Automated sleep staging with the Waveband at-home EEG system achieves or exceeds expert-level performance

Jason Manley, Eric Hanson, Kendal Sandridge, Jack Sampiere, Antoine Guillot, Alexander M. Chan
Beacon Biosignals, Boston, MA

BEACONBIOSIGNALS

P-01 Board 23, #361

We demonstrate that Waveband — an FDA-cleared, at-home, dry-electrode EEG device — captures sleep physiology with expert-level performance, with targeted improvements in REM and wake accuracy delivered through an FDA-authorized Predetermined Change Control Plan (PCCP) update to the original 510(k)-cleared algorithm.

Introduction

Disturbed sleep is a hallmark of many neuropsychiatric conditions, but the gold standard for assessment — in-clinic polysomnography (PSG) scored by expert technologists — is costly, burdensome to patients, and subject to inter-rater variability. Waveband is an FDA-cleared, at-home, dry-electrode EEG device that captures overnight sleep physiology with automated sleep staging. Here, we evaluate an updated algorithm deployed under an FDA-authorized PCCP, comparing its performance against six registered PSG technologists (RPSGTs) on simultaneously acquired recordings.

Methods

OCTAVE-3 study (NCT05438017): 38 subjects with simultaneous overnight Waveband and in-clinic PSG recordings (ages 23-66; 50% female).

Reference scoring: each 30-second epoch scored by six RPSGTs per AASM guidelines. Consensus computed as weighted majority consensus across all scorers. Each scorer is weighted by their average agreement with the other scorers.

Leave-one-out evaluation: each rater (expert or algorithm) compared against weighted majority consensus of the remaining 5 experts, yielding leave-one-out estimates.

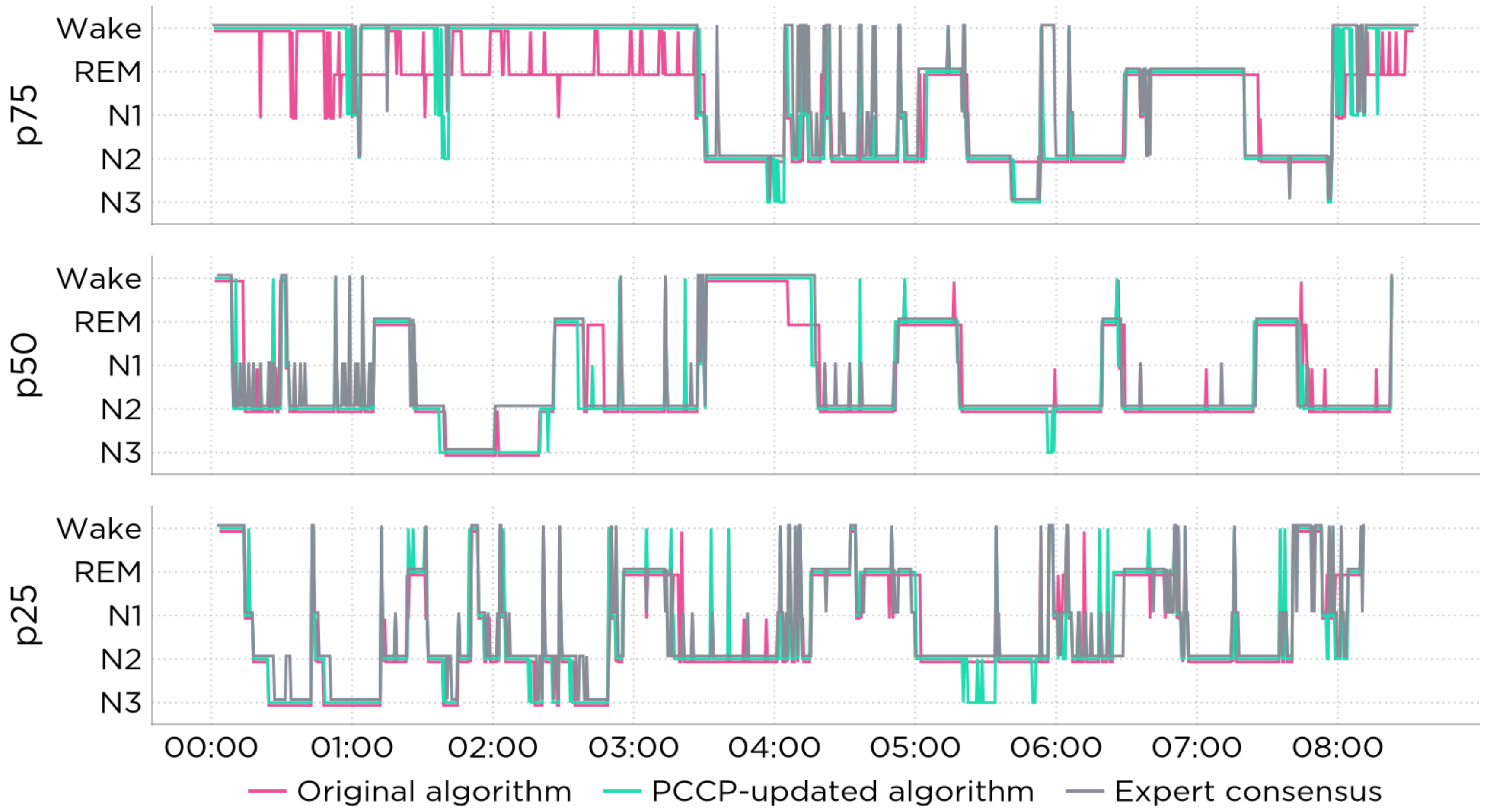

p75
p50
p25
Wake
REM
N1
N2
N3
00:00
01:00
02:00
03:00
04:00
05:00
06:00
07:00
08:00
Original algorithm
PCCP-updated algorithm
Expert consensus

Figure 1: Example hypnograms across the distribution of algorithm performance. Three overnight recordings selected at the 25th, 50th, and 75th percentiles of PCCP-Updated algorithm overall agreement with expert consensus on OCTAVE-3. Each panel overlays the Original algorithm (pink), the PCCP-Updated algorithm (teal), and the weighted majority consensus of 6 experts (gray). The PCCP-Updated algorithm tracks expert consensus closely across the performance distribution; the largest visible improvements over the Original occur in REM and wake detection during the initial sleep period (p75, top).

Results

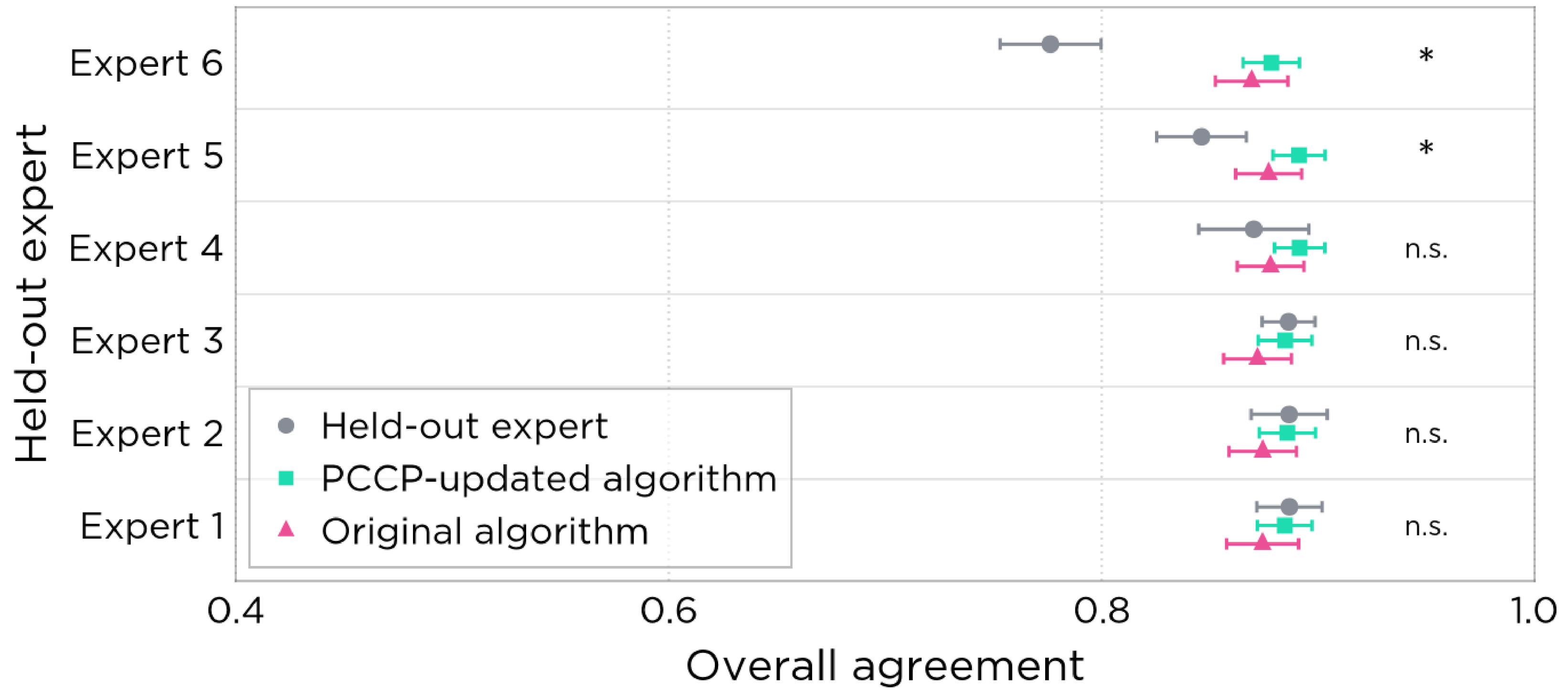

Held-out expert
Expert 6
Expert 5
Expert 4
Expert 3
Expert 2
Expert 1
*
*
n.s.
n.s.
n.s.
n.s.
Held-out expert
PCCP-updated algorithm
Original algorithm
0.4
0.6
0.8
1.0
Overall agreement

Figure 2: Overall epoch-level agreement matches or exceeds individual experts. For each held-out expert (rows), overall agreement (mean ± 95% CI across n=38 OCTAVE-3 recordings) is shown for that expert (gray), the Original algorithm (pink), and the PCCP-Updated algorithm (teal), each compared against the weighted majority consensus of the remaining 5 experts. The PCCP-Updated algorithm significantly exceeds 2 experts (*, $p<0.05$; one-sided paired Wilcoxon signed-rank test) and is statistically indistinguishable from the remaining 4 (n.s., $p>0.05$; two-sided). The point estimate of the PCCP-Updated algorithm also exceeds the Original algorithm across all 6 leave-one-out comparisons.

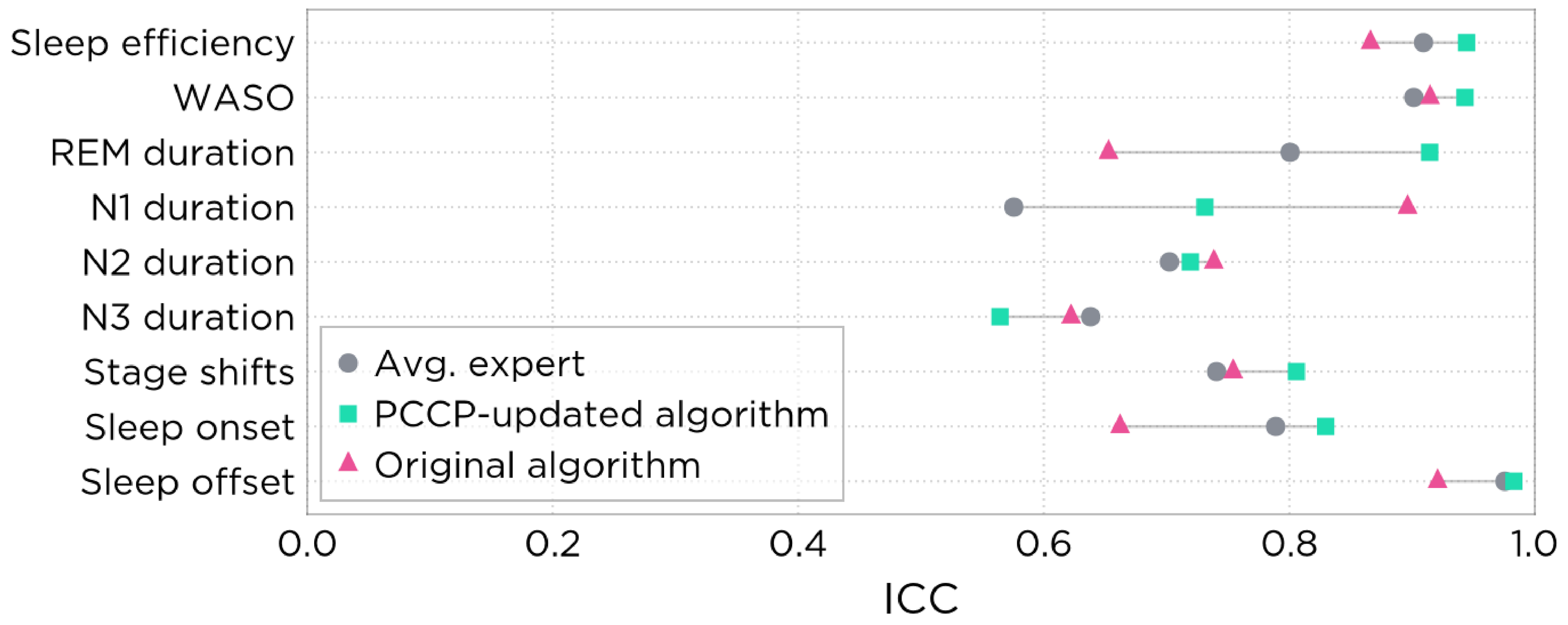

Sleep efficiency
WASO
REM duration
N1 duration
N2 duration
N3 duration
Stage shifts
Sleep onset
Sleep offset
Avg. expert
PCCP-updated algorithm
Original algorithm
0.0
0.2
0.4
0.6
0.8
1.0
ICC

Figure 3: Macro-quantity reliability vs. PSG expert consensus. Intraclass correlation coefficients (ICC) for sleep macroquantities in the OCTAVE-3 dataset (n=38 recordings). The point estimate for the PCCP-Updated algorithm matches or exceeds that of the average human expert on 7 of 9 metrics, with the largest improvement over the Original algorithm in REM duration (ICC 0.65 → 0.91). Avg. expert: each scorer vs. weighted majority consensus of the remaining 5 experts, averaged across scorers. Original and PCCP-Updated algorithms: vs. weighted majority consensus of all 6 experts. Grey lines span the observed range across the three predictors for each macroquantity.

PCCP Validation

The updated algorithm was validated under an FDA-authorized PCCP on a newly-collected, independent dataset of 50 subjects from four clinical sites (ages 22-65; 56% female). Staging was performed by three RPSGTs, and both the original and the PCCP-updated algorithms were compared against their weighted majority consensus.

Updated algorithm (rows: Predicted class; columns: Consensus class)

Predicted class	wake	n1	n2	n3	rem
wake	**87.1%**	12.5%	1.3%	0.3%	1.6%
n1	5.8%	**50.0%**	1.7%	0.0%	0.7%
n2	6.0%	34.2%	**90.6%**	18.3%	4.8%
n3	0.2%	0.3%	4.5%	**81.3%**	0.0%
rem	1.0%	3.0%	1.9%	0.1%	**92.9%**

Updated − Original (rows: Predicted class; columns: Consensus class)

Predicted class	wake	n1	n2	n3	rem
wake	**+6.2%**	+4.1%	+0.3%	−0.2%	+0.2%
n1	−4.0%	**0.0%**	−0.4%	0.0%	−0.5%
n2	+1.1%	−0.3%	**+2.6%**	−2.8%	−1.4%
n3	−0.2%	−0.5%	−1.6%	**+3.2%**	0.0%
rem	−2.9%	−3.3%	−0.9%	−0.3%	**+1.8%**

Figure 4: Confusion matrix and associated improvement on the PCCP validation dataset. [A] Sleep stage confusion matrix for the PCCP-updated algorithm against weighted majority consensus (n=41,705 epochs). Each column is normalized to sum to 1, such that diagonal values indicate per-stage positive percent agreement (PPA). [B] Element-wise difference (PCCP-updated − original) between the two confusion matrices. Blue indicates an improvement in the PCCP-update algorithm, whereas red indicates a regression. The PCCP-updated algorithm was non-inferior to the Original algorithm on each individual stage and overall agreement, with the largest gains in wake (+6.2pp) and N3 (+3.2pp). Overall multi-stage agreement improved by +3.3pp.

Conclusions

- → The PCCP-Updated algorithm achieves expert-level epoch agreement and clinically reliable macroquantities on OCTAVE-3.
- → On a held-out cohort, the Updated algorithm shows PCCP-validated improvements over the Original, with largest gains in clinically actionable stages (wake, N3, REM).
- → Demonstrates that PCCP enables continuous, FDA-authorized improvement of AI/ML in cleared medical devices.

Contact

Jason Manley, PhD
Algorithm Engineer,
Beacon Biosignals
jason.manley@beacon.bio