

Supplementary Materials

Tables

Supplementary Table 1: List of hand-crafted features extracted by the Walmsley accelerometer¹ pip package for the baseline random forest² model. ENMO is the Euclidean Norm Minus One³. g is acceleration due to gravity. v is the Euclidean norm of the signal in x , y and z . nats are units of information (e.g., entropy), Hz is frequency per second, and s is seconds.

Feature Name	Description	Unit
mean	Mean value of v	g
sd	Standard deviation of v	g
min, max	Minimum and maximum of v	g
25thp, median, 75thp	25th, 50th, 75th percentile of v	g
coefvariation	Coefficient of variation of v	
MAD	Mean absolute deviation of v	g
MPD	Mean pairwise difference of v	g
skew, kurt	Skewness, Kurtosis of v	
enmoTrunc, enmoAbs	Truncated, Absolute ENMO	g
autocorr	Auto-correlation of v	
fmax	Maximum frequency	Hz
pmax	Maximum power	g^2/s
fmaxband	Frequency band of maximum power	Hz
pmaxband	Power of maximum frequency band	g^2/s
entropy	Spectral entropy	nats
fft1, fft2, ..., fft10	Power spectral density for frequencies 1Hz, 2Hz, ..., 10Hz	g^2/s
f1, f2	1st and 2nd dominant frequencies	Hz
p1, p2	Power spectral densities of respective dominant frequencies	g^2/s
f625	Dominant frequency between 0.6 and 2.5Hz	Hz
p625	Power at dominant frequency between 0.6 and 2.5 Hz	g^2/s
totalPower	Total power	g^2/s

Supplementary Table 2: Test participant IDs for each fold in the cross-validation training and evaluation on the CAPTURE-24⁺ dataset in this study. The remaining participant IDs (P001-P151) not listed as test IDs for the respective fold were used for model training, preventing data leakage.

Fold	Test Participant IDs
1	P001, P003, P006, P010, P012, P013, P017, P025, P030, P044, P056, P057, P059, P065, P070, P075, P079, P080, P082, P090, P092, P114, P115, P116, P117, P130, P131, P142, P145
2	P004, P008, P014, P015, P016, P021, P023, P024, P028, P029, P032, P035, P037, P038, P043, P047, P048, P050, P074, P081, P085, P093, P094, P098, P099, P112, P118, P121, P127, P128, P146
3	P026, P036, P039, P040, P042, P045, P046, P049, P052, P055, P064, P068, P069, P071, P072, P088, P091, P095, P100, P101, P103, P106, P109, P111, P126, P141, P144, P148, P149, P150, P151
4	P002, P009, P011, P018, P019, P027, P031, P051, P053, P054, P060, P061, P067, P076, P077, P078, P083, P086, P108, P110, P113, P119, P122, P123, P132, P133, P135, P136, P138, P139, P147
5	P005, P007, P020, P022, P033, P034, P041, P058, P062, P063, P066, P073, P084, P087, P089, P096, P097, P102, P104, P105, P107, P120, P124, P125, P129, P134, P137, P140, P143

Hidden Markov models

Below shows the transition, and emission matrices from the training of ActiNet model on the full CAPTURE-24 population. Transition matrices represent the probability of moving from a state **S** at time **t** to a state at time **t+1**. Emission matrices represent the probability of observing an activity intensity **O** at time **t** given the state **S** at time **t**. A state **S** and corresponding observation **O** respectively represent the actual and predicted activity intensity label: sleep, sedentary, light or moderate-vigorous activity (MVPA), for a 30-second time window. These models assume a uniform prior distribution.

Transition matrix:

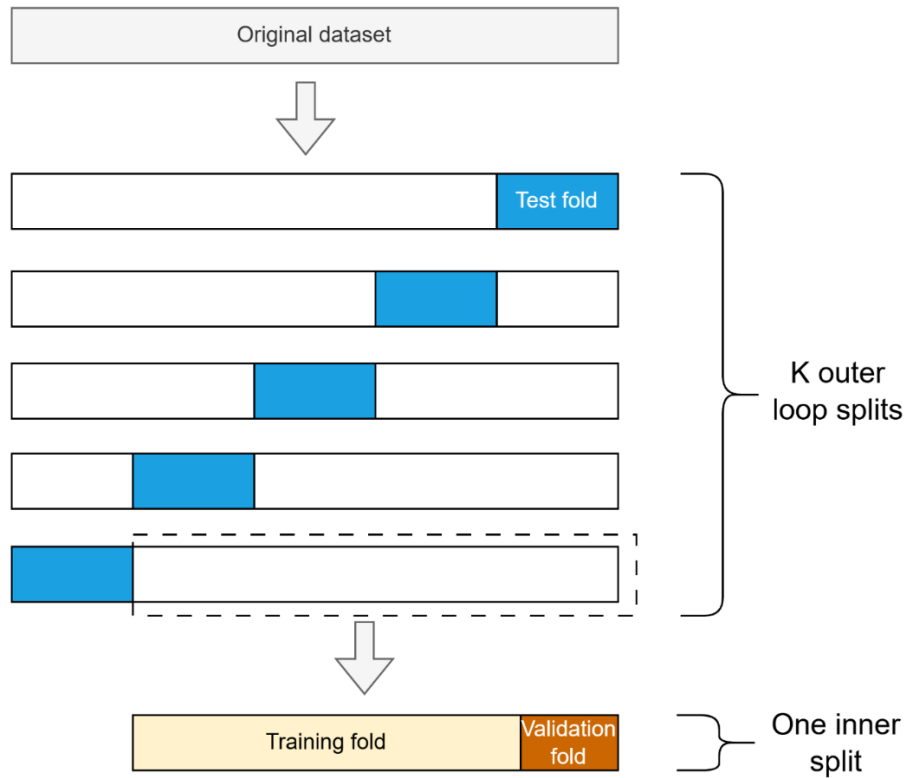
S(t) \ S(t+1)	Sleep	Sedentary	Light	MVPA
Sleep	0.999	0.001	0.000	0.000
Sedentary	0.001	0.986	0.012	0.000
Light	0.000	0.028	0.963	0.009
MVPA	0.000	0.003	0.025	0.972

Emission matrix:

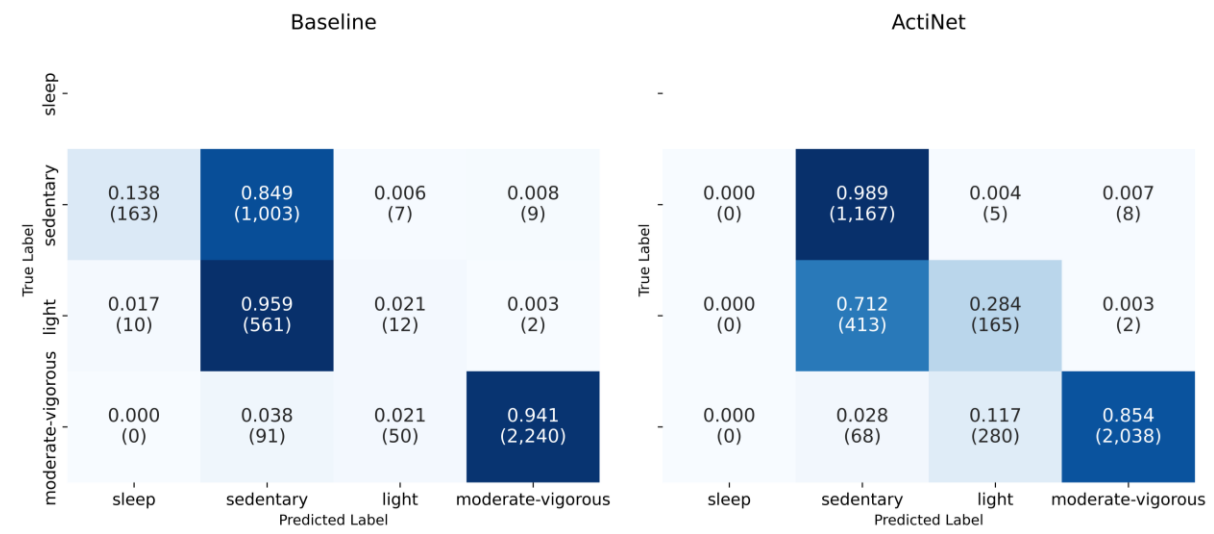
S(t) \ O(t)	Sleep	Sedentary	Light	MVPA
Sleep	0.861	0.126	0.010	0.003
Sedentary	0.085	0.700	0.153	0.062
Light	0.017	0.202	0.514	0.266
MVPA	0.003	0.036	0.166	0.795

Figures

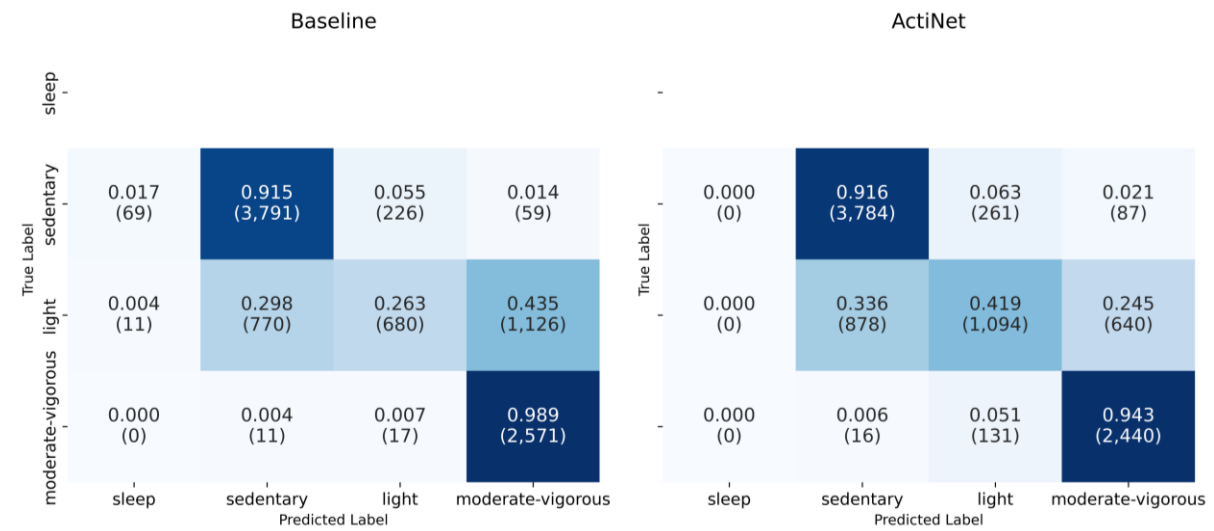
Supplementary Figure 1: Illustration of nested cross-validation, for the training and evaluation of the activity recognition models on the CAPTURE-24 dataset. K outer loop splits are a 5-fold stratified group cross validation. The one inner split is a group shuffle split with random state of 42.



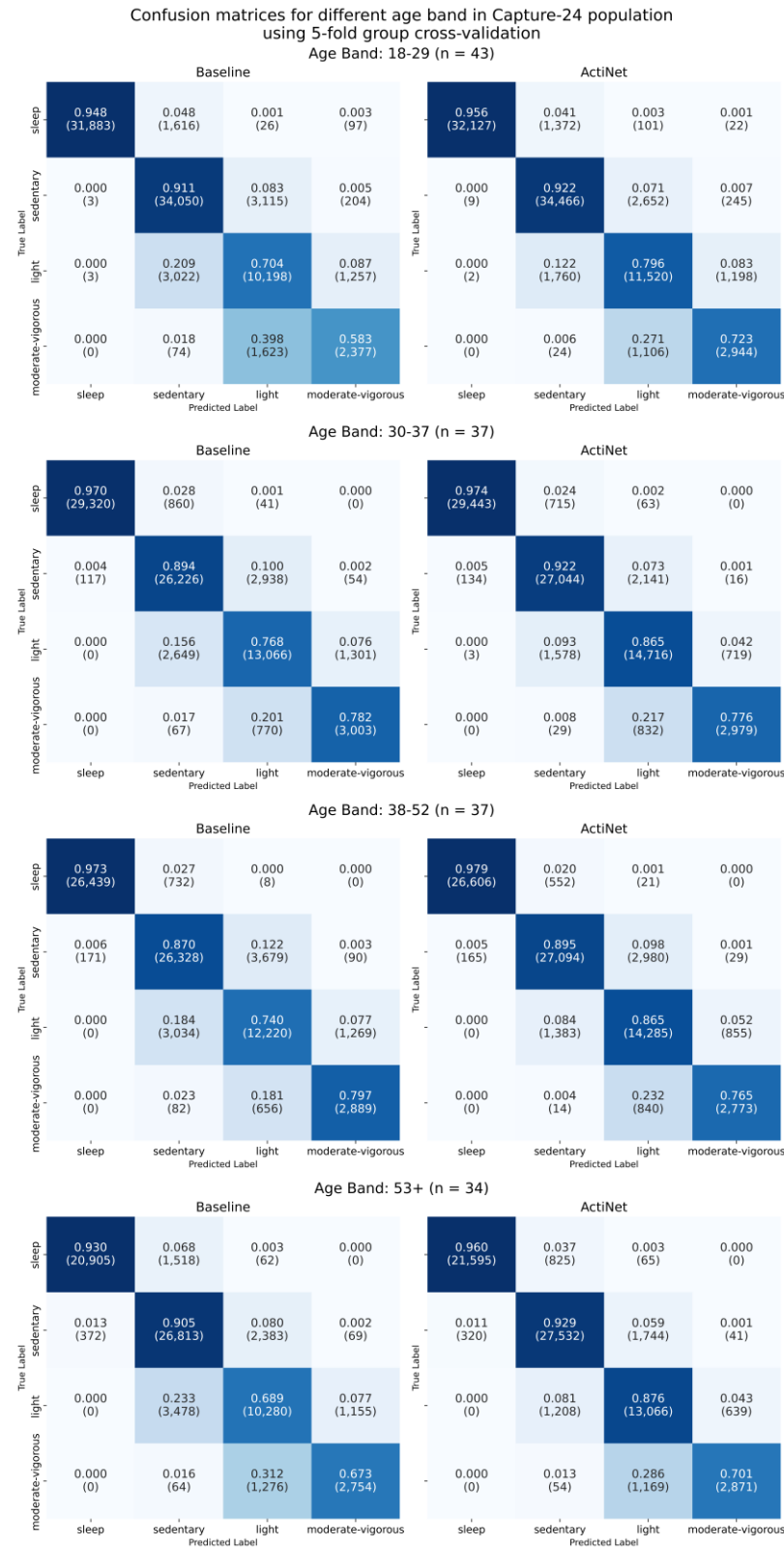
Supplementary Figure 2: Confusion matrices for the performance of CAPTURE-24 trained activity recognition models, tested on the Realworld dataset ($n = 14$). The baseline model is a random forest-based model, ActiNet is a self-supervised HARNet-based model. Each cell shows the proportion of 30-second windows with a given true activity intensity label that were predicted as each activity intensity label, with corresponding window counts in parentheses.



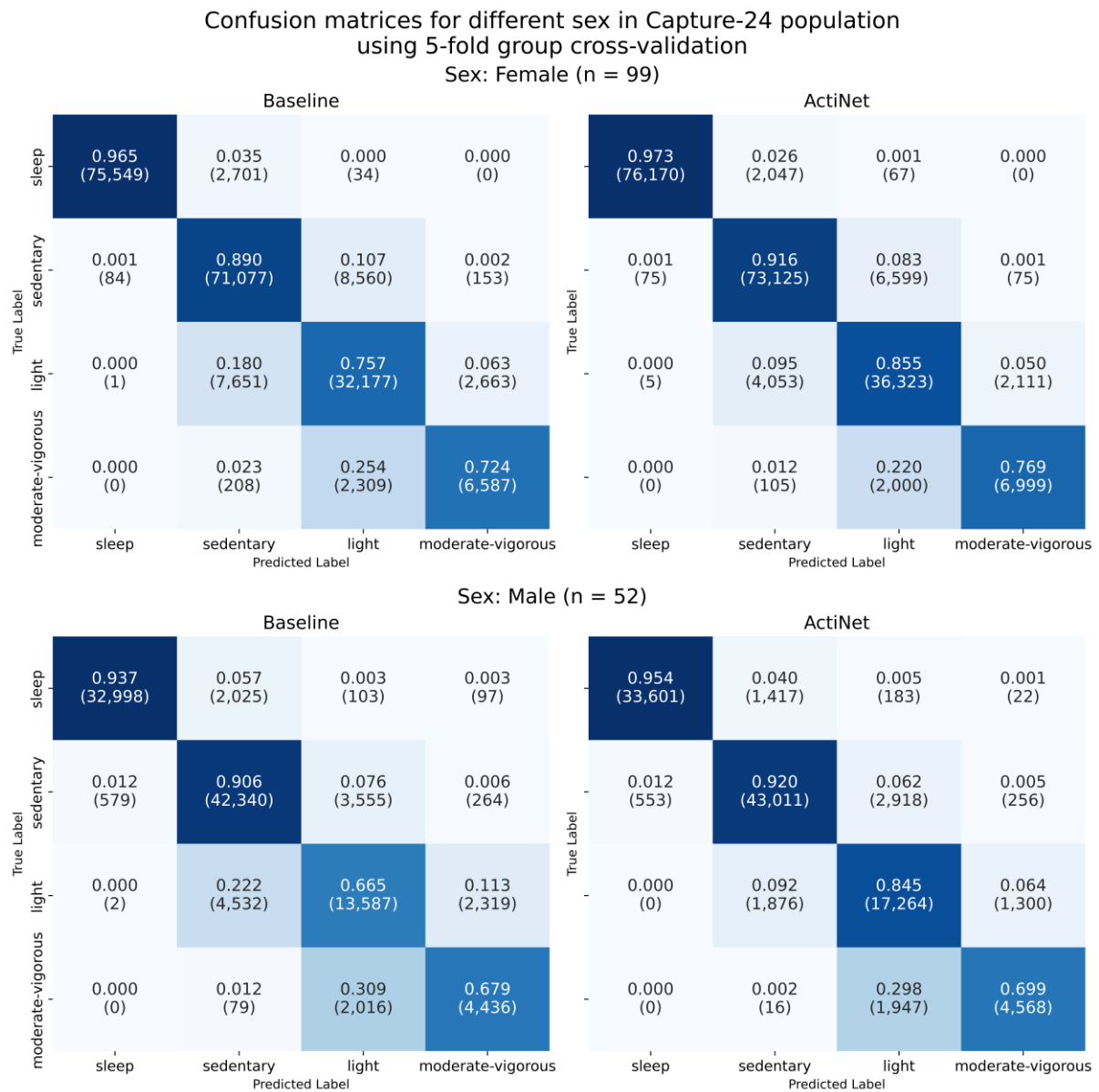
Supplementary Figure 3: Confusion matrices for the performance of CAPTURE-24 trained activity recognition models, tested on the WISDM dataset ($n = 46$). The baseline model is a random forest-based model, ActiNet is a self-supervised HARNet-based model. Each cell shows the proportion of 30-second windows with a given true activity intensity label that were predicted as each activity intensity label, with corresponding window counts in parentheses.



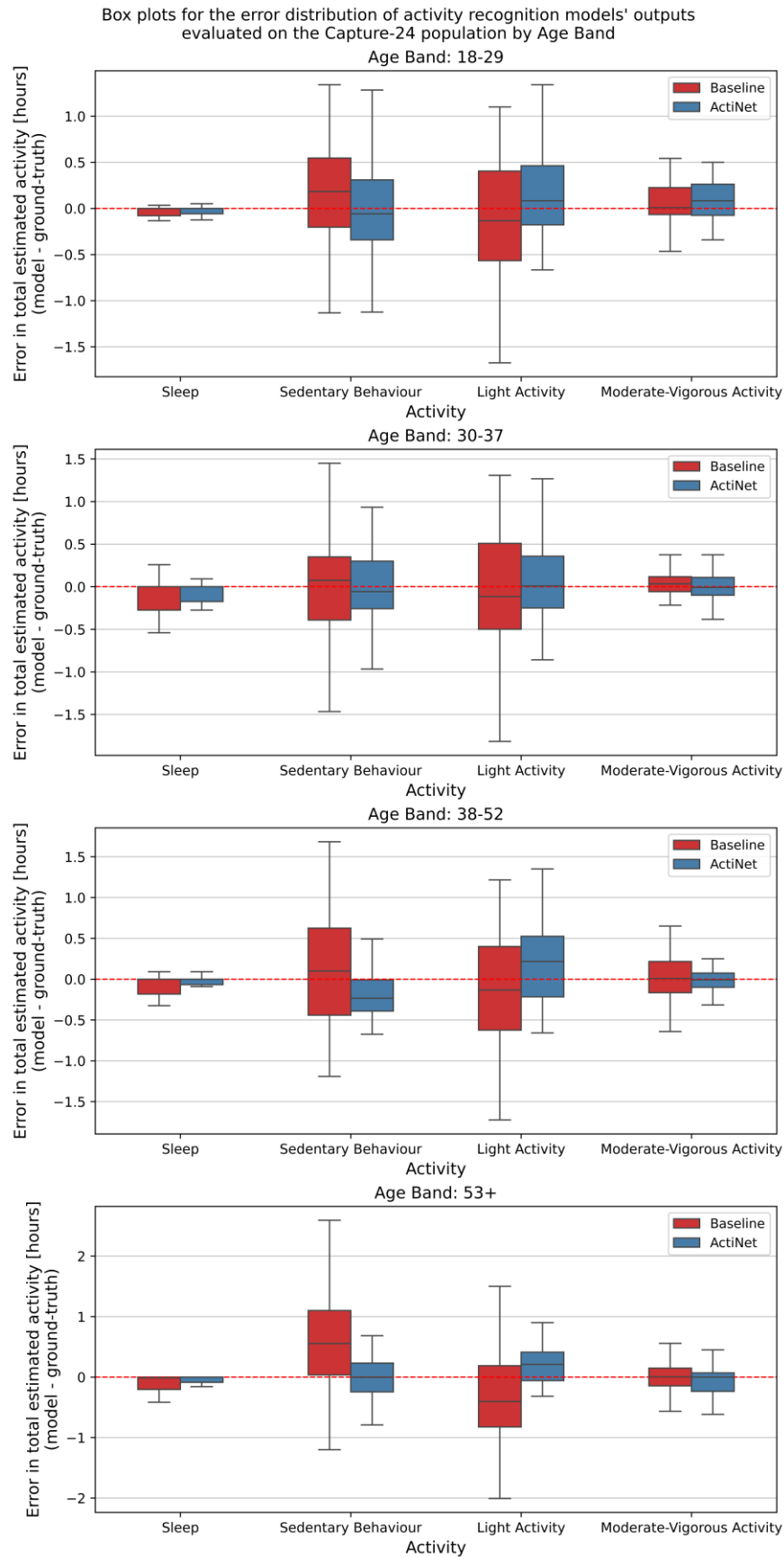
Supplementary Figure 4: Confusion matrices for activity intensity classifications made by baseline and ActiNet models using 5-fold cross validation on CAPTURE-24 dataset, evaluated in age bands: 18-29, 30-37, 38-52 and 53+. Each cell shows the proportion of 30-second windows with a given true activity intensity label that were predicted as each activity intensity label, with corresponding window counts in parentheses.



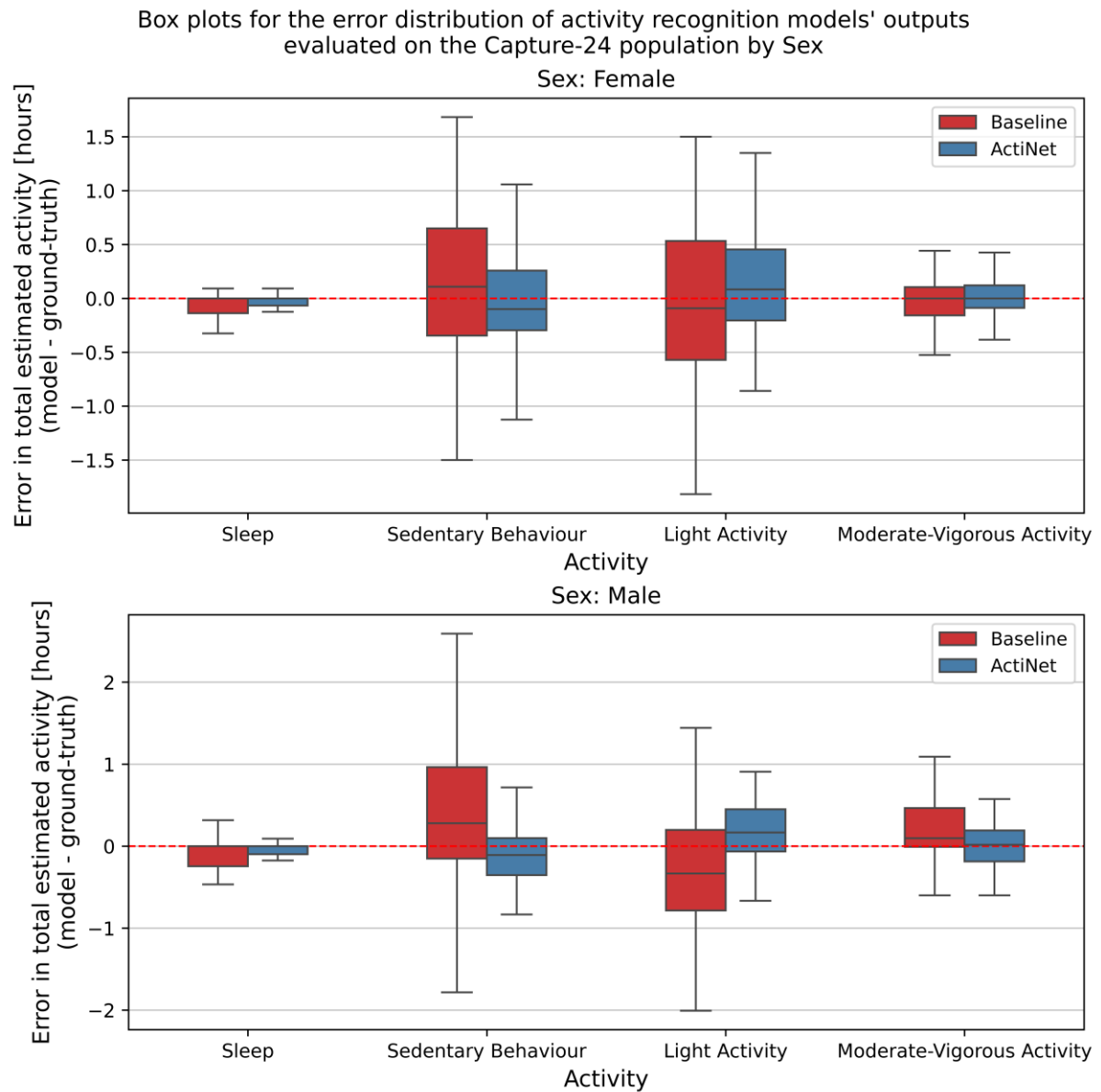
Supplementary Figure 5: Confusion matrices for activity intensity classifications made by baseline and ActiNet models using 5-fold cross validation on CAPTURE-24 dataset, evaluated for different sexes: female and male. Each cell shows the proportion of 30-second windows with a given true activity intensity label that were predicted as each activity intensity label, with corresponding window counts in parentheses.



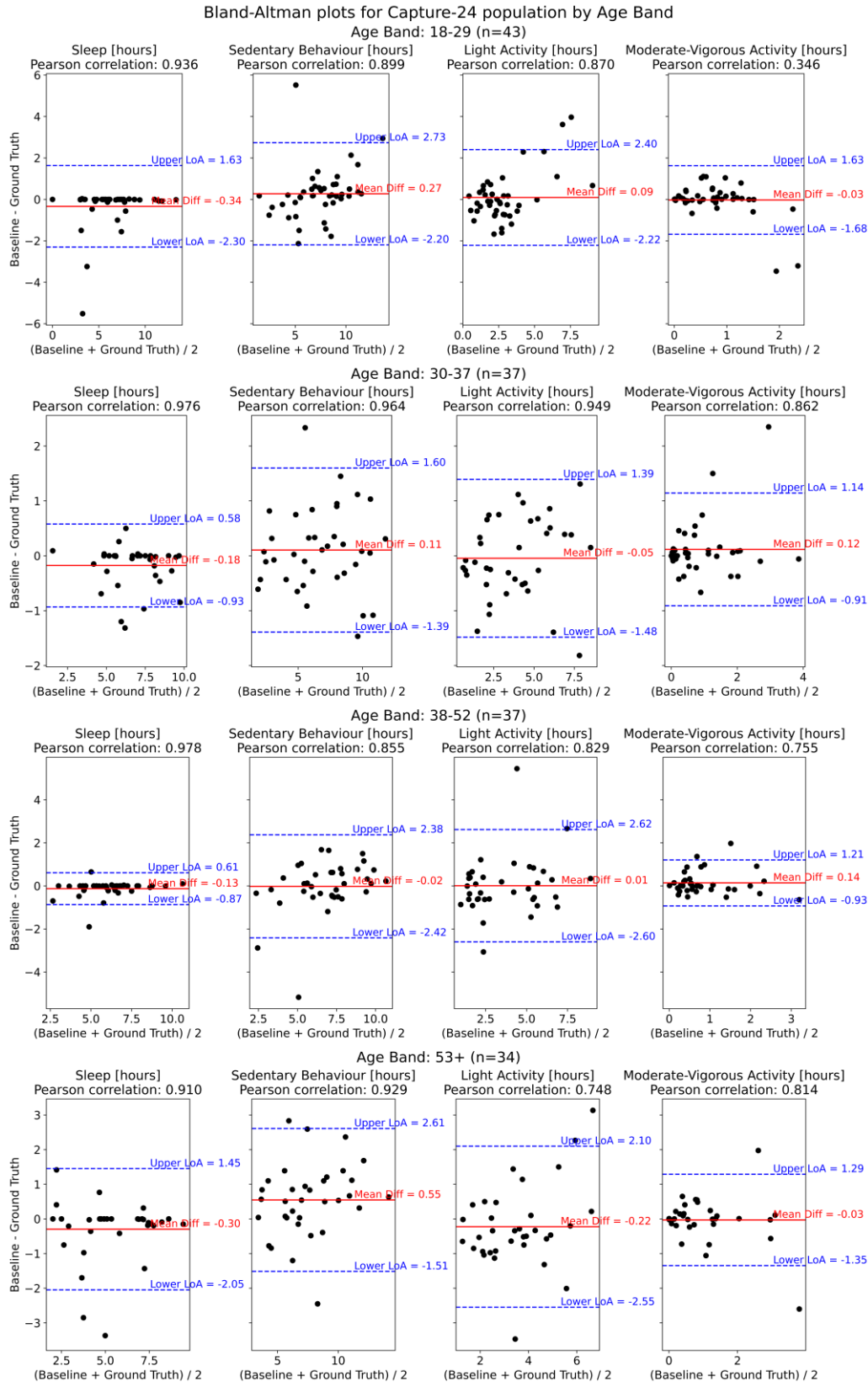
Supplementary Figure 6: Box plots for the distribution of errors in total activity intensity estimation by the baseline and ActiNet models, evaluated on different age bands of the CAPTURE-24 dataset using 5-fold group cross-validation.



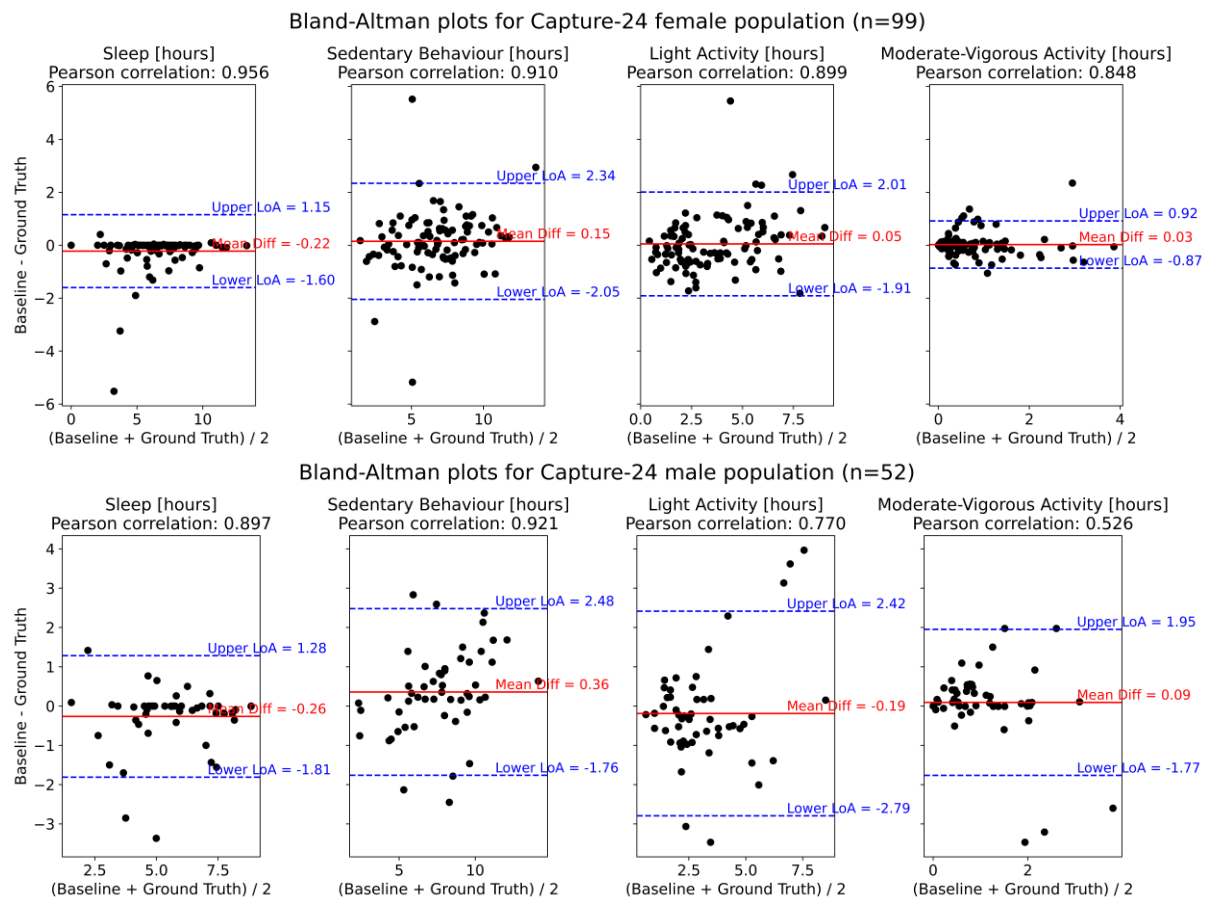
Supplementary Figure 7: Box plots for the error in total activity intensity estimation by the baseline and ActiNet models, evaluated on different sexes of the CAPTURE-24 dataset using 5-fold group cross-validation.



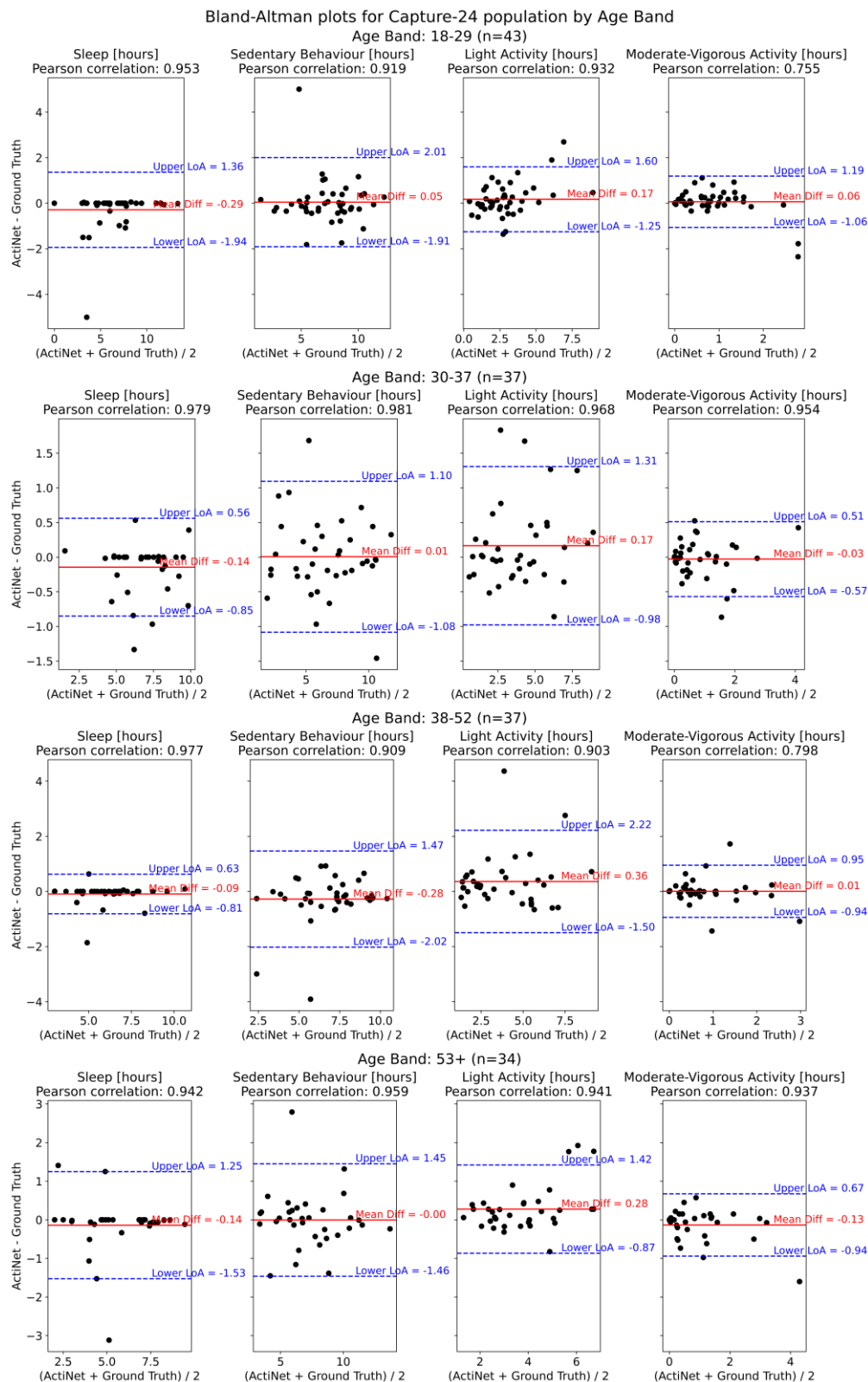
Supplementary Figure 8: Bland-Altman plots comparing the amount of activity intensity from the ground-truth annotations to the classifications by the baseline random forest-based model, evaluated on different age bands of the CAPTURE-24 dataset using 5-fold group cross-validation. Each Bland-Altman plot shows the difference between two activity intensity estimates against their mean, with the mean bias (solid red line) and 95% limits of agreement (dashed blue lines) indicating agreement.



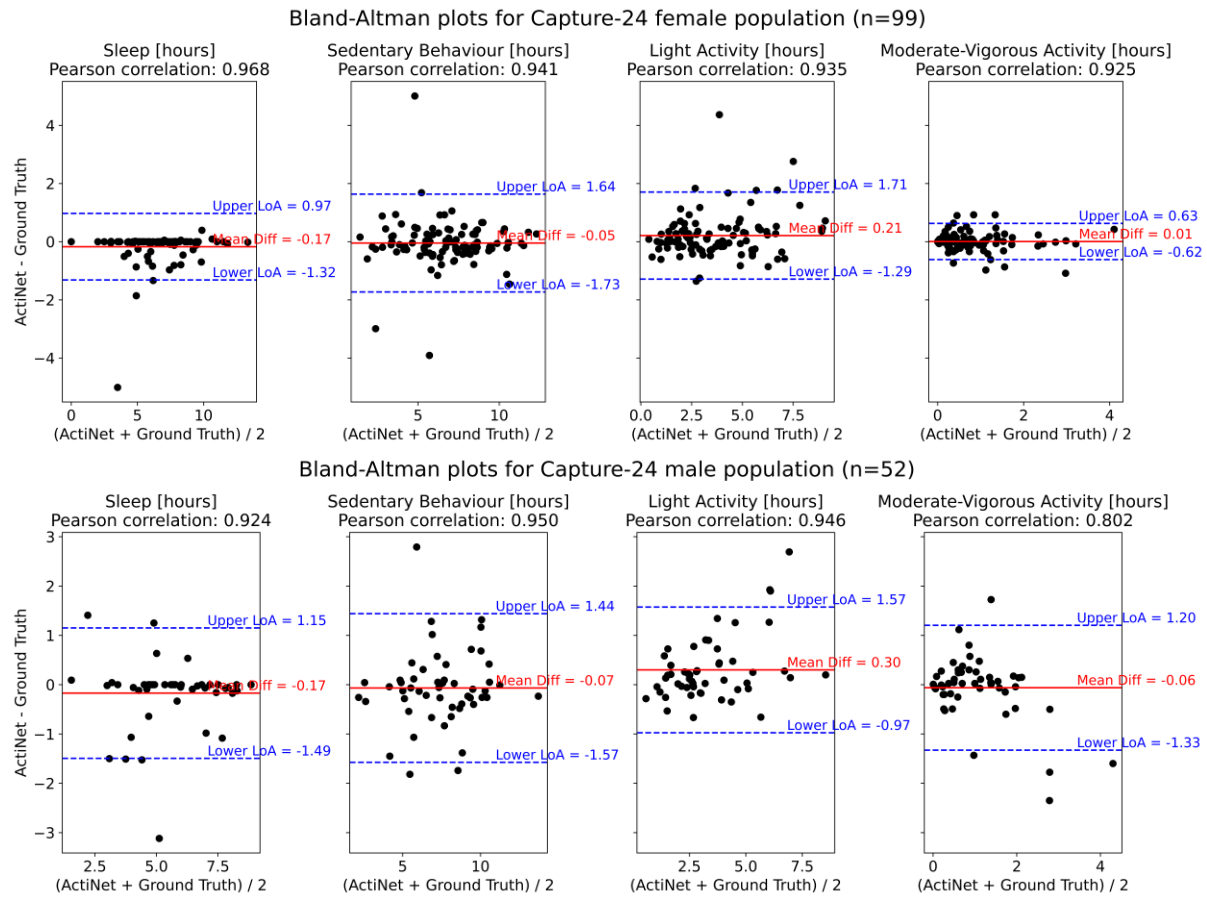
Supplementary Figure 9: Bland-Altman plots comparing the amount of activity intensity from the ground-truth annotations to the classifications by the baseline random forest-based model, evaluated on different sexes of the CAPTURE-24 dataset using 5-fold group cross-validation. Each Bland-Altman plot shows the difference between two activity intensity estimates against their mean, with the mean bias (solid red line) and 95% limits of agreement (dashed blue lines) indicating agreement.



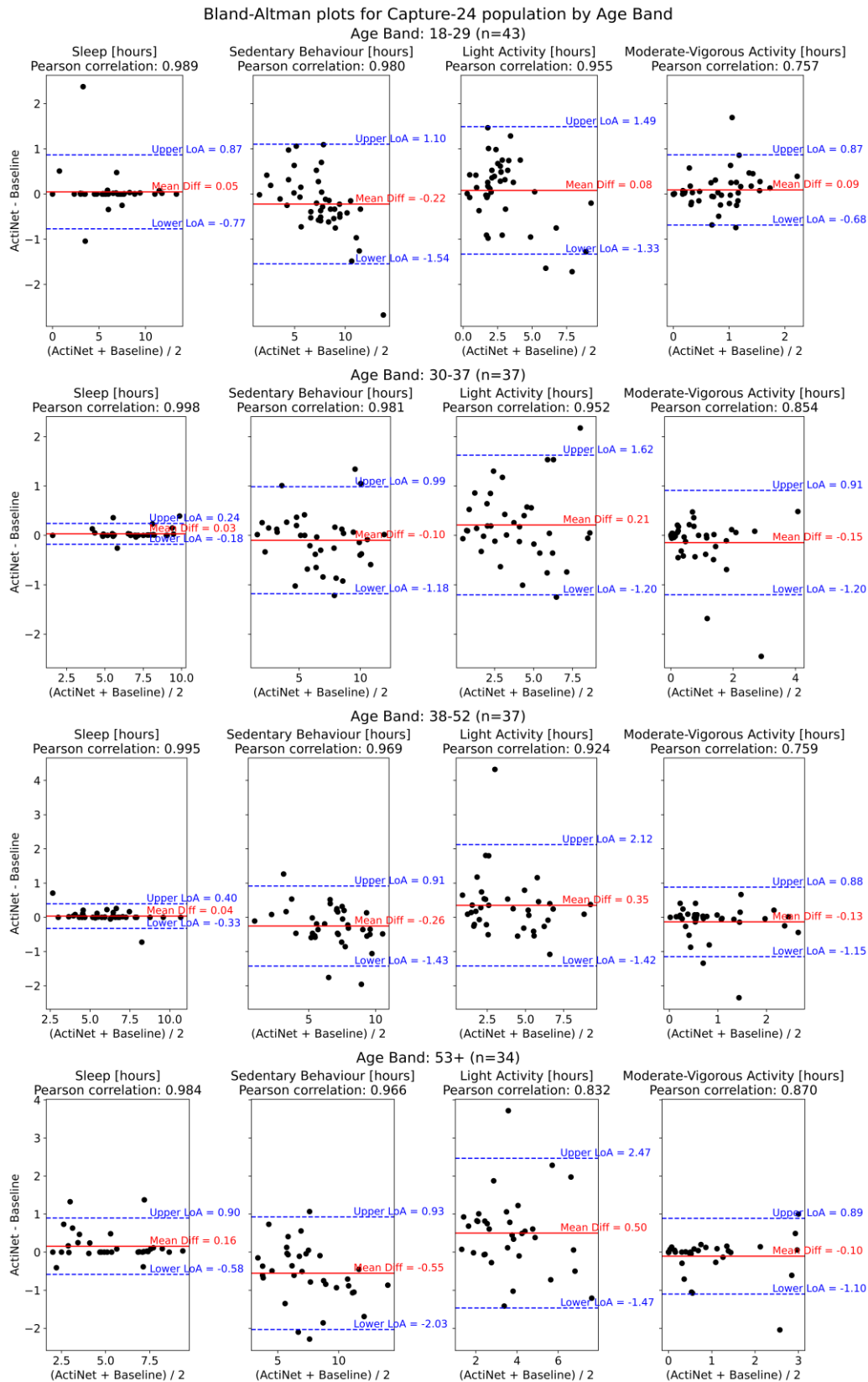
Supplementary Figure 10: Bland-Altman plots comparing the amount of activity intensity from the ground-truth annotations to the classifications by the ActiNet model, evaluated on different age bands of the CAPTURE-24 dataset using 5-fold group cross-validation. Each Bland-Altman plot shows the difference between two activity intensity estimates against their mean, with the mean bias (solid red line) and 95% limits of agreement (dashed blue lines) indicating agreement.



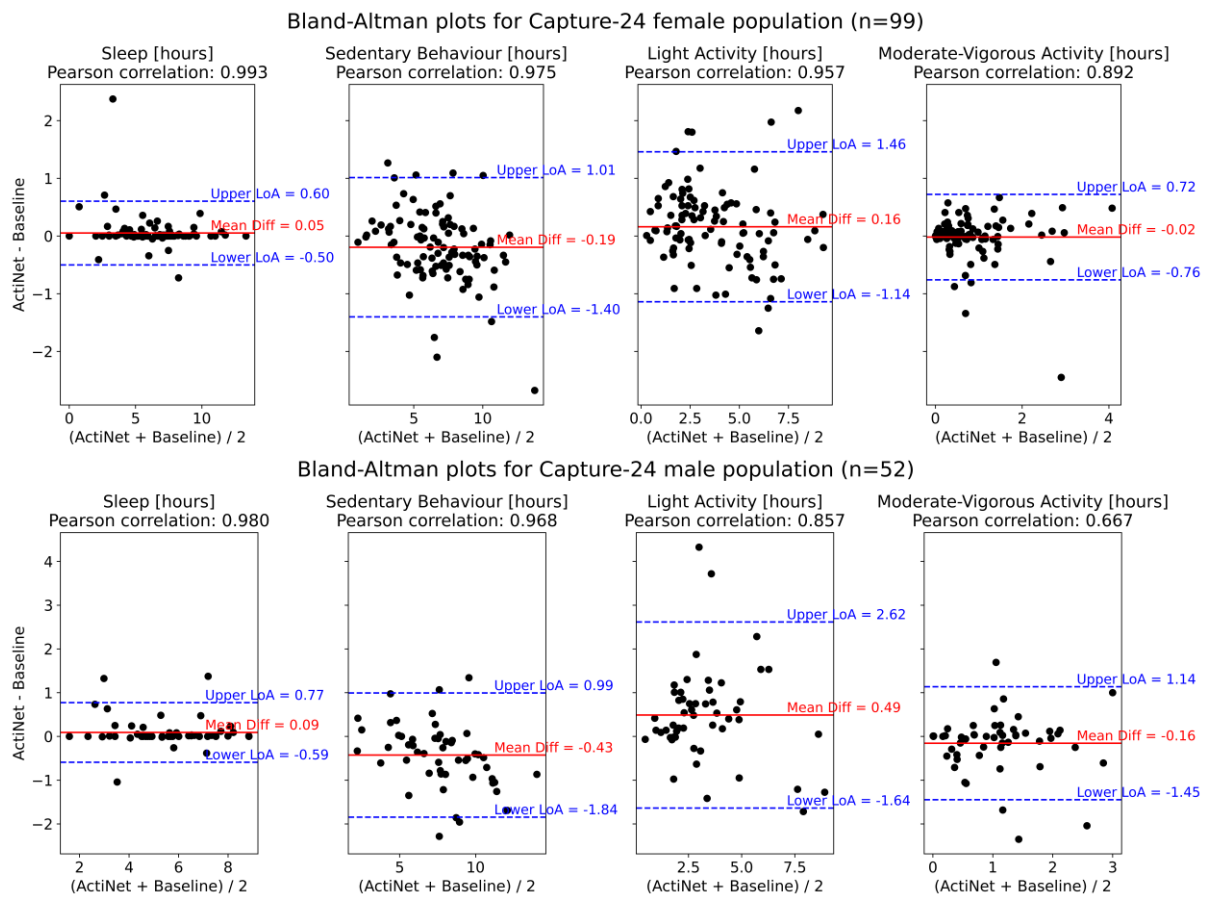
Supplementary Figure 11: Bland-Altman plots comparing the amount of activity intensity from the ground-truth annotations to the classifications by the ActiNet model, evaluated on different sexes of the CAPTURE-24 dataset using 5-fold group cross-validation. Each Bland-Altman plot shows the difference between two activity intensity estimates against their mean, with the mean bias (solid red line) and 95% limits of agreement (dashed blue lines) indicating agreement.



Supplementary Figure 12: Bland-Altman plots comparing the amount of activity intensity classified by the Baseline and ActiNet models, evaluated on different age bands of the CAPTURE-24 dataset using 5-fold group cross-validation. Each Bland-Altman plot shows the difference between two activity intensity estimates against their mean, with the mean bias (solid red line) and 95% limits of agreement (dashed blue lines) indicating agreement.



Supplementary Figure 13: Bland-Altman plots comparing the amount of activity intensity classified by Baseline and ActiNet models, evaluated on difference sexes of the CAPTURE-24 dataset using 5-fold group cross-validation. Each Bland-Altman plot shows the difference between two activity intensity estimates against their mean, with the mean bias (solid red line) and 95% limits of agreement (dashed blue lines) indicating agreement.



References

1. Doherty, A., Chan, S., Yuan, H. & Walmsley, R. accelerometer: A Python Toolkit for Extracting Physical Activity and Behavior Metrics from Wearable Sensor Data.
<https://doi.org/10.5281/zenodo.14515076> (2020).
2. Breiman, L. Random Forests. *Machine Learning* **45**, 5–32 (2001).
3. van Hees, V. T. *et al.* Separating Movement and Gravity Components in an Acceleration Signal and Implications for the Assessment of Human Daily Physical Activity. *PLoS ONE* **8**, e61691 (2013).
4. Chang, S. C. *et al.* Capture-24: Activity tracker dataset for human activity recognition.
<https://ora.ox.ac.uk/objects/uuid:99d7c092-d865-4a19-b096-cc16440cd001> (2021).