

# Supplementary Material for DOI:

2023-07-07

Table 1: Summary of Bias, Limits of Agreement (LOA), and Pearson Correlation for various Sleep Parameter Predictions (SPT, TST, SE, LPS, WASO) using different Machine Learning Models (Decision Tree, Logistic Regression, Feed-Forward Neural Net, XG-Boost) with Raw ZM Predictions, 5-Min and 10-Min Median as predictors. Each value is provided with its 95% Confidence Interval (CI).

	Bias (95% CI)	LOA (95% CI)	LOA (95% CI)	Pearson, $r$ (95% CI)
Raw ZM Predictions - Decision Tree				
SPT (min)	-21.6 (-25.6;-17.6)	-117.5 (-125.6;-110.7)	74.2 (63.9;85.9)	0.54 (0.48;0.6)
TST (min)	-148 (-153.8;-142.4)	-283 (-295.5;-272.6)	-13.1 (-22.8;-1)	0.3 (0.22;0.37)
SE (%)	-22.7 (-23.7;-21.8)	-45.5 (-47.5;-43.8)	0 (-1.6;1.9)	0.17 (0.09;0.24)
LPS (min)	28.9 (24.5;33.2)	-76 (-87.6;-69.8)	133.8 (124.6;144.7)	0.13 (0.05;0.21)
WASO (min)	46.1 (43;49.4)	-33.2 (-43.4;-26.2)	125.4 (117.7;138.8)	0.29 (0.22;0.37)
5-Min Median - Decision Tree				
SPT (min)	-21.6 (-25.6;-17.6)	-117.5 (-125.6;-110.7)	74.2 (63.9;85.9)	0.54 (0.48;0.6)
TST (min)	-50.5 (-55.2;-46)	-161.4 (-175.8;-151.3)	60.4 (51.5;71.7)	0.48 (0.42;0.54)
SE (%)	-5.5 (-6.3;-4.7)	-23.9 (-26.4;-22.2)	12.9 (11.6;14.6)	0.22 (0.14;0.29)
LPS (min)	24.6 (19.7;29.1)	-88.8 (-115;-77.3)	138 (126.2;156.7)	0.06 (-0.02;0.14)
WASO (min)	9.9 (6.5;14)	-79.4 (-109;-63.1)	99.2 (80;136.1)	0.15 (0.07;0.22)
10-Min Median - Decision Tree				
SPT (min)	-21.8 (-25.7;-17.8)	-117.3 (-125.2;-110.4)	73.7 (63.4;85.4)	0.54 (0.48;0.6)
TST (min)	-31.5 (-35.7;-27.4)	-129.9 (-140.9;-121.8)	67 (58.3;77.7)	0.56 (0.5;0.61)
SE (%)	-2.1 (-2.8;-1.4)	-18 (-19.9;-16.6)	13.9 (12.6;15.3)	0.22 (0.14;0.29)
LPS (min)	22.8 (17.1;27.6)	-102.7 (-137.1;-83.3)	148.4 (131.9;173.6)	0.06 (-0.02;0.14)
WASO (min)	9 (5.2;14.3)	-97.4 (-133.5;-72.1)	115.3 (85.2;163)	0.07 (-0.02;0.15)
Raw ZM Predictions - Logistic Regression				
SPT (min)	-4 (-8.3;0.7)	-113.5 (-122.7;-106.1)	105.5 (95;118.7)	0.37 (0.29;0.43)

TST (min)	-139.2 (-145.7;-132.8)	-291.6 (-306.1;-279.2)	13.1 (3.8;23.5)	0.12 (0.04;0.2)
SE (%)	-23.1 (-24;-22.1)	-45.6 (-47.4;-44)	-0.6 (-2;1)	0.18 (0.1;0.26)
LPS (min)	47.5 (43.6;51.4)	-46.2 (-57;-38.6)	141.2 (131;154.5)	0.1 (0.01;0.18)
WASO (min)	48.7 (45.3;52.1)	-34.7 (-46.2;-28)	132.1 (124.6;147.3)	0.25 (0.17;0.33)
5-Min Median - Logistic Regression				
SPT (min)	-3.7 (-8;1)	-112.2 (-120.9;-105.2)	104.8 (94;117.4)	0.38 (0.3;0.44)
TST (min)	-139.7 (-146.9;-133)	-305.6 (-323.6;-291.8)	26.2 (16.1;38.6)	0.09 (0.01;0.17)
SE (%)	-23.2 (-24.3;-22.2)	-48.1 (-50.9;-46.1)	1.7 (0.1;3.8)	0.13 (0.05;0.21)
LPS (min)	58.1 (53.4;62.6)	-52.3 (-75;-40.1)	168.6 (155.9;187.7)	0.05 (-0.03;0.13)
WASO (min)	45.4 (41.7;49.7)	-50.7 (-74.4;-38.4)	141.5 (126.8;173)	0.19 (0.11;0.27)
10-Min Median - Logistic Regression				
SPT (min)	-4.2 (-8.6;0.5)	-113.4 (-122.4;-106)	105 (94.2;118)	0.37 (0.3;0.44)
TST (min)	-130.9 (-138;-124.2)	-295.1 (-311.8;-281.4)	33.2 (23.3;45.1)	0.09 (0.01;0.17)
SE (%)	-21.6 (-22.6;-20.6)	-45.7 (-48.2;-43.8)	2.5 (1;4.3)	0.13 (0.05;0.21)
LPS (min)	60.7 (54.9;65.6)	-64.8 (-100.8;-43.9)	186.2 (168.1;213.6)	0.02 (-0.06;0.1)
WASO (min)	44.8 (40.8;50)	-66 (-98.3;-45.1)	155.7 (130.2;197.8)	0.17 (0.09;0.25)
Raw ZM Predictions - Feed-Forward Neural Net				
SPT (min)	-3.9 (-8.1;0.9)	-112.7 (-122;-105.2)	104.9 (94.1;118.4)	0.38 (0.3;0.44)
TST (min)	-154 (-159.9;-148)	-297 (-308.6;-287)	-10.9 (-20;-0.5)	0.25 (0.17;0.32)
SE (%)	-25.6 (-26.5;-24.7)	-48.2 (-50;-46.6)	-3 (-4.5;-1.2)	0.23 (0.15;0.31)
LPS (min)	34.3 (30.2;38.6)	-67.7 (-80.1;-60.5)	136.4 (126.4;149.3)	0.11 (0.03;0.19)
WASO (min)	58.7 (55.4;62.1)	-23.8 (-33.9;-17.5)	141.2 (133.9;155.6)	0.33 (0.26;0.4)
5-Min Median - Feed-Forward Neural Net				
SPT (min)	-3.9 (-8.1;0.9)	-112.7 (-122;-105.2)	104.9 (94.1;118.4)	0.38 (0.3;0.44)
TST (min)	-126.5 (-132.8;-120.3)	-276.8 (-291.3;-264.7)	23.9 (14.8;33.9)	0.25 (0.17;0.32)
SE (%)	-20.9 (-21.9;-19.9)	-44.3 (-46.3;-42.5)	2.5 (1.1;4)	0.21 (0.13;0.29)
LPS (min)	35.3 (30.7;39.8)	-75.8 (-102.3;-63.4)	146.5 (134.4;166.9)	0.07 (-0.01;0.15)
WASO (min)	45 (41.2;49.2)	-51.8 (-76.4;-39.1)	141.7 (125.8;174.1)	0.21 (0.14;0.29)
10-Min Median - Feed-Forward Neural Net				
SPT (min)	-4.1 (-8.5;0.6)	-112.6 (-121.7;-105)	104.5 (93.5;117.6)	0.38 (0.31;0.45)
TST (min)	-116.3 (-122.9;-110.3)	-266.2 (-280.4;-254.2)	33.6 (24.7;43.4)	0.29 (0.21;0.36)
SE (%)	-19.1 (-20.1;-18.1)	-42.9 (-44.8;-41.1)	4.7 (3.3;6.2)	0.25 (0.17;0.33)
LPS (min)	33.8 (28;38.6)	-91.1 (-127.2;-70.2)	158.6 (141.2;184.7)	0.05 (-0.03;0.13)
WASO (min)	53.4 (49.2;58.7)	-58.6 (-89.6;-38.6)	165.4 (140.4;206.7)	0.22 (0.14;0.3)
Raw ZM Predictions - XGboost				
SPT (min)	0.2 (-3.7;4.5)	-97.4 (-106.2;-90.3)	97.8 (86.6;111)	0.56 (0.5;0.61)
TST (min)	-66 (-70.8;-61.4)	-178.1 (-187.9;-169.6)	46.1 (38.9;54.5)	0.47 (0.4;0.53)
SE (%)	-11.1 (-11.8;-10.4)	-28.8 (-30.2;-27.5)	6.5 (5.5;7.7)	0.37 (0.29;0.44)

LPS (min)	34.5 (30.6;38.5)	-62.4 (-75.8;-55.2)	131.3 (121.1;143.9)	0.2 (0.12;0.28)
WASO (min)	18.4 (15.6;21.2)	-50.2 (-62.7;-43.1)	86.9 (79.8;104.2)	0.36 (0.28;0.43)
<b>5-Min Median - XGboost</b>				
SPT (min)	0.2 (-3.7;4.5)	-97.4 (-106.2;-90.3)	97.8 (86.6;111)	0.56 (0.5;0.61)
TST (min)	-7 (-10.8;-3.3)	-95.5 (-105.2;-88)	81.4 (72.4;92.5)	0.66 (0.61;0.7)
SE (%)	-1.1 (-1.7;-0.5)	-15.6 (-17;-14.4)	13.3 (12.2;14.7)	0.44 (0.38;0.51)
LPS (min)	28.5 (23.9;32.6)	-76.4 (-104.2;-63.3)	133.4 (120.4;154.2)	0.12 (0.04;0.2)
WASO (min)	-0.9 (-3.9;3)	-83.4 (-113.1;-66)	81.7 (62;119.6)	0.26 (0.18;0.33)
<b>10-Min Median - XGboost</b>				
SPT (min)	0.2 (-3.8;4.4)	-97.4 (-106.1;-90)	97.9 (86.7;111.1)	0.56 (0.5;0.61)
TST (min)	-4.2 (-7.7;-0.5)	-90.6 (-101.3;-82.9)	82.3 (72.3;95.3)	0.67 (0.62;0.71)
SE (%)	-0.6 (-1.2;-0.1)	-14.5 (-16;-13.3)	13.2 (12.1;14.9)	0.43 (0.36;0.49)
LPS (min)	26.4 (21;30.8)	-92.2 (-130;-69.5)	145 (125.5;173.7)	0.1 (0.02;0.18)
WASO (min)	3.8 (0.3;9.1)	-98.1 (-135.4;-71.9)	105.7 (74.5;153.4)	0.2 (0.13;0.28)
<b>Raw ZM Predictions - biLSTM</b>				
SPT (min)	-36.7 (-42.6;-30.3)	-141.4 (-153.2;-132)	68 (54.5;85.5)	0.5 (0.4;0.58)
TST (min)	39 (33.3;44.9)	-60.1 (-72.9;-51.1)	138 (126;152)	0.53 (0.44;0.61)
SE (%)	12.6 (11.8;13.3)	0 (-1.6;1.1)	25.2 (23.6;27.2)	0.07 (-0.05;0.18)
LPS (min)	-17.6 (-24.1;-11.3)	-127.2 (-177.4;-97.4)	92.1 (63.4;143.8)	0.05 (-0.06;0.17)
WASO (min)	-15.9 (-21;-8.9)	-116.1 (-158.9;-95.4)	84.3 (58.9;138.2)	0.04 (-0.07;0.16)
<b>5-Min Median - biLSTM</b>				
SPT (min)	-36.1 (-41.7;-30)	-136.1 (-146.3;-126.9)	64 (51.1;78.6)	0.54 (0.45;0.62)
TST (min)	12.8 (7.4;18.3)	-80.1 (-89.8;-72.3)	105.8 (94.3;118.8)	0.63 (0.55;0.69)
SE (%)	8 (7.2;8.8)	-5.1 (-6.8;-3.8)	21.1 (19.5;23.1)	0.16 (0.04;0.27)
LPS (min)	-15.7 (-25.9;-7.5)	-169 (-230.7;-127.9)	137.6 (101.1;184.9)	0.09 (-0.02;0.2)
WASO (min)	-3 (-9.9;7.7)	-144.1 (-197.2;-107.2)	138.1 (90.8;211.4)	0.02 (-0.1;0.13)
<b>10-Min Median - biLSTM</b>				
SPT (min)	-83.7 (-90.7;-76.1)	-207.4 (-221.3;-195.2)	40 (27.7;57.1)	0.3 (0.19;0.4)
TST (min)	-42.2 (-49.3;-35.1)	-162 (-176.9;-149.6)	77.6 (66.3;90.4)	0.4 (0.29;0.49)
SE (%)	6.4 (5.7;7.2)	-6.5 (-7.8;-5.3)	19.2 (17.7;21.2)	0.16 (0.04;0.27)
LPS (min)	-21.5 (-32.7;-12.8)	-187.2 (-253.4;-138.6)	144.3 (104.3;192.6)	0.06 (-0.05;0.18)
WASO (min)	26.8 (19.2;38)	-128.2 (-176.3;-90.8)	181.8 (132.8;250.7)	0.12 (0.01;0.23)

The analysis of precision-recall and ROC curves across different models and ZM prediction types shows varying performance. In terms of precision-recall AUC, the Decision Tree model consistently outperforms others, indicating its superior predictive accuracy (see Figure 1). Conversely, the Neural Network model generally shows weaker performance. However, for ROC AUC, the XGBoost model consistently excels across all data types, indicating a strong ability

to differentiate between classes, while the Neural Network model tends to underperform (see Figure 2). The F-measure (F1 score) shows variable performance across different configurations but generally, the Decision Tree model yields higher scores.

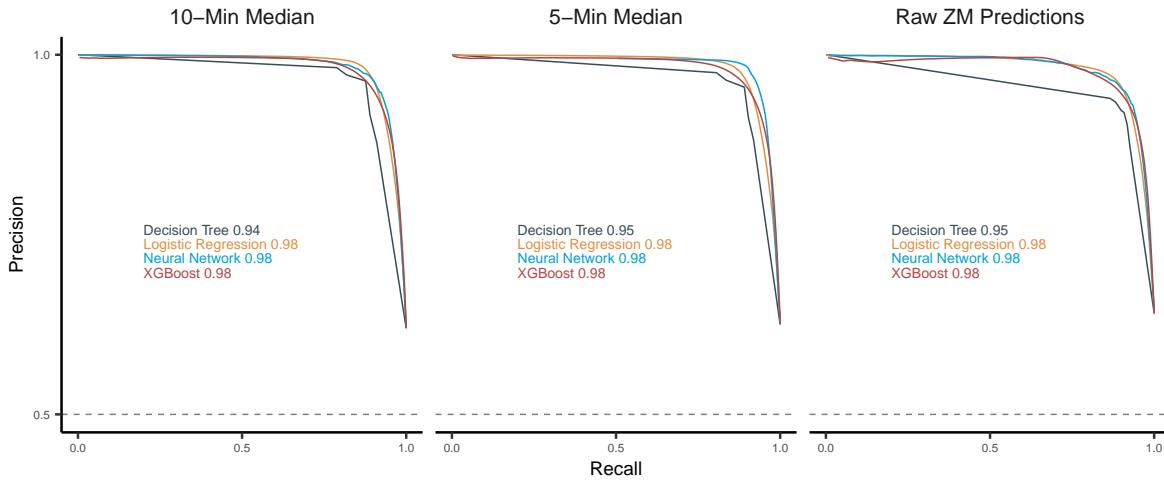


Figure 1: Precision-recall curves of the models evaluated across the different ZM predictions, including raw ZM predictions, as well as 5-minute and 10-minute median smoothing of the ZM raw predictions. The x-axis of the plot represents the proportion of true wake epochs that were correctly classified as wake, while the y-axis represents the proportion of all epochs labeled as wake by the classifier that were classified correctly. The area under the curve values are displayed as color-coded text in the plot indicate the area under the Precision-Recall curve for each model and condition.

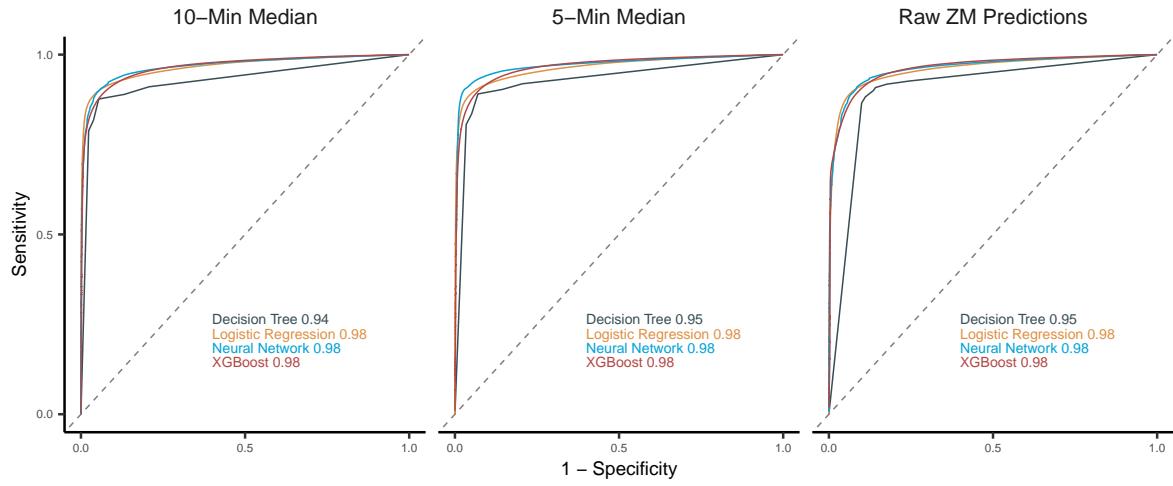


Figure 2: Receiver operating characteristic curves of the models evaluated across the different ZM predictions, including raw ZM predictions, as well as 5-minute and 10-minute median smoothing of the ZM raw predictions. The x-axis of the plot represents the proportion of true asleep epochs that were incorrectly classified as awake, while the y-axis represents the proportion of all epochs labeled as awake by the classifier that were correctly classified. The area under the curve values displayed are displayed as color-coded text in the plot to indicate the area under the receiver operating characteristic curve for each model and condition.

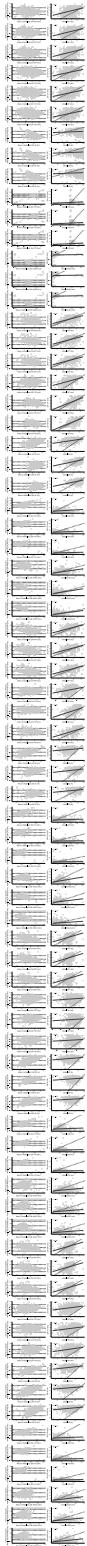


Figure 3: all ba and scatterplots