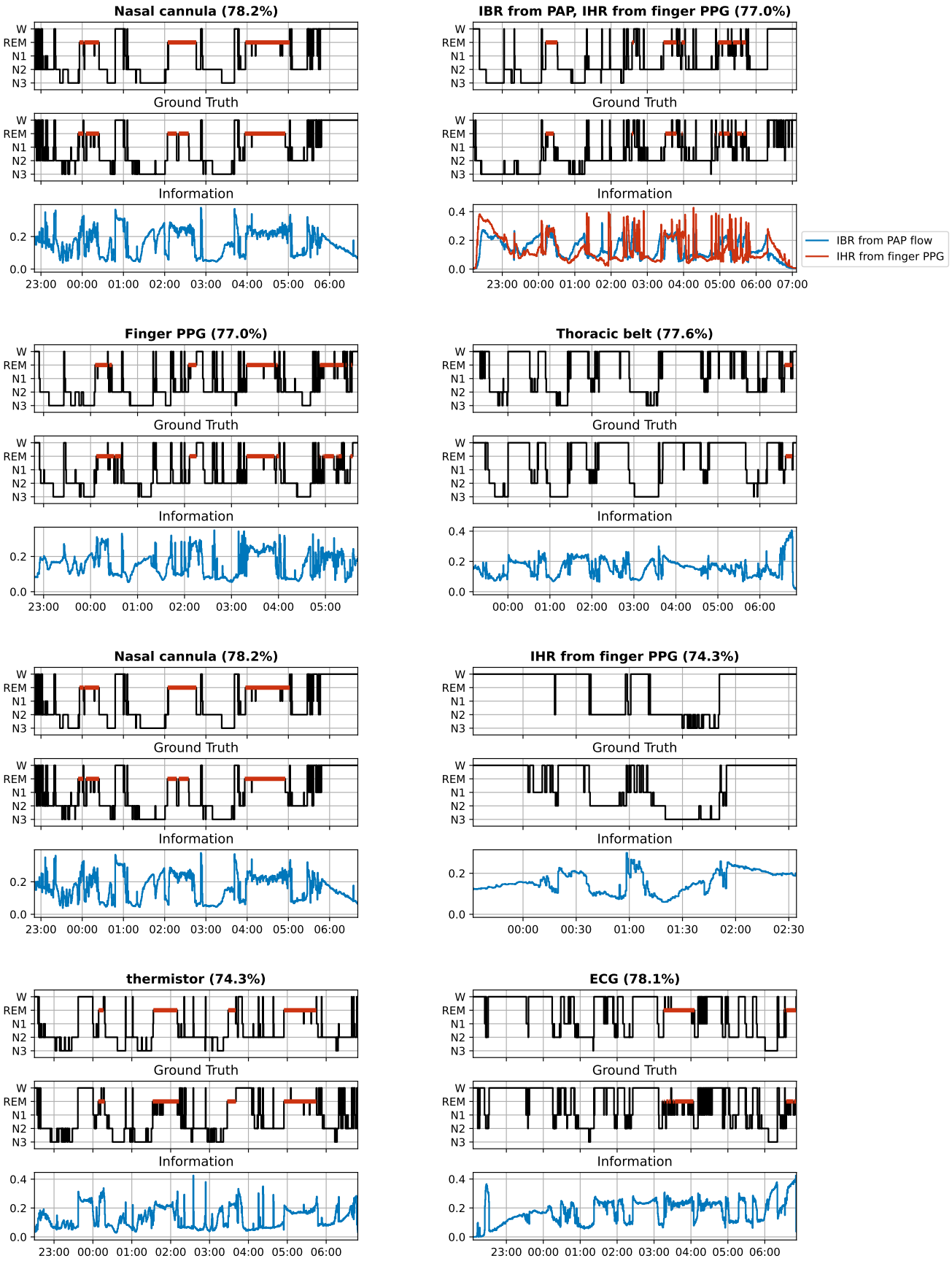


Fig. S.4. Some additional qualitative examples.

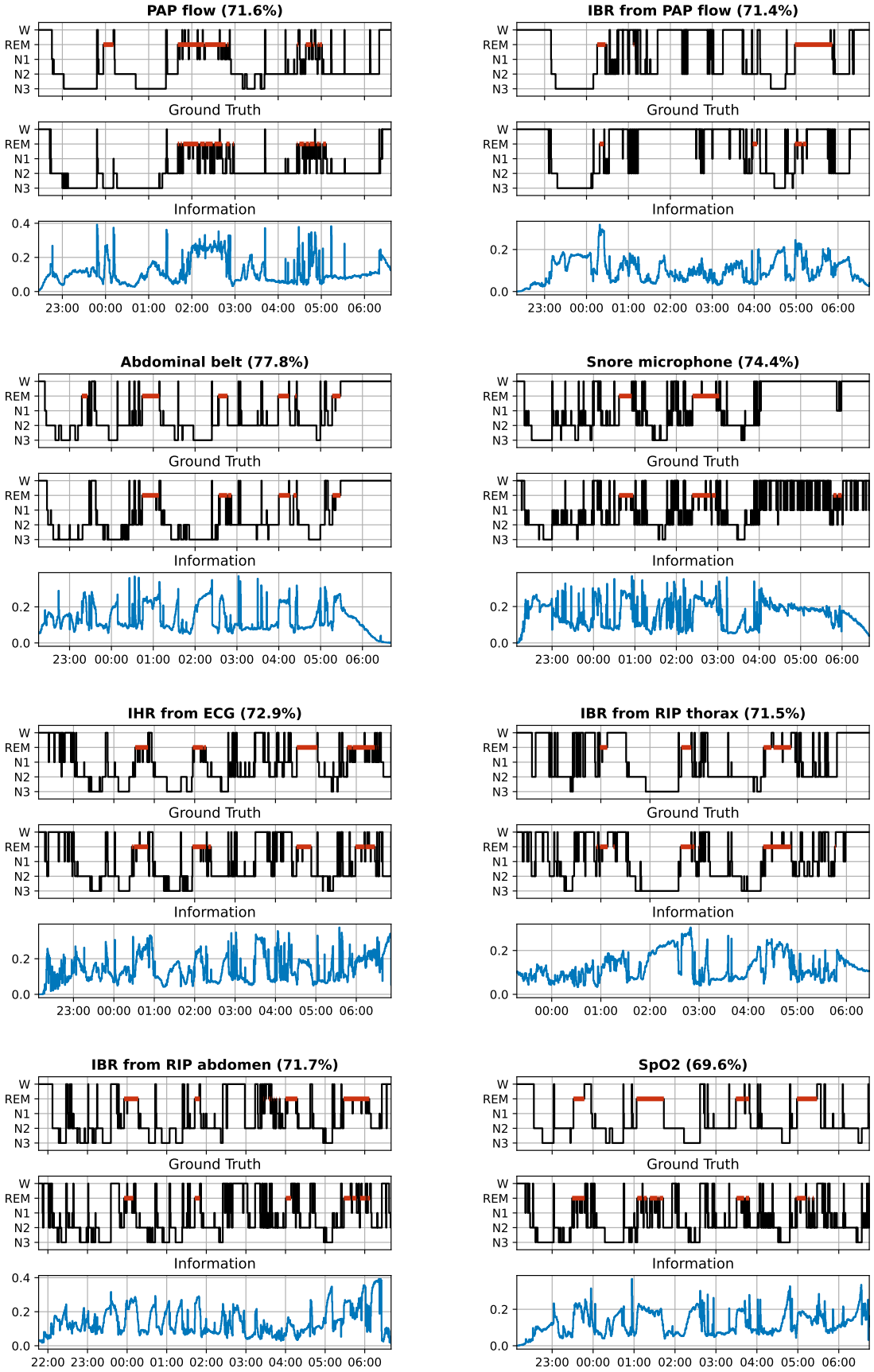


Fig. S.5. Some additional qualitative examples.

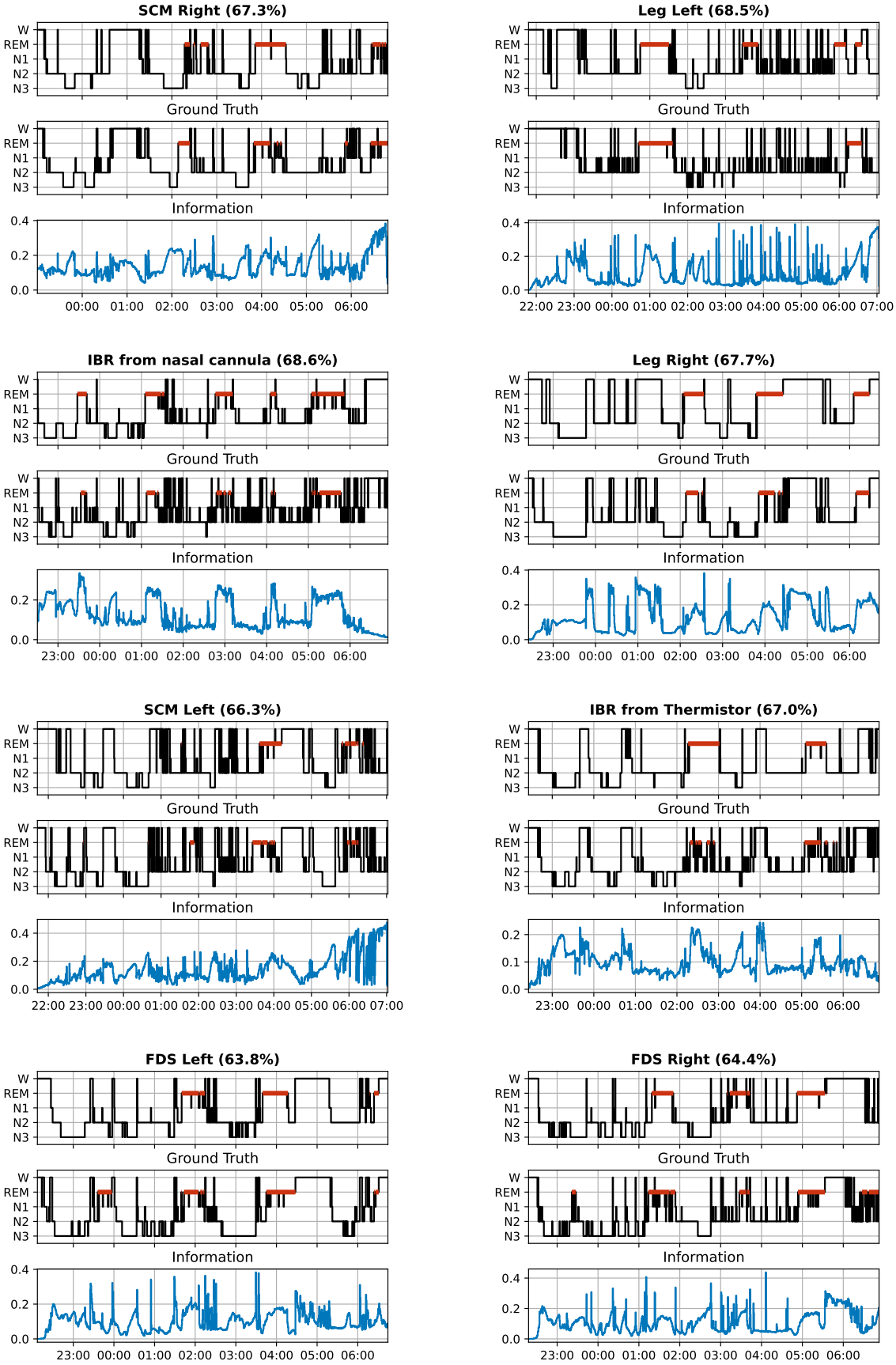


Fig. S.6. Some additional qualitative examples.

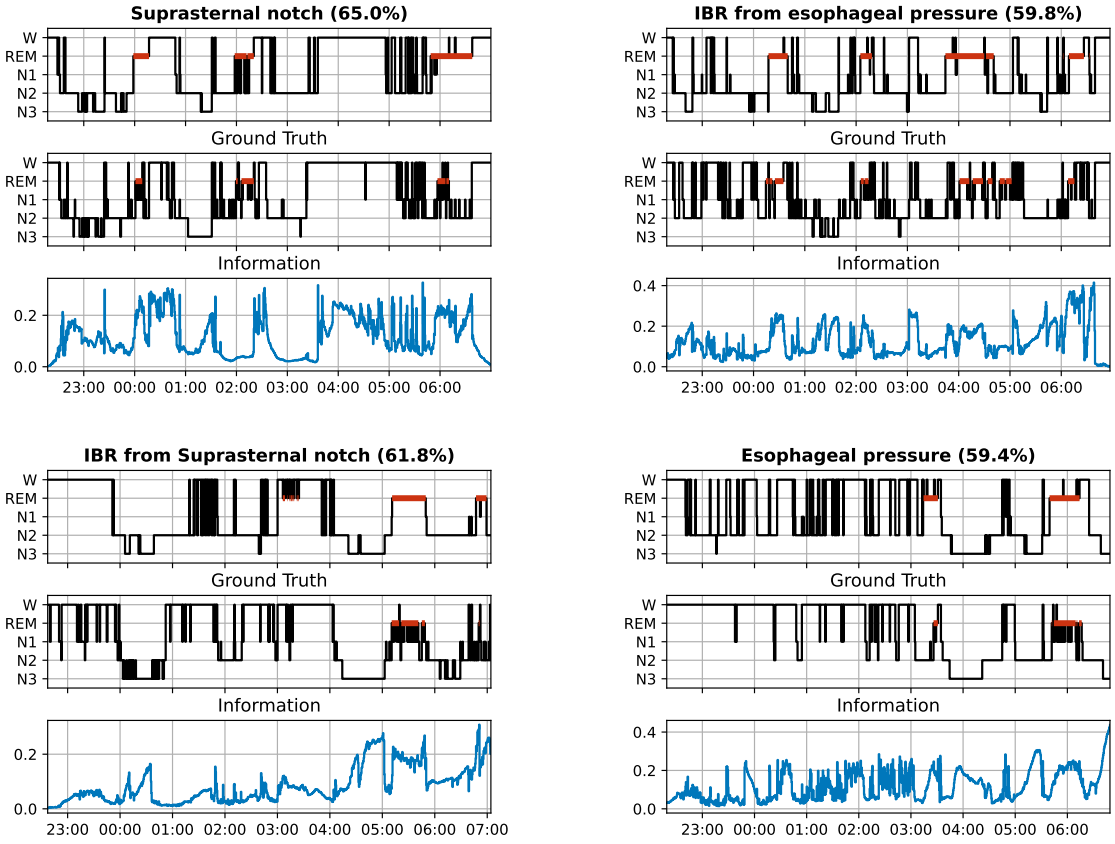


Fig. S.7. Some additional qualitative examples.

D. SAMPLES FROM THE PRIOR

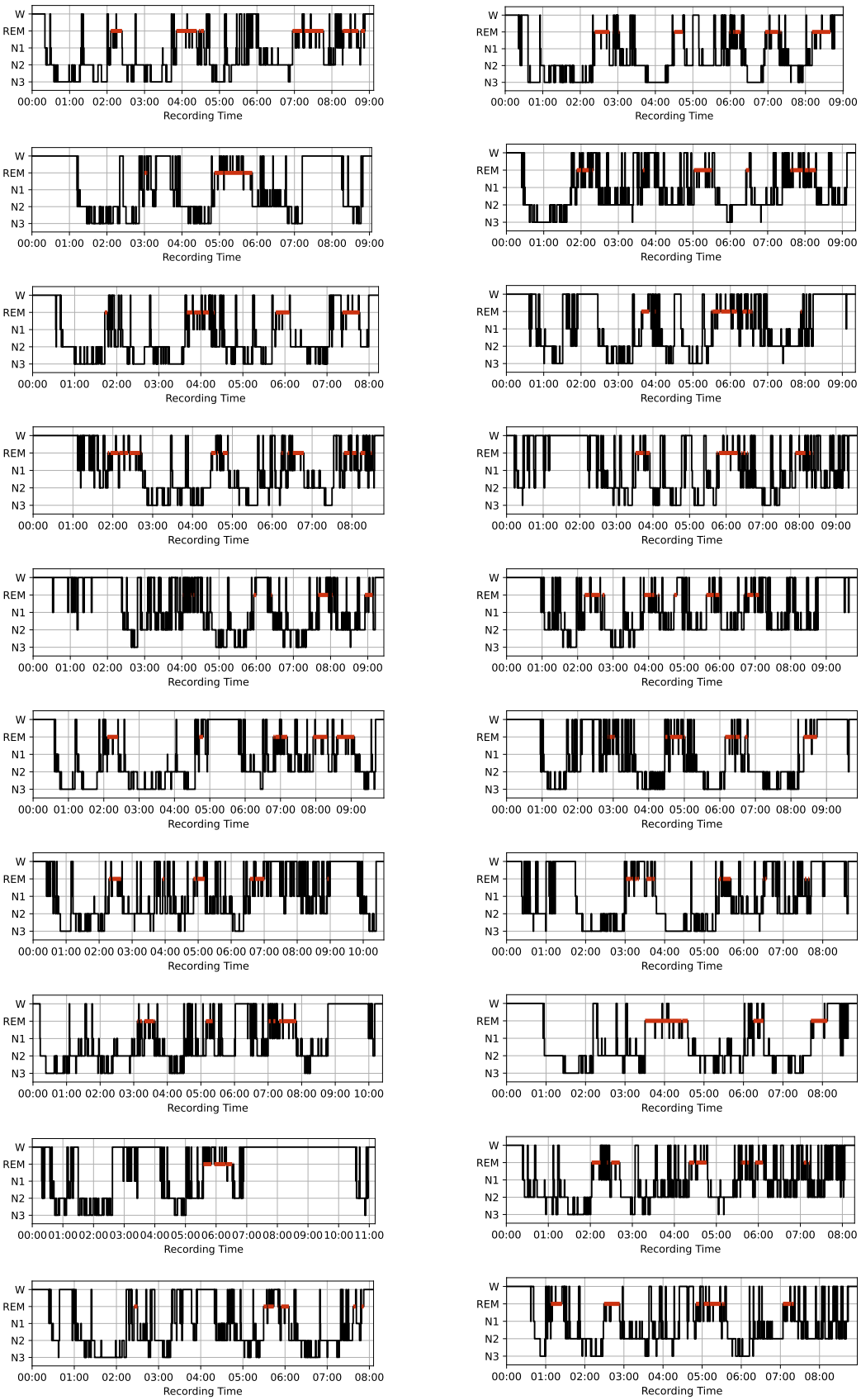
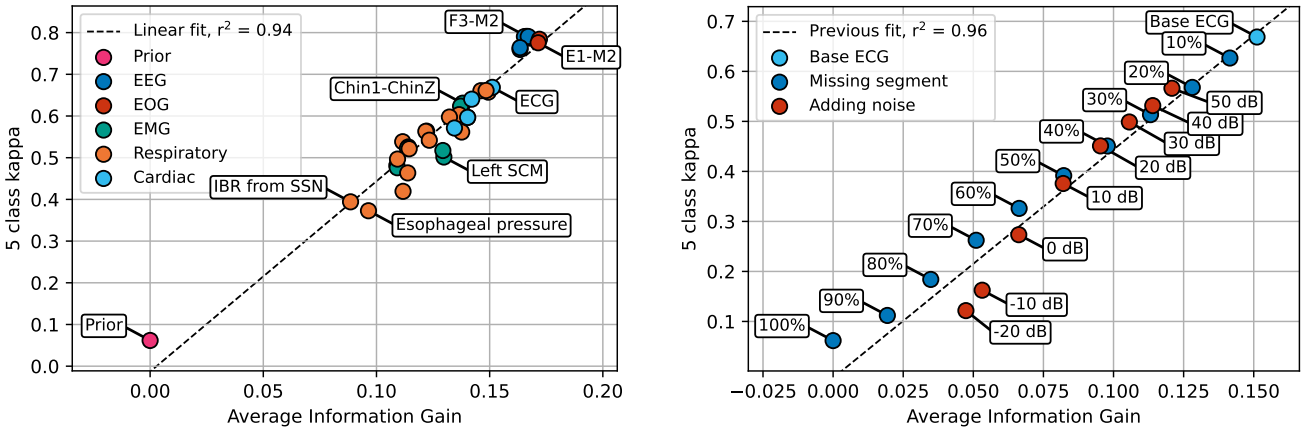


Fig. S.8. We here show samples taken from the global prior score network (without the use of any measurement data).

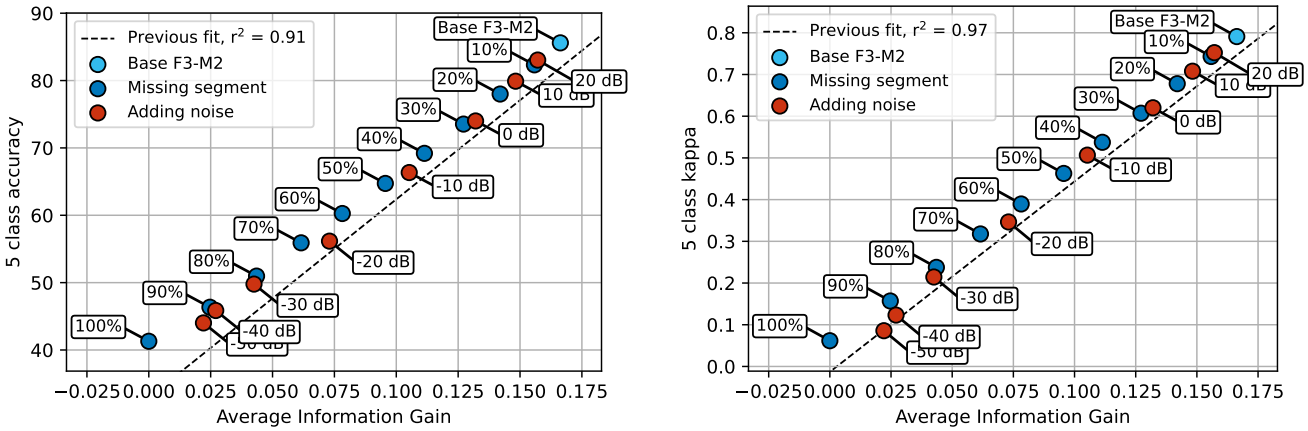
E. Information gain vs performance

We here show the same experiment as shown in Fig.7 from the manuscript, but now for Cohen's kappa instead of accuracy. Additionally, we show the results for this experiment for the F3-M2 sensor.



(A) **(B)**

Fig. S.9. Quantitative results for information gain. **(A)** The average information gain per sensor over all test recordings shows a clear linear correlation with respect to Cohen's kappa. **(B)** Reducing the usefulness of the ECG signal by removing segments or adding noise reduces down-stream accuracy and information gain. The linear relationship as fitted on the data from (A) still provides a good fit here.



(A) **(B)**

Fig. S.10. Quantitative results for information gain for the F3-M2 sensor. **(A)** Results of noise and sensor disconnection for the 5-class accuracy and information gain of the F3-M2 sensor. **(B)** The same experiment as (A), but now expressed for Cohen's kappa.

F. Disconnection of another sensor

We here show the same experiment as shown in Fig.8 from the manuscript, but now for a nasal cannula that is disconnected halfway through the night.

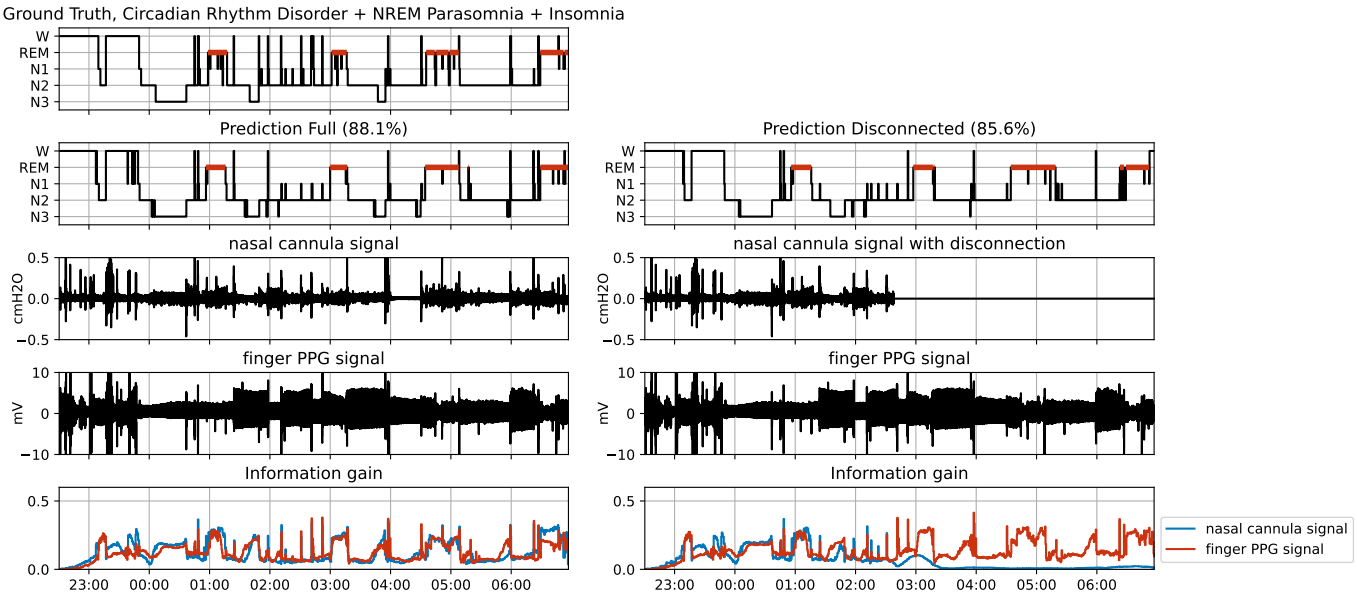


Fig. S.11. Example of disconnecting the nasal cannula signal halfway through the night. Left, output of using the nasal cannula and finger PPG signals over the entire night. Right, artificially created example of what would happen if the user took of the nasal cannula halfway through the night at 3:15. Note that in the original signal there are already two periods of disconnection for the nasal cannula (2:00-2:30 and 4:00-4:45). Additionally, this subject was diagnosed with circadian rhythm disorder, NREM parasomnia, and insomnia.

G. Individual diagnoses and their clusters

TABLE S.V: Specific diagnoses and how they were clustered. Subjects could have multiple sleep disorders within the same cluster, e.g., a pediatric obstructive sleep apnea diagnosis and an adult obstructive sleep apnea diagnsosis.

Diagnosis	**#**	**Cluster**	**#**
Chronic insomnia disorder	217	Insomnia disorders	613
Psychophysiological insomnia	206		
Idiopathic insomnia	11		
Paradoxical insomnia	23		
Inadequate sleep hygiene	55		
Behavioral insomnia of childhood	0		
Insomnia due to (another) mental disorder	69		
Insomnia due to (a) medical condition	86		
Other insomnia disorder	31		
Obstructive sleep apnea, adult	1037	Obstructive Sleep Apnea	1037
Obstructive sleep apnea, pediatric	4		
Central sleep apnea with Cheyne-Stokes breathing	19	Central Sleep Apnea	42
Central sleep apnea due to a medical disorder without Cheyne-Stokes breathing	9		
Central sleep apnea due to a medication or substance	3		
Primary central sleep apnea	11		
Treatment emergent central sleep apnea	6	Treatment emergent central sleep apnea	6
Obesity hypoventilation syndrome	1	Hypoventilation	8
Idiopathic central alveolar hypoventilation	1		
Sleep related hypoventilation due to a medication or substance	1		
Sleep related hypoventilation due to a medical disorder	0		
Sleep related hypoxemia disorder	5		
Narcolepsy type 1	25	Narcolepsy	31
Narcolepsy type 1 due to a medical condition	1		
Narcolepsy type 2	5		
Narcolepsy type 2 due to a medical condition	1		
Idiopathic hypersomnia	8	Other Hypersomnolence Disorders	54
Idiopathic hypersomnia with normal sleep time	14		
Idiopathic hypersomnia with long sleep time	8		
Kleine-Levin syndrome	1		
Hypersomnia due to a medical disorder	11		
Hypersomnia secondary to Parkinson disease	4		
Residual hypersomnia in OSA patients with adequately treated OSA	2		
Hypersomnia associated with a psychiatric disorder	7		
Hypersomnia associated with mood disorder	1		
Hypersomnia associated with a conversion disorder or somatic symptom disorder	1		
Insufficient sleep syndrome	66	Insufficient sleep syndrome	66

TABLE S.V: Specific diagnoses and how they were clustered. Subjects could have multiple sleep disorders within the same cluster, e.g., a pediatric obstructive sleep apnea diagnosis and an adult obstructive sleep apnea diagnsosis.

Diagnosis	**#**	**Cluster**	**#**
Delayed sleep-wake phase disorder	33	Circadian rhythm disorders	46
Advanced sleep-wake phase disorder	2		
Irregular sleep-wake rhythm disorder	1		
Shift work disorder	9		
Circadian sleep-wake disorder NOS	1		
Confusional arousals	74	NREM parasomnias	115
Sleepwalking	48		
Sleep terrors	24		
Sleep related abnormal sexual behaviors	2		
Sleep related eating disorder	1		
REM sleep behavior disorder	122	RBD	122
Recurrent isolated sleep paralysis	11	REM parasomnias other than RBD	55
Nightmare disorder	39		
Sleep related hallucinations	12		
Parasomnia overlap disorder	7	Other parasomnias	45
Exploding head syndrome	1		
Sleep enuresis	2		
Parasomnia due to a medical disorder	4		
Parasomnia, unspecified	31		
Restless legs syndrome	185	RLS/PLMD	268
Periodic limb movement disorder	114		
Sleep related leg cramps	2	Other movement disorders	58
Sleep related bruxism	27		
Sleep related rhythmic movement disorder	5		
Propriospinal myoclonus at sleep onset	2		
Sleep related movement disorder, unspecified	18		
Sleep starts (hypnic jerks)	5		
Other sleep disorder	5	Other	16
Sleep related epilepsy	2		
Sleep related headache	1		
Sleep related laryngospasm	4		
Sleep related gastroesophageal reflux	3		
Sleep disorder due to sedative, hypnotic or anxiolytic	1		
No primary sleep diagnosis	45	No primary sleep diagnosis and/or normal variants	99
Short sleeper	2		
Snoring	31		
Catathrenia	7		
Long sleeper	14		

TABLE S.VI
PERFORMANCE OF THE 'RECOMMENDED PSG' SETUP ON THE DIFFERENT DIAGNOSTIC GROUPS. DIFFERENCE WERE TESTED WITH A MANN-WHITNEY U RANK TEST AT SIGNIFICANCE LEVEL P = 0.05, COMPARING EACH GROUP WITH RESPECT TO ALL OTHER RECORDINGS.

Diagnostic Group	#	5-class Accuracy	5-class Kappa	F1 Wake	F1 N1	F1 N2	F1 N3	F1 REM
Insomnia disorders	166	87.3	0.817	0.911	0.604	0.891	0.842	0.884
Obstructive sleep apnea	278	85.4	† 0.787	0.887	0.596	0.877	0.830	0.878
Central sleep apnea	14	85.3	0.779	0.910	0.570	0.870	0.836	0.870
Hypoventilation	2	77.7	0.688	0.899	0.416	0.784	0.613	0.861
Narcolepsy	10	86.0	0.797	0.824	0.654	0.897	0.849	0.874
Other hypersomnolence disorders	11	88.2	0.828	0.873	0.623	0.920	0.901	0.866
Insufficient sleep syndrome	11	88.4	0.836	0.908	0.668	0.899	0.906	0.899
Circadian rythm disorder	8	86.8	0.808	0.843	0.617	0.898	0.838	0.838
NREM Parasomnias	24	88.2	0.832	0.897	0.655	0.898	0.893	0.901
RBD	30	81.9	† 0.744	0.869	0.510	0.849	0.808	0.836
REM Parasomnias other than RBD	6	85.2	0.794	0.906	0.655	0.871	0.884	0.883
Other Parasomnia	11	84.1	0.702	0.828	0.604	0.877	0.859	0.880
RLS/PLMD	59	87.1	0.799	0.913	0.596	0.888	0.853	0.886
Other movement disorders	18	86.6	0.808	0.872	0.585	0.896	0.828	0.899
Other sleep disorders	2	85.1	0.421	0.908	0.650	0.885	0.922	0.907
No primary sleep diagnosis and/or normal variants	18	87.6	0.821	0.888	0.606	0.902	0.868	0.875
Healthy	27	89.0	0.842	0.865	0.657	0.905	0.902	0.928

† Performance is significantly different from the rest.

TABLE S.VII
PERFORMANCE OF AN HSAT SETUP (CANNULA FLOW, THORACIC BELT, FINGER PPG) ON THE DIFFERENT DIAGNOSTIC GROUPS. DIFFERENCE WERE TESTED WITH A MANN-WHITNEY U RANK TEST AT SIGNIFICANCE LEVEL P = 0.05, COMPARING EACH GROUP WITH RESPECT TO ALL OTHER RECORDINGS.

diagnostic group	#	5-class Accuracy	5-class Kappa	F1 Wake	F1 N1	F1 N2	F1 N3	F1 REM
Insomnia disorders	149	80.2	0.715	0.864	0.424	0.816	0.734	0.850
Obstructive sleep apnea	219	78.0	† 0.683	0.834	0.444	0.801	0.712	0.838
Central sleep apnea	7	71.6	0.575	0.742	0.433	0.770	0.568	0.835
Hypoventilation	2	65.7	0.521	0.815	0.172	0.594	0.544	0.845
Narcolepsy	10	77.3	0.672	0.730	0.465	0.826	0.669	0.847
Other hypersomnolence disorders	10	83.0	0.750	0.821	0.513	0.869	0.832	0.843
Insufficient sleep syndrome	11	80.6	0.728	0.823	0.493	0.825	0.803	0.861
Circadian rythm disorder	8	78.3	0.679	0.764	0.352	0.831	0.720	0.843
NREM Parasomnias	24	79.9	0.711	0.833	0.502	0.825	0.727	0.876
RBD	30	72.6	† 0.609	0.807	0.362	0.755	0.624	0.747
REM Parasomnias other than RBD	6	79.0	0.703	0.849	0.508	0.820	0.787	0.799
Other Parasomnia	11	79.2	0.696	0.804	0.444	0.802	0.693	0.818
RLS/PLMD	56	80.2	0.712	0.857	0.409	0.818	0.706	0.843
Other movement disorders	18	80.3	0.714	0.821	0.391	0.828	0.733	0.864
Other sleep disorders	3	89.3	0.850	0.927	0.548	0.859	0.812	0.872
No primary sleep diagnosis and/or normal variants	17	81.5	0.733	0.849	0.440	0.841	0.801	0.853
Healthy	27	81.3	0.730	0.795	0.472	0.835	0.782	0.902

† Performance is significantly different from the rest.

H. PERFORMANCE PER DIAGNOSIS

To test whether the underlying sleep disorder impacts sleep staging results, we calculated the performance of two signal combinations for each group separately. We used the groups as indicated in Table S.V in Supplement F. The results for the recommended PSG setup are shown in Table S.VI, while the results for an HSAT (cannula flow, thoracic belt, finger PPG) are shown in Table S.VII.

To test whether differences in performance were statistically significant, we used a Mann-Whitney U rank test. We compared the 5-class Kappa of subjects with a certain sleep disorder to those without that sleep disorder. We used a significance level of p = 0.05 and applied a Bonferroni correction for repeated testing. Using this setup, two significant differences were found in both sensor setups. The performance of the OSA and RBD groups was found to be significantly different from the rest. All other diagnostic groups did not show significant differences.

The lower performance for the OSA group is probably due to the increased sleep fragmentation found in this group. The differences in performance for the RBD group are probably due to the large differences in sleep structure of these subjects. By definition, these subjects display REM sleep without atonia (RSWA), which has completely different characteristics from the REM sleep found in all other diagnostic groups. Additionally, this group tends to contain older subjects, whose sleep, especially stage N3, is also more difficult to score [49].

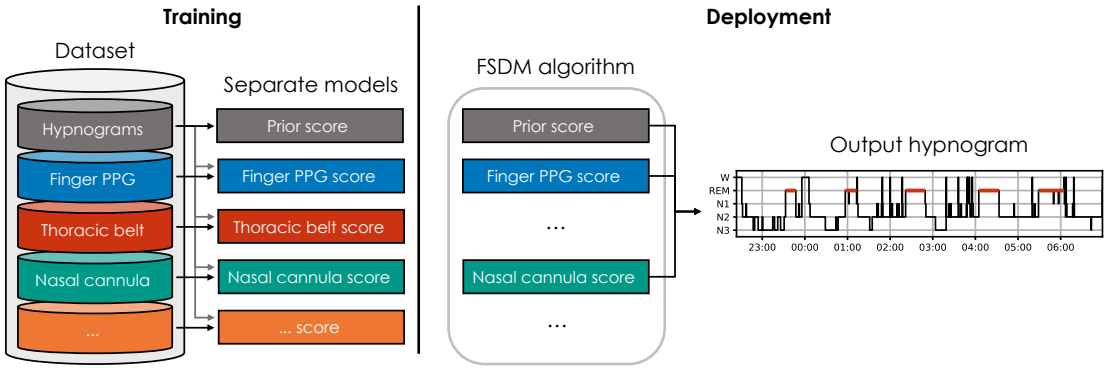


Fig. S.12. Visual overview of the proposed FSDM pipeline. During training, each signal-specific score-network is trained on the subset of data where its sensor was used, including the ground-truth hypnogram. During deployment, any combination of signal modalities can occur. Each signal that was present in the measurement is used with its specific score-network. The proposed FSDM algorithm then fuses the results with a prior score to obtain a posterior using equation (1), from which the hypnogram is sampled.

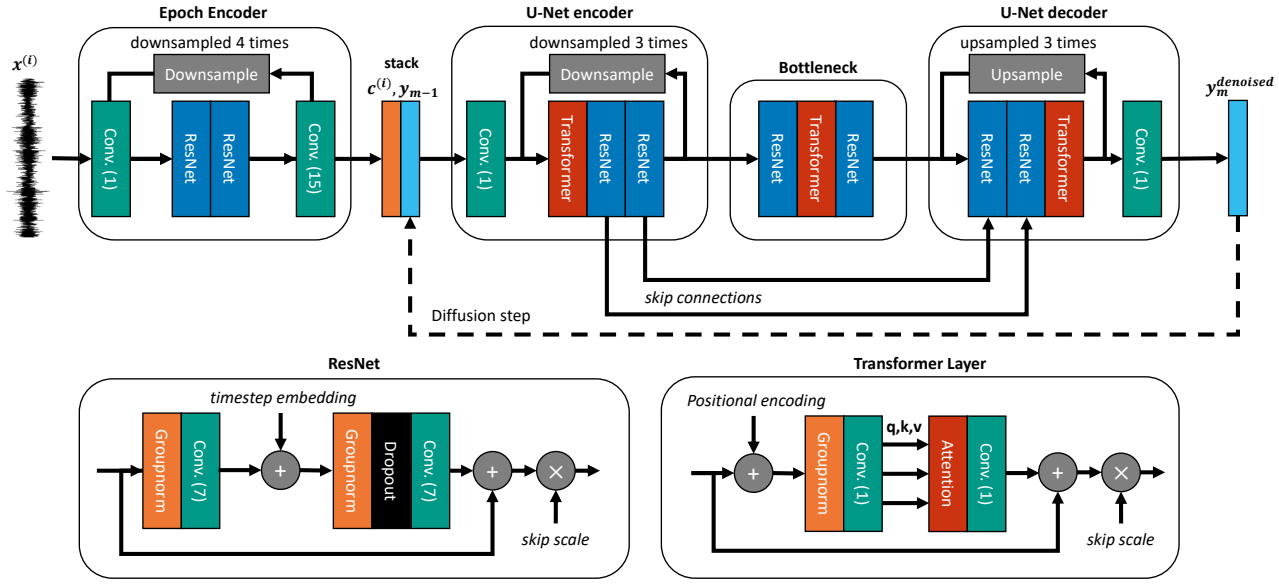


Fig. S.13. Overview of a the neural network used for each denoiser $\boldsymbol{D}_{\boldsymbol{\theta}^{(i)}}\left(\boldsymbol{y_m}, \boldsymbol{x}^{(i)}, \boldsymbol{\sigma}(\boldsymbol{t_m})\right)$. The signal, $\boldsymbol{x}^{(i)}$, is used as the inital input to the network and encoded into a context vector. This is then stacked with the sample at the current timestep $\boldsymbol{y_m}$ and fed though a U-Net structure. At the end, a hypnodensity is given as output through the use of a softmax activation function. The current noise variance, $\boldsymbol{\sigma}(\boldsymbol{t_m})$, is additionally embedded into a timestep embedding, which is added inside the ResNet layers of the U-Net encoder and U-Net decoder. To avoid needing to run the Epoch Encoder $\boldsymbol{M}$ times, the timestep embedding is not added to its ResNet layers.

I. Details regarding the network architecture

We leveraged the DDPM++ model as implemented by Karras *et al.* [11], and modified to work on 1D timeseries in our previous work on EOG-driven sleep staging [24]. See Fig. S.13 for an overview. The neural network architecture used in this work differs from [24] in two aspects. Firstly, there is an additional input between the epoch encoder and U-Net encoder in order to add the $\boldsymbol{y}_{m-1}$ of the previous output (in order to solve the ODE). Secondly, there is an additional embedding in ResNets to add the current diffusion timestep, a common practice in score-based diffusion models [11]. We will now discuss each neural network component.

1. Epoch encoder

Because the signals and the hypnograms had different sampling frequencies (128 Hz vs. 1/30 Hz), we first needed to downsample the input signal before we could use the U-Net structure of our model. To that end, we employed a context encoder, which downsampled the signals from $\mathbb{R}^{1792\cdot30\cdot128\times1}$ to $\mathbb{R}^{1792\times16}$, i.e. a context encoding of length number of epochs with 16 channels.

The context encoder worked as follows. First, a convolution of kernel size 1 expanded the number of channels from 1 to 16. Then, a series of two ResNets was employed to extract meaningful features from the input signal (see the ResNet section for further details). This pattern was repeated 5 times with 4 downsampling operations between the 5 blocks. Each downsampling operation used a kernel of [1,1,1,1] and a stride of 4, to effectively downsample the input by a factor of 4. At the end of the epoch encoder, another convolution of kernel and stride 15 was used, thus compressing the signals to a feature map of size $\mathbb{R}^{1792\times16}$ which was used as input to the U-Net encoder.

2. Stacking

After the epoch encoder, the feature map is concatenated channel wise with the previous estimate of the diffusion step. Following [11], we apply input scaling to the previous estimate of $\boldsymbol{y}_m$ as:

$$\tilde{\boldsymbol{y}}_m = \frac{1}{\sqrt{\sigma_{data}^2 + \sigma_t^2}} \cdot \boldsymbol{y}_m, \tag{23}$$

Where σ_{data} was estimated from data as $\sigma_{data} = 0.3160$ and σ_t is the current variance of the diffusion ODE. The stacking then results in an input of $\mathbb{R}^{1792\times(16+5)} = \mathbb{R}^{1792\times21}$ for the U-Net encoder.

3. Note on prior networks

When we are using the network as a prior network, no input conditioning data is used. Thus, we skip the epoch encoder and the stacking operation, and we only input the previous diffused hypnogram $\tilde{\boldsymbol{y}}_m$ into the rest of the network.

4. U-Net encoder

The U-Net encoder first employed a convolution of kernel size 1 to increase the channel size from 21 to 32. Then, a Transformer layer together with two ResNet blocks was employed (see the Transformer layer section for further details). After each ResNet block, a skip connection was added to the U-Net decoder at the same resolution. This pattern of a transformer with two ResNets was repeated 4 times with 3 downsampling operations in between. Again, a kernel of [1,1,1,1] and a stride of 4 was used in the downsampling operations. The number of channels was left the same throughout the network, at a fixed 32 channels. Note that in the original DDPM++ implementation [10], an attention layer was added after each ResNet in the encoder. However, to bring down the computational complexity of our method and to make the encoder symmetric with the decoder, we employed only a single transformer layer at the start of each resolution level in the U-Net encoder.

5. Bottleneck

In the bottleneck, the feature map was of its smallest size, namely $\mathbb{R}^{28\times32}$. Here, one transformer layer sandwiched between two ResNet blocks was used to learn the highest-level features of the hypnogram.

6. U-Net decoder

The decoder followed a mirrored structure to the encoder. The skip connections from the corresponding resolution levels were concatenated to the inputs of each ResNet block. These connections allowed the feature maps to skip the downward path of the 'U' and enabled the model to learn both high-and-low level features of the hypnogram. The upsampling operation of the decoder was implemented using a transposed convolution with the same filter of [1,1,1,1].

As a final step toward creating the $\boldsymbol{y}_m^{denoised}$, the U-Net decoder employed a convolution of kernel size 1 to map the input to 5 channels, where each channel corresponded to one of the five sleeps stages. A softmax activation function was then used to map each channel to a class probability. This creates a 'hypnodensity', a soft version of the hypnogram where each epoch is partially associated with each sleep stage according to some probability [50].

7. ResNet

The ResNet, or Residual Network, was repeated throughout the architecture. It consists of two group normalization layers and two convolutions in an alternating pattern. Group normalization, as described by [51], applies a learned normalization across groups of channels, enabling faster training. In our case, each group consisted of 4 channels. The 1D convolutions of the ResNet each used a kernel of size 7 and zero-padding set to 'same'. Each convolution was followed by SiLU (Sigmoid Linear Unit) activation [52]. Additionally, a spatial dropout layer was added before the second convolution, which drops out entire channels during training with a probability of 10%. Spatial dropout is a better regularizer for convolutional neural networks, since neighbouring samples are often highly correlated [53]. Finally, a residual connection was added to help combat vanishing gradient problems. To limit the magnitude of the signals, scaling with a factor of $skip\ scale = \sqrt{0.5}$ was applied.

In the case that the ResNet was part of the epoch encoder, the additive timestep embedding was equal to zero, and we effectively do not add it. Otherwise. an additive timestep embedding is generated from the current noise level of the diffusion process to tell it about what kind of noise level to expect in $\boldsymbol{y}_{m-1}$, which has been found to be helpful in the score-based diffusion literature. This additive timestep embedding is explained in the sequel.

8. Timestep embedding

Following [11], the current noise level of the diffusion process, σ_t, is given as an additional input to the network in a scaled and embedded form. To that end, it is first scaled as follows:

$$\tilde{\sigma}_t = 0.25 \log(\sigma_t). \tag{24}$$

After this scaling, a sine-cosine embedding scheme was used that embedded the noise level as follows:

$$\boldsymbol{c} = [0, 1, \ldots, C/2 - 1]^T, \tag{25}$$

$$\boldsymbol{f} = 1000\hat{}(-\boldsymbol{c}/(C/2 - 1)), \tag{26}$$

$$\boldsymbol{z} = [\cos(\tilde{\sigma}_t \cdot \boldsymbol{f}),\ \sin(\tilde{\sigma}_t \cdot \boldsymbol{f})]^T, \tag{27}$$

where C is the number of channels in the ResNet, which is equal to 32 throughout the U-Net. Subsequently, a multilayer perceptron (MLP) was applied of 2 layers, with 8 hidden nodes each and SiLU activation. Then, for each ResNet separately, a local linear layer was applied to increase the noise level embedding to 32 again. As a final step it is broadcasted into the input length of the feature map at the ResNet and added to it element-wise.

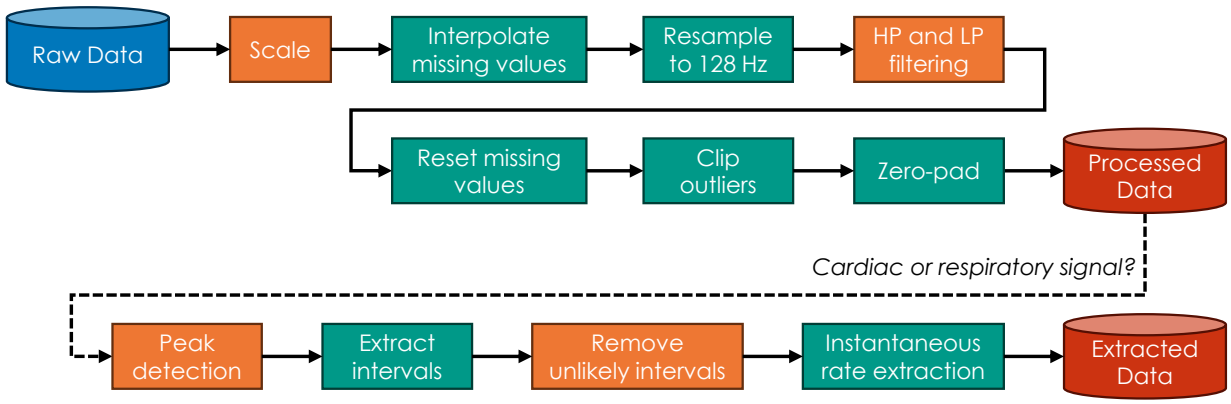


Fig. S.14. The preprocessing pipeline is the same for all the signals. The orange blocks are set by signal-type specific parameters, e.g. the cutoff frequencies of the filters can be different for different signal types, see Table III. If a signal is a cardiac or respiratory signal we also extract the instantaneous heart-rate or breathing-rate.

9. Transformer

The original transformer architecture is a sequence-to-sequence model composed of both an encoder and a decoder [54]. Where each element consists of a scaled dot-product attention layer and an element-wise feed-forward network. Additionally, positional encoding is added at the start of the encoding and decoding stacks. We adapt the transformer architecture to be suited for our network. Firstly, we did not use the decoder, since it is used to generate new sequence in an auto-regressive manner. Secondly, since we embedded the layers within a larger convolutional neural network, there was no need for separate element-wise feed-forward networks. lastly, because the attention layers operated at different time scales, we added positional encoding to each of them.

The positional encoding was also implemented using sine-cosine embedding. The encoding scheme used is similar to the timestep embedding, with some differences. Namely, we here create a full matrix embedding instead of only a vector embedding, no MLP is applied, and the sine and cosine terms are interleaved.

In the transformer layer positional encoding scheme, a positional encoding matrix is added element-wise to the input sequence of the transformer. To that end, the input sequence $\mathbf{S}$ and positional encoding matrix $\mathbf{P}$ should be of the same size: $\mathbf{S}, \mathbf{P} \in \mathbb{R}^{L \times C}$, where L is the length of the input sequence and C is the number of channels (32 in our case). The positional encoding matrix for the transformer layers is given by:

$$\begin{aligned} \mathbf{P}_{(l,2c)} &= \sin\left(l \cdot 1000^{-2c/C}\right) \\ \mathbf{P}_{(l,2c+1)} &= \cos\left(l \cdot 1000^{-2c/C}\right), \end{aligned} \tag{28}$$

with $l \in [0, 1, \ldots, L-1]$ and $c \in [0, 1, \ldots, C/2-1]$. This type of encoding enables the transformer to exploit information about both the absolute and relative positions of samples along the night.

Each of the transformer layers used scaled dot-product self-attention. While the attention mechanism can be implemented using multiple attention-heads for added complexity, we here only made use of a single head. In scaled dot-product self-attention, three linear projections are applied to transform the sequence to a query, key, and value matrix:

$$\mathbf{Q} = \mathbf{S}\mathbf{W}_Q, \ \mathbf{K} = \mathbf{S}\mathbf{W}_K, \ \mathbf{V} = \mathbf{S}\mathbf{W}_V, \tag{29}$$

where $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V \in \mathbb{R}^{C \times C}$ are learned linear projection weights and $\mathbf{Q}, \mathbf{K}, \mathbf{V} \in \mathbb{R}^{L \times C}$ are the query, key, and value matrices, respectively. These linear projections can be implemented efficiently by a single convolutional layer of kernel size 1 and output channel size of $3C$, as its output can be split along the channel dimension into the three separate components.

Following a database analogy, the queries are going to look for matching keys and propagate the associated values to the output, where each individual query, key, and value are found along the rows of their respective matrices. This process is defined by the scaled dot-product self-attention mechanism:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{C}}\right)\mathbf{V}, \tag{30}$$

where $\mathbf{K}^T$ denotes the transpose of the key matrix. Moreover, $\mathbf{Q}\mathbf{K}^T \in \mathbb{R}^{L \times L}$ denotes the attention map. To ensure that the magnitudes in the attention map do not grow too large, it is scaled down by a factor of $1/\sqrt{C}$. Additionally, a softmax activation is applied along the rows of the attention map in order to ensure that the attention sums to 1.

After the scaled dot-product attention layer, another linear projection using a 1D convolution was applied. Similar to the ResNet, a residual connection was applied with a scaling of $skip\ scale = \sqrt{0.5}$.

J. Preprocessing

We applied a common preprocessing pipeline to all signals as shown in Fig. S.14 and Table S.VIII. We will briefly describe each preprocessing operation. First, each of the signals is scaled by a constant value in order to bring its approximate magnitude around 1. The scaling factor is chosen specific to each signal type and is shown in Table III. We for example scale all the EEG channels by a factor 10^4, making it so that an amplitude of 100μV corresponds to the value 1 and the slow-wave amplitude threshold of 75μV corresponds to a value of 0.75 [1]. This scaling enables faster training of the neural networks.

TABLE S.VIII
OVERVIEW OF THE SIGNALS EXTRACTED FROM THE DATASETS. WE CLUSTERED THE SIGNALS INTO GROUPS SUCH AS EEG AND RIP BELTS. 'HP' AND 'LP' DENOTE THE HIGH-PASS AND LOW-PASS FILTER RESPECTIVELY WHERE WE SHOW THE CUT-OFF FREQUENCY IN HZ. 'FDS' AND 'SCM' REFER TO THE FLEXOR DIGITORUM SUPERFICIALIS AND STERNOCLEIDOMASTOID MUSCLES, RESPECTIVELY.

Signal group	Signals	Associated physiology	Unit	Scale	HP	LP	Total †	Train †	Val †	Test †
EEG	F3-M2, F4-M1, C3-M3, C4-M1, O1-M2, O2-M1	Cortical activity	V	10^4	0.3	49	11681 (1947)	8081 (1347)	600 (100)	3000 (500)
EOG	E1-M2, E2-M2	Eye activity	V	10^4	0.3	49	3886 (1947)	2689 (1347)	200 (100)	997 (500)
EMG chin	Chin1-ChinZ, Chin2-ChinZ, Chin1-Chin2	Muscle activity	V	10^4	10	49	5838 (1946)	4038 (1346)	300 (100)	1500 (500)
ECG	ECG	Heartrate	V	10^3	0.3	49	1947 (1947)	1347 (1347)	100 (100)	500 (500)
RIP belts	Abdomen, Thorax	Respiration	V	10^{-2}	0.1	15	3892 (1946)	2692 (1346)	200 (100)	1000 (500)
Thermistor	Thermistor	Respiration	V	10^4	0.1	15	1706 (1706)	1184 (1184)	88 (88)	434 (434)
Nasal cannula	Nasal cannula	Respiration	cmH2O	1	0.03	49	1706 (1706)	1184 (1184)	88 (88)	434 (434)
PAP flow	PAP flow	Respiration	cmH2O	10	0.03	49	241 (241)	163 (163)	12 (12)	66 (66)
Suprasternal notch	Suprasternal notch	Respiration	V	10	0.03	49	289 (289)	199 (199)	18 (18)	72 (72)
Esophageal pressure	Esophageal pressure	Respiration	mmHg	10^{-1}	0.03	49	97 (97)	65 (65)	8 (8)	24 (24)
Snore microphone	snore microphone	Respiration	V	10^3	10	49	1947 (1947)	1347 (1347)	100 (100)	500 (500)
PPG	PPG	Heartrate	V	10^{-2}	0.3	49	1976 (1943)	1364 (1343)	101 (100)	511 (500)
SpO2	SpO2	Oxygen saturation	%	10^{-2}	-	-	1944 (1944)	1344 (1344)	100 (100)	500 (500)
EMG FDS	FDS L, FDS R	Muscle activity	V	10^4	10	49	508 (254)	368 (184)	20 (10)	120 (60)
EMG legs	Leg L, Leg R	Muscle activity	V	10^4	10	49	3894 (1947)	2694 (1347)	200 (100)	1000 (500)
EMG SCM	SCM L, SCM R	Muscle activity	V	10^4	10	49	296 (148)	206 (103)	24 (12)	66 (33)
Instantaneous heart rate	ECG, PPG	Heartrate	Bpm	1/60	-	-	3923 (1947)	2711 (1347)	201 (100)	1011 (500)
Instantaneous breath rate	RIP Belts, Thermistor, Nasal cannula, PAP flow, Suprasternal notch, Esophageal pressure	Respiration	Brpm	1/60	-	-	8163 (1947)	5642 (1347)	426 (100)	2095 (500)

† The listed numbers indicate the total number of individual signals, after which we list the unique number of recordings between brackets. For example, 6 individual EEG signals were extracted from each recording.

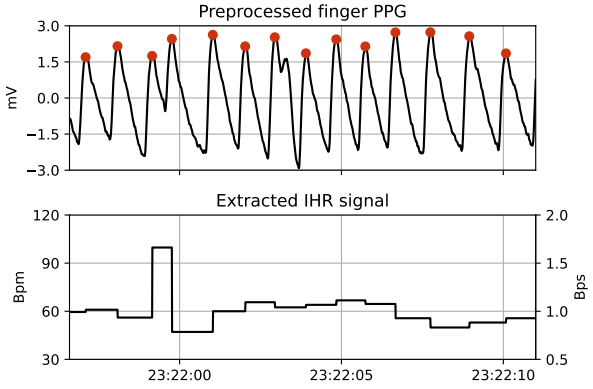


Fig. S.15. Example of IHR extraction from a finger PPG signal. The red dots denote the peaks found by the automatic multi-scale peak detection algorithm.

Second, we identify missing values in the signals as those locations where they are exactly equal to 0. We linearly interpolate these values by using neighbouring sample values. This interpolation is performed in order to reduce filtering artifacts that can occur by the large jumps in magnitude that these missing values cause. After all the filtering operations, the samples with the missing values are reset to exact zeroes, this enables the neural networks to understand which values are missing in the signal.

Third, we resample all signals to 128 Hz using a polyphase filter, except for the SpO2 signal which was up-sampled from 32 Hz to 128 Hz using a sample and hold scheme. After resampling, we apply a low-pass and a high-pass filter, both implemented using fifth order butterworth filter. The high-pass and low-pass filter settings are signal type specific and reflect the recommendations of the AASM manual [1] with some minor adjustments to the low-pass filter settings.

After filtering we reset the samples values of the missing value indices back to zero. We then clip the signals between -5 and 5. For example, the EEG signals are clipped between -500μV and +500μV. Lastly, we zero-pad all the signals to a common length of 1792 sleep epochs, or 1792·30·128 = 6,881,280 samples. This zero-padding was solely done for implementation purposes, as it allows to stack the signals of

Algorithm 1: Training a single score network

Input:

signal index $i \in [0, 1, \dots, N]$, Densoising function $D_{\theta^{(i)}}$, dataset sampler $p^{(i)}_{data}$, noise sampling scheme p_σ, optimizer $opt()$, probabilities p_{augment} and p_{zero}

while *not converged* **do**

  `// expectation through Monte-Carlo`

  sample $\boldsymbol{x}^{(i)}, \boldsymbol{y} \sim p^{(i)}_{data}$

  sample *augment* $\sim p_{\text{augment}}$, *zero* $\sim p_{\text{zero}}$

  sample $\sigma \sim p_\sigma$

  sample $\boldsymbol{n} \sim \mathcal{N}(0, \sigma^2 I)$

  `// augment the data`

  if *augment* **then**

    $k \sim u(1, |\boldsymbol{x}^{(i)}|)$, $l \sim u(1, |\boldsymbol{x}^{(i)}|)$

    $\boldsymbol{x}^{(i)}[k : l] = \mathbf{0}$

  if *zero* **then**

    $\boldsymbol{x}^{(i)} = \mathbf{0}$

  `// Optimizer step`

  $\boldsymbol{y}_{noisy} = \boldsymbol{y} + \boldsymbol{n}$

  $\boldsymbol{y}_{denoised} = D_{\theta^{(i)}}\left(\boldsymbol{y}_{noisy}, \boldsymbol{x}^{(i)}, \sigma\right)$

  $J_i = -\sum \boldsymbol{y} \log \boldsymbol{y}_{denoised}$

  $opt(J_i, D_{\theta^{(i)}})$

Output: $D_{\theta^{(i)}}$

different nights, and the zero-padded segment was not used to calculate any of the overnight statistics or performance metrics such as accuracy and Cohen's kappa.

In the case that the extracted signal was a cardiac or respiratory signal, such as ECG, finger PPG, or RIP Belt, we also extracted the instantaneous heart-rate (IHR) or the instantaneous breath-rate (IBR). This extraction was performed by extracting the peaks of the cardiac pulses and the breaths using the automatic multi-scale peak detection algorithm [48]. We used the following settings: window length = 600 seconds, window overlap = 120 seconds, and maximum scale = 5 seconds for cardiac signals, while maximum scale = 60 seconds for respiratory signals. After peak detection, we convert the peaks to the inter-beat and inter-breath interval length using a sample and hold scheme. To reduce the impact of artifacts, biologically implausible intervals are removed from the sequence. We remove all cardiac inter-beat intervals outside the range of [0.3s - 2s], and we remove all respiratory inter-breath intervals outside the range of [1s - 30s]. As a final step, the intervals are converted into the IHR or IBR. While typically expressed in beats per minute (Bpm) or breaths per minute (Brpm), we scale the signals by 1/60 to get the beats/breaths per second, resulting in a better magnitude for use by the neural networks. Fig. S.15 shows an example of how we extract the peaks of a finger PPG signal, which we then convert to inter-beat intervals, to subsequently convert to the IHR.

K. Pseudo-code for training and sampling

The pseudo-code used to train a single score-based diffusion model is shown in algorithm 1. When training the prior network

Algorithm 2: Sampling from an FSDM

Input: Measured signals $X = [\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_N]$, Denoising functions $D = [D_{\theta^{(0)}}, \dots, D_{\theta^{(N)}}]$, noise schedule $\sigma(t)$, time $t_{m \in \{0,1,\dots,M\}}$, projection function $\tau()$

`// Factorized score calculation`

Function $\texttt{FS}(\boldsymbol{y}_{nosiy}, \sigma)$**:**

  $\boldsymbol{y}_{denoised} = D_{\theta^{(0)}}(\boldsymbol{y}_{nosiy}, \sigma, \mathbf{0})$

  foreach $\boldsymbol{x}^{(i)} \in X$ **do**

    likelihood $= D_{\theta^{(i)}}(\boldsymbol{y}_{nosiy}, \sigma, \boldsymbol{x}^{(i)})$

    individual prior $= D_{\theta^{(i)}}(\boldsymbol{y}_{nosiy}, \sigma, \mathbf{0})$

    $\boldsymbol{y}_{denoised} = \boldsymbol{y}_{denoised} + \lambda($ likelihood $-$ individual prior $)$

  $\boldsymbol{y}_{denoised} = \tau(\boldsymbol{y}_{denoised})$

  score $= (\boldsymbol{y}_{denoised} - \boldsymbol{y}_{nosiy})/\sigma^2$

  return score

`// Main sampling algorithm`

sample $\boldsymbol{y}_0 \sim \mathcal{N}(0; \sigma(t_0)^2 I)$

for $m = 0$ **to** $M - 1$ **do**

  `// Gradient step`

  $d\boldsymbol{y} = -\sigma(t_m)\ \texttt{FS}(\boldsymbol{y}_m, \sigma(t_m))$

  $\boldsymbol{y}_{m+1} = \boldsymbol{y}_m + (t_{m+1} - t_m) d\boldsymbol{y}$

  `// Second order correction`

  if $\sigma(t_{m+1}) \neq 0$ **then**

    $d\boldsymbol{y}' = -\sigma(t_{m+1})\ \texttt{FS}(\boldsymbol{y}_{m+1}, \sigma(t_{m+1}))$

    $\boldsymbol{y}_{m+1} = \boldsymbol{y}_m + 0.5(t_{m+1} - t_m)(d\boldsymbol{y} + d\boldsymbol{y}')$

Output: $\boldsymbol{y}_M$

the same algorithm is used, however, p_{zero} is set to always be true.

The pseudo-code used to sample from an FSDM model is shown in algorithm 2. To use this algorithm, the denoising neural networks as trained using algorithm 1 are needed. We need $N + 1$ such neural networks. With $D_{\theta^{(1)}}, \dots, D_{\theta^{(N)}}$ corresponding to the N input signals, and $D_{\theta^{(0)}}$ the global prior network.

L. Computational cost

To analyze the computational cost of sampling from the system we timed the procedure 100 times on an NVIDIA GeForce RTX 3080 TI, to provide some indication of the computational burden of the system. Both theoretically and empirically, the computational complexity of sampling from our model can be described as:

$$T = (2n + 1) \cdot c, \tag{31}$$

with T the time it takes to sample from the model for a single recording, and c a hardware dependent constant. In the case of our machine with an NVIDIA GeForce RTX 3080 TI, c was found to be equal to 1.6 seconds. This means that for the recommended PSG ($n = 6$), our method takes about 20.8 seconds to come to an output.

TABLE S.IX
SLEEP STAGING RESULTS FOR SOME ADDITIONAL COMBINATIONS OF SENSORS. WE SHOW BOTH THE 5-CLASS (W/N1/N2/N3/REM) AND THE 4-CLASS (W/N1-N2/N3/REM) PERFORMANCE.

	Accuracy		Kappa	
Signals	5	4	5	4
Odd EEG	85.5	89.5	0.789	0.822
ECG	76.9	82.2	0.669	0.711
Chin1-ChinZ	74.9	80.9	0.630	0.677
Chin2-ChinZ	74.9	80.9	0.631	0.678
Chin1-Chin2	74.5	80.6	0.624	0.672
Odd EEG + ECG	85.5	89.5	0.788	0.822
ECG + Chin1-ChinZ	78.8	83.9	0.691	0.732
ECG + Chin2-ChinZ	78.7	83.9	0.689	0.732
ECG + Chin1-Chin2	78.6	84.0	0.688	0.735
Odd EEG + ECG + Chin1-ChinZ	85.2	89.1	0.783	0.816
Odd EEG + ECG + Chin2-ChinZ	85.1	89.1	0.782	0.816
Odd EEG + ECG + Chin1-Chin2	85.1	89.1	0.782	0.815

M. ADDITIONAL COMBINATIONS OF SENSORS

Table S.IX presents the sleep staging results for various combinations of sensors, specifically EEG, ECG, and EMG. Since EEG alone already reaches the upper limit of performance, equivalent to human inter-rater agreement, adding additional sensor modalities does not enhance its sleep staging performance. For instance, the combination of EEG and ECG achieves the same performance as EEG alone. Moreover, if an additional sensor introduces confusion, such as chin EMG, the combination with EEG may result in slightly lower performance. However, combining two sensors that have not yet reached the performance limit can be very beneficial. As shown in Table S.IX, the combination of ECG and chin EMG achieves better sleep staging performance than either sensor alone. This effect was also observed in the combinations tested in the main manuscript, where any combination involving EEG reached the performance limit, while HSAT combinations outperformed their individual constituent signals.

TABLE S.X

COMPARISON OF THE FSDM MODEL ON SOMNIA TO OTHER METHODS FROM LITERATURE THAT USE CARDIO-RESPIRATORY SIGNAL COMBINATIONS. 5-CLASS ENTAILS THE FULL AASM LABELING, 4-CLASS MERGES N1 AND N2 INTO A JOINT N1-N2 CLASS, AND 3-CLASS MERGES N1, N2, AND N3 INTO A JOINT NREM CLASS. CRESS [5] AND SUN ET AL. [55] EACH USED TWO DATASETS, RESULTING IN THE DUPLICATED ROWS FOR THE LATTER. NOTE THAT RESULTS BETWEEN THE METHODS ARE NOT DIRECTLY COMPARABLE DUE TO THE USE OF DIFFERENT DATASETS.

		Kappa [-]		
Method	**Signals**	**5-class**	**4-class**	**3-class**
FSDM	Finger PPG + Nasal cannula + Thoracic belt	0.697	0.740	0.801
FSDM	Finger PPG + Nasal cannula	0.686	0.731	0.731
FSDM	IHR from finger PPG + Nasal cannula	0.676	0.719	0.777
FSDM	ECG + Thermistor + Thoracic belt	0.687	0.733	0.797
FSDM	ECG + Thermistor	0.674	0.720	0.785
FSDM	Finger PPG + PAP flow + Thoracic belt	0.678	0.722	0.797
FSDM	Finger PPG + PAP flow	0.652	0.690	0.759
FSDM	IHR from finger PPG + IBR from PAP flow	0.626	0.666	0.719
Kazemi *et al.* [56] (2024)	Finger PPG + Respiratory flow + Abdominal belt + Thoracic belt	0.66	0.63	0.68
Kazemi *et al.* [56] (2024)	Finger PPG + Abdominal belt + Thoracic belt	0.68	0.67	0.71
Kazemi *et al.* [56] (2024)	Finger PPG + Respiratory flow	0.60	0.60	0.70
Huttunen *et al.* [57] (2023)	Finger PPG + Nasal cannula + SpO2	0.60	-	-
CReSS [5] (2021)	IHR from finger PPG + Nasal cannula + Thoracic belt	-	0.680	0.762
CReSS [5] (2021)	IHR from finger PPG + Nasal cannula	-	0.643	0.719
CReSS [5] (2021)	IHR from finger ECG + Thermistor + Thoracic belt	-	0.635	0.729
CReSS [5] (2021)	IHR from finger ECG + Thermistor	-	0.578	0.665
Sun *et al.* [55] (2020)	IHR from ECG + Thoracic belt	0.586	-	0.750
Sun *et al.* [55] (2020)	IHR from ECG + Abdominal belt	0.585	-	0.760
Sun *et al.* [55] (2020)	IHR from ECG + Thoracic belt	0.514	-	0.688
Sun *et al.* [55] (2020)	IHR from ECG + Abdominal belt	0.533	-	0.697

N. COMPARISONS OF CARDIO-RESPIRATORY RESULTS

We compare the results for the HSAT combinations to those found in the literature in Table S.X in terms of Cohen's kappa, as this metric is most often reported. Our FSDM method obtains the best sleep staging results for these signal combinations, although it should be noted that there are methodological differences between datasets, so direct comparisons should be interpreted with caution.

O. Additional results for subjects from Fig. 2

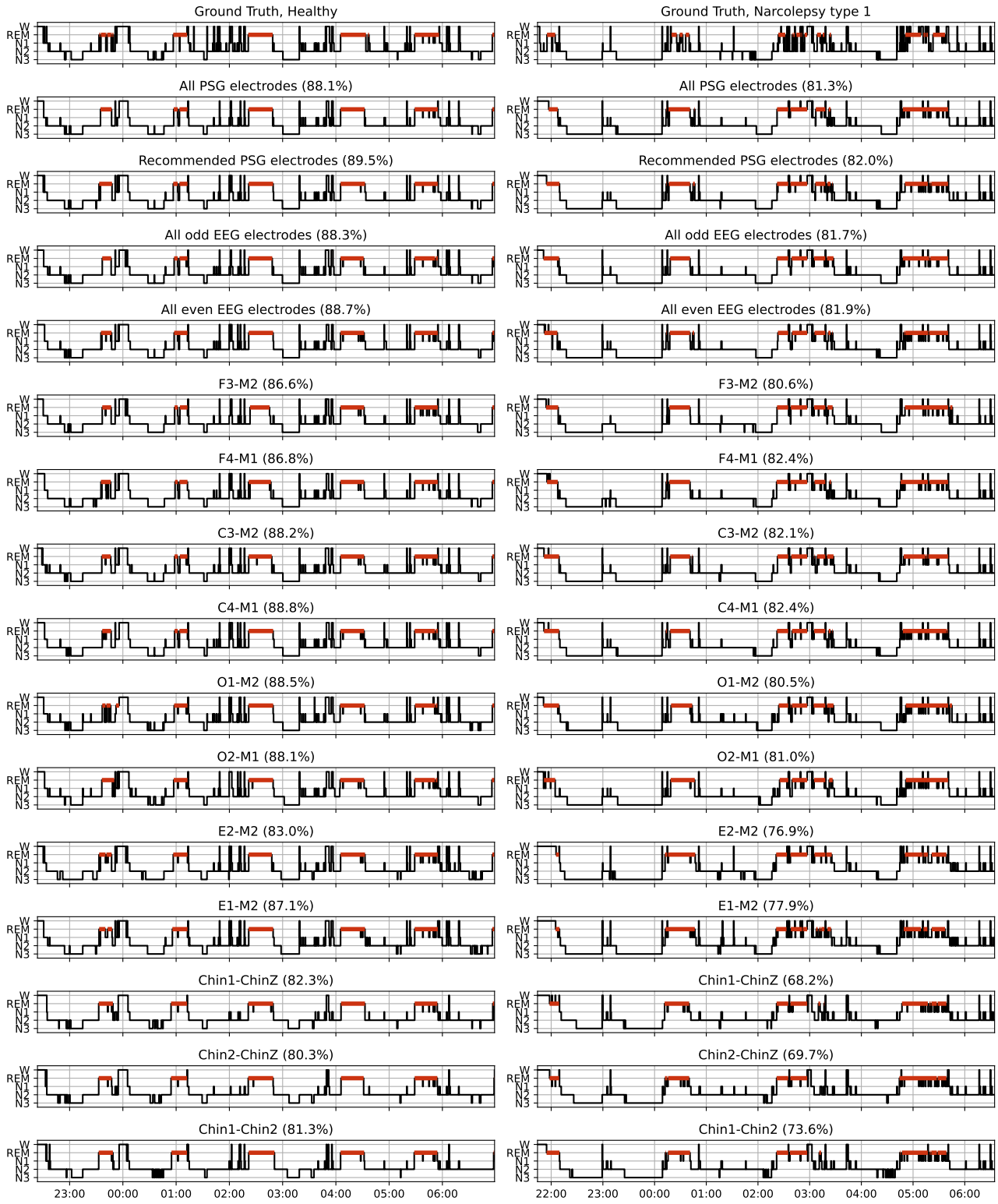


Fig. S.16. Additional Results for the two subjects from Fig. 2. Between brackets the 5 class accuracy for that specific recording is listed. Part 1 of 3.

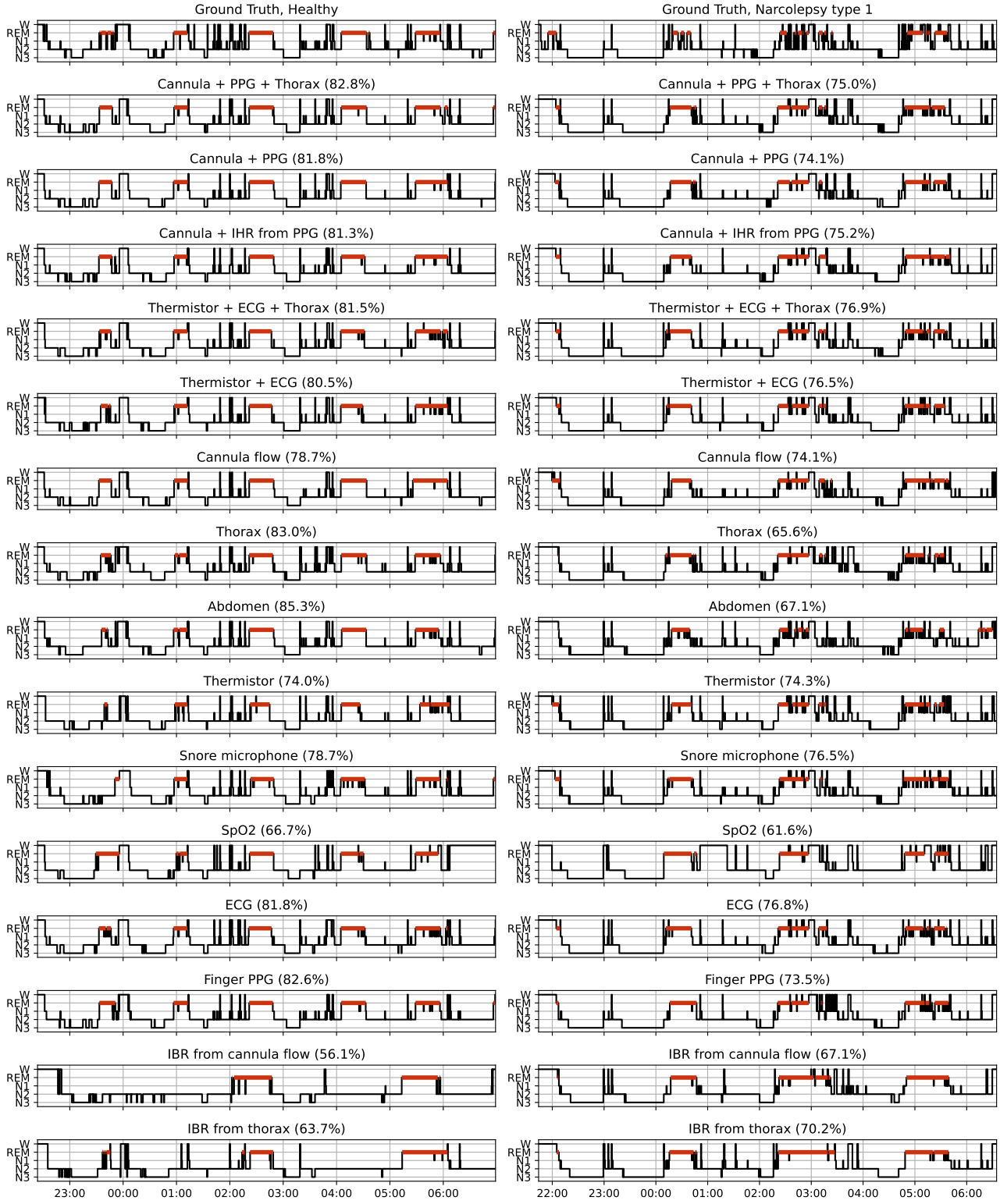


Fig. S.17. Additional Results for the two subjects from Fig. 2. Between brackets the 5 class accuracy for that specific recording is listed. Part 2 of 3.

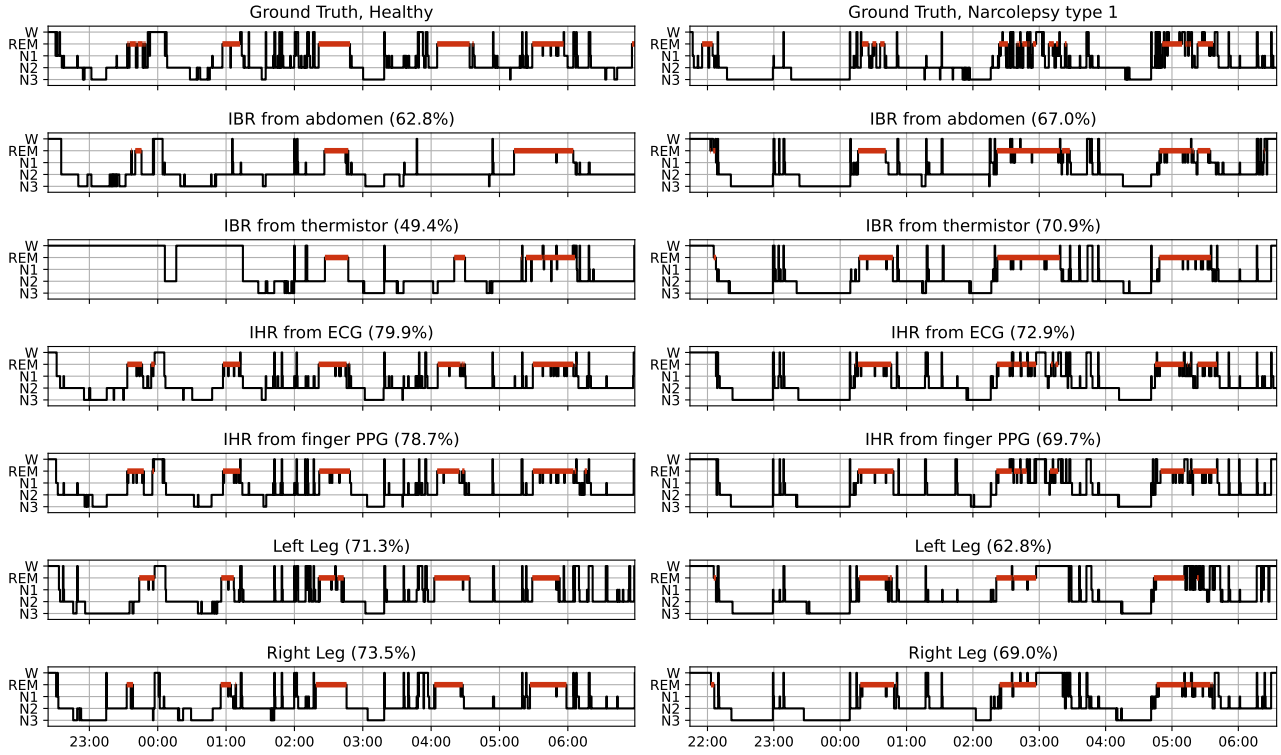


Fig. S.18. Additional Results for the two subjects from Fig. 2. Between brackets the 5 class accuracy for that specific recording is listed. Part 3 of 3.

REFERENCES

[1] M. M. Troester, S. F. Quan, and R. B. Berry, *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*, Version 3. Darien, IL, USA: American Academy of Sleep Medicine, 2023.

[2] C. van der Woerd *et al.*, "Studying sleep: towards the identification of hypnogram features that drive expert interpretation," *Sleep*, vol. 47, no. 3, zsad306, Dec. 2023, ISSN: 0161-8105.

[3] H. Phan and K. Mikkelsen, "Automatic sleep staging of EEG signals: Recent development, challenges, and future directions," *Physiological Measurement*, vol. 43, no. 4, 04TR01, Apr. 2022.

[4] P. Fonseca *et al.*, "A computationally efficient algorithm for wearable sleep staging in clinical populations," *Scientific Reports*, vol. 13, no. 1, p. 9182, Jun. 2023, ISSN: 2045-2322.

[5] J. P. Bakker *et al.*, "Estimating sleep stages using cardiorespiratory signals: Validation of a novel algorithm across a wide range of sleep-disordered breathing severity," *Journal of Clinical Sleep Medicine*, vol. 17, no. 7, pp. 1343–1354, Jul. 2021.

[6] H. Zhai, Y. Yan, S. He, P. Zhao, and B. Zhang, "Evaluation of the accuracy of contactless consumer sleep-tracking devices application in human experiment: A systematic review and meta-analysis," *Sensors*, vol. 23, no. 10, May 2023.

[7] L. Cerina *et al.*, "A sleep stage estimation algorithm based on cardiorespiratory signals derived from a suprasternal pressure sensor," *Journal of Sleep Research*, e14015, 2023.

[8] M. Perslev, S. Darkner, L. Kempfner, M. Nikolic, P. J. Jennum, and C. Igel, "U-Sleep: Resilient high-frequency sleep staging," *npj Digital Medicine*, vol. 4, no. 1, p. 72, Apr. 2021, ISSN: 2398-6352.

[9] L. Fiorillo *et al.*, "U-Sleep's resilience to AASM guidelines," *NPJ digital medicine*, vol. 6, no. 1, p. 33, 2023.

[10] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," in *International Conference on Learning Representations (ICLR)*, 2020.

[11] T. Karras, M. Aittala, T. Aila, and S. Laine, "Elucidating the design space of diffusion-based generative models," in *Advances in Neural Information Processing Systems*, vol. 35, Dec. 2022, pp. 26 565–26 577.

[12] P. Vincent, "A connection between score matching and denoising autoencoders," *Neural Computation*, vol. 23, no. 7, pp. 1661–1674, 2011.

[13] B. Efron, "Tweedie's formula and selection bias," *Journal of the American Statistical Association*, vol. 106, no. 496, pp. 1602–1614, 2011.

[14] P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," in *Advances in Neural Information Processing Systems*, vol. 34, 2021, pp. 8780–8794.

[15] J. Ho and T. Salimans, "Classifier-free diffusion guidance," in *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021.

[16] J. Song, A. Vahdat, M. Mardani, and J. Kautz, "Pseudoinverse-guided diffusion models for inverse problems," in *International Conference on Learning Representations*, 2023.

[17] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye, "Diffusion posterior sampling for general noisy inverse problems," *arXiv preprint arXiv:2209.14687*, 2022.

[18] H. van Gorp, M. M. van Gilst, P. Fonseca, S. Overeem, and R. J. G. van Sloun, "Modeling the impact of inter-rater disagreement on sleep statistics using deep generative learning," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–12, Aug. 2023.

[19] A. Caticha, "Entropic inference," in *AIP Conference Proceedings*, American Institute of Physics, vol. 1305, 2011, pp. 20–29.

[20] L. Fiorillo, D. Pedroncelli, V. Agostini, P. Favaro, and F. D. Faraci, "Multi-scored sleep databases: How to exploit the multiple-labels in automated sleep scoring," *Sleep*, vol. 46, no. 5, zsad028, 2023.

[21] P. Anderer, M. Ross, A. Cerny, R. Vasko, E. Shaw, and P. Fonseca, "Overview of the hypnodensity approach to scoring sleep for polysomnography and home sleep testing," *Frontiers in Sleep*, vol. 2, Apr. 2023.

[22] M. M. van Gilst *et al.*, "Protocol of the SOMNIA project: An observational study to create a neurophysiological database for advanced clinical sleep monitoring," *BMJ open*, vol. 9, no. 11, Nov. 2019.

[23] F. B. van Meulen *et al.*, "Contactless camera-based sleep staging: The healthbed study," *Bioengineering*, vol. 10, no. 1, 2023, ISSN: 2306-5354.

[24] H. van Gorp, M. M. van Gilst, P. Fonseca, S. Overeem, and R. J. G. van Sloun, "Single-channel EOG sleep staging on a heterogeneous cohort of subjects with sleep disorders," *Physiological measurement*, pp. 1–14, 2024.

[25] B. Kemp, A. H. Zwinderman, B. Tuk, H. A. Kamphuisen, and J. J. Oberye, "Analysis of a sleep-dependent neuronal feedback loop: The slow-wave microcontinuity of the EEG," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 9, pp. 1185–1194, 2000.

[26] A. L. Goldberger, L. A. Amaral, L. Glass, *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, e215–e220, 2000.

[27] A. Rechtschaffen and A. Kales, "A manual for standardized terminology, techniques and scoring system for sleep stages in human subjects," *Brain information service*, 1968.

[28] H. Phan, K. B. Mikkelsen, O. Chen, P. Koch, A. Mertins, and M. de Vos, "Sleeptransformer: Automatic sleep staging with interpretability and uncertainty quantification," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 8, pp. 2456–2467, 2022.

[29] American Academy of Sleep Medicine, *International classification of sleep disorders*, 3rd ed, text revision. Darien, IL, USA: American Academy of Sleep Medicine, 2023.

[30] J. Van Der Donckt, J. Van Der Donckt, E. Deprost, *et al.*, "Do not sleep on traditional machine learning: Simple and interpretable techniques are competitive to deep learning for sleep scoring," *Biomedical Signal Processing and Control*, vol. 81, p. 104 429, 2023.

[31] J. Wang, S. Zhao, H. Jiang, *et al.*, "Caresleepnet: A hybrid deep learning network for automatic sleep staging," *IEEE Journal of Biomedical and Health Informatics*, 2024.

[32] H. Phan, O. Y. Chén, M. C. Tran, P. Koch, A. Mertins, and M. De Vos, "XSleepNet: Multi-view sequential model for automatic sleep staging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 9, pp. 5903–5915, 2021.

[33] Q. Shen, J. Xin, X. Liu, *et al.*, "LGSleepNet: An automatic sleep staging model based on local and global representation learning," *IEEE Transactions on Instrumentation and Measurement*, 2023.

[34] A. Supratak and Y. Guo, "TinySleepNet: An efficient deep learning model for sleep stage scoring based on raw single-channel EEG," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, 2020, pp. 641–644.

[35] L. Fiorillo, P. Favaro, and F. D. Faraci, "Deepsleepnet-lite: A simplified automatic sleep stage scoring model with uncertainty estimates," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 29, pp. 2076–2085, 2021.

[36] M. Perslev, M. Jensen, S. Darkner, P. J. Jennum, and C. Igel, "U-time: A fully convolutional network for time series segmentation applied to sleep staging," *Advances in neural information processing systems*, vol. 32, 2019.

[37] S. Mousavi, F. Afghah, and U. R. Acharya, "SleepEEGNet: Automated sleep stage scoring with sequence to sequence deep learning approach," *PloS one*, vol. 14, no. 5, e0216456, 2019.

[38] F. Andreotti, H. Phan, and M. De Vos, "Visualising convolutional neural network decisions in automatic sleep scoring," in

CEUR Workshop Proceedings, CEUR Workshop Proceedings, 2018, pp. 70–81.

[39] M. Dutt, S. Redhu, M. Goodwin, and C. W. Omlin, "SleepXAI: An explainable deep learning approach for multi-class sleep stage identification," *Applied Intelligence*, vol. 53, no. 13, pp. 16 830–16 843, 2023.

[40] I. A. M. Huijben, S. Overeem, M. M. van Gilst, and R. J. G. van Sloun, "Attention on sleep stage specific characteristics," in *46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2024.

[41] A. Brink-Kjaer, K. M. Gunter, E. Mignot, E. During, P. Jennum, and H. B. D. Sorensen, "End-to-end deep learning of polysomnograms for classification of REM sleep behavior disorder," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2022, pp. 2941–2944.

[42] W. Cai, J. Jiang, F. Wang, J. Tang, S. Kim, and J. Huang, "A survey on mixture of experts," *Authorea Preprints*, 2024.

[43] H. van Gorp, I. A. M. Huijben, P. Fonseca, R. J. G. van Sloun, S. Overeem, and M. M. van Gilst, "Certainty about uncertainty in sleep staging: a theoretical framework," *Sleep*, vol. 45, no. 8, Jun. 2022.

[44] R. S. Rosenberg and S. Van Hout, "The American Academy of Sleep Medicine inter-scorer reliability program: sleep stage scoring," *Journal of Clinical Sleep Medicine*, vol. 9, no. 1, pp. 81–87, Jan. 2013.

[45] H. Zhu, C. Fu, F. Shu, H. Yu, C. Chen, and W. Chen, "The effect of coupled electroencephalography signals in electrooculography signals on sleep staging based on deep learning methods," *Bioengineering*, vol. 10, no. 5, p. 573, 2023.

[46] Y. Song, P. Dhariwal, M. Chen, and I. Sutskever, "Consistency models," in *Proceedings of the 40th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 202, PMLR, Jul. 2023, pp. 32 211–32 252.

[47] J. Kohler *et al.*, "Imagine flash: Accelerating emu diffusion models with backward distillation," *Facebook preprint*, Apr. 2024.

[48] F. Scholkmann, J. Boss, and M. Wolf, "An efficient algorithm for automatic peak detection in noisy periodic and quasi-periodic signals," *Algorithms*, vol. 5, no. 4, pp. 588–603, 2012.

[49] M. Baumert, S. Hartmann, and H. Phan, "Automatic sleep staging for the young and the old – evaluating age bias in deep learning," *Sleep Medicine*, vol. 107, pp. 18–25, 2023, ISSN: 1389-9457.

[50] J. B. Stephansen *et al.*, "Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy," *Nature Communications*, vol. 9, no. 1, p. 5229, Dec. 2018, ISSN: 2041-1723.

[51] Y. Wu and K. He, "Group normalization," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[52] D. Hendrycks and K. Gimpel, "Gaussian error linear units (gelus)," *arXiv preprint arXiv:1606.08415*, 2016.

[53] J. Tompson, R. Goroshin, A. Jain, Y. LeCun, and C. Bregler, "Efficient object localization using convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 648–656.

[54] A. Vaswani *et al.*, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[55] H. Sun, W. Ganglberger, E. Panneerselvam, *et al.*, "Sleep staging from electrocardiography and respiration with deep learning," *Sleep*, vol. 43, no. 7, zsz306, 2020.

[56] K. Kazemi, A. Abiri, Y. Zhou, *et al.*, "Improved sleep stage predictions by deep learning of photoplethysmogram and respiration patterns," *Computers in Biology and Medicine*, vol. 179, p. 108 679, 2024.

[57] R. Huttunen, T. Leppänen, B. Duce, *et al.*, "A comparison of signal combinations for deep learning-based simultaneous sleep staging and respiratory event detection," *IEEE Transactions on Biomedical Engineering*, vol. 70, no. 5, pp. 1704–1714, 2022.