Supporting Information for Multimodal wearable sensing reveals personalized digital signatures of momentary happiness and stress

Table 1 Participant-specific Pearson correlation summaries for pre-15 min CGM and affect.

Analysis	Happiness Median r	Stress Median r
Raw CGM vs. actual affect	-0.03	0.04
ΔCGM vs. actual affect	-0.04	0.07
Raw CGM vs. Δaffect	-0.03	0.07
ΔCGM vs. Δaffect	-0.04	0.07

Table 2 EMA model-family summary for participant-specific held-out performance. Median RMSE, MAE, held-out R^2, and ΔRMSE are summarized across eligible participant-specific analyses. Positive ΔRMSE indicates lower held-out error than the participant-specific time-bin baseline.

Outcome	Model	Feature set	Participants	Median RMSE	Median MAE	Median R2	Median dRMSE	Improved (%)
happiness	Linear ridge	Pre	38	15.509	11.902	-0.247	-0.574	23.684
happiness	Spline ridge	Pre	37	14.283	11.275	-0.174	-0.286	29.730
happiness	Elastic net	Pre	37	14.463	11.992	-0.149	-0.304	24.324
happiness	Linear ridge	Post	38	14.093	11.494	-0.143	-0.127	39.474
happiness	Spline ridge	Post	37	14.827	11.163	-0.150	-0.428	32.432
happiness	Elastic net	Post	37	14.093	10.916	-0.116	-0.051	35.135
happiness	Linear ridge	Pre+post	38	14.508	11.036	-0.161	-0.175	36.842
happiness	Spline ridge	Pre+post	37	14.772	11.498	-0.140	-0.158	35.135
happiness	Elastic net	Pre+post	37	14.276	10.869	-0.123	-0.114	24.324
happiness	Linear ridge	Pre+context	38	16.401	12.529	-0.336	-0.992	18.421
happiness	Spline ridge	Pre+context	37	15.284	11.868	-0.282	-0.946	16.216
happiness	Elastic net	Pre+context	37	15.984	11.930	-0.199	-0.321	27.027
happiness	Random forest	Pre+context	37	13.609	10.692	-0.126	-0.149	43.243
happiness	SVR-RBF	Pre+context	37	14.213	10.768	-0.125	-0.066	45.946
stress	Linear ridge	Pre	38	18.808	13.785	-0.173	-0.334	23.684
stress	Spline ridge	Pre	37	18.432	14.209	-0.184	-0.443	29.730
stress	Elastic net	Pre	37	18.307	13.938	-0.151	-0.152	10.811
stress	Linear ridge	Post	38	18.296	12.918	-0.159	-0.326	23.684
stress	Spline ridge	Post	37	17.645	13.766	-0.150	-0.238	37.838
stress	Elastic net	Post	37	17.437	12.960	-0.117	-0.098	24.324
stress	Linear ridge	Pre+post	38	18.103	14.162	-0.150	-0.261	18.421
stress	Spline ridge	Pre+post	37	18.838	14.158	-0.184	-0.329	27.027
stress	Elastic net	Pre+post	37	18.392	13.917	-0.142	-0.156	10.811
stress	Linear ridge	Pre+context	38	19.319	14.451	-0.243	-0.464	26.316
stress	Spline ridge	Pre+context	37	19.978	14.721	-0.222	-0.732	24.324
stress	Elastic net	Pre+context	37	18.106	14.069	-0.159	-0.162	18.919
stress	Random forest	Pre+context	37	17.681	13.277	-0.136	-0.149	40.541
stress	SVR-RBF	Pre+context	37	17.269	12.782	-0.136	-0.038	43.243

Table 3 SHAP feature-class frequency for EMA random-forest models using pre+context features. Percentages indicate the fraction of top-10 SHAP-ranked feature-class appearances assigned to each physiological or behavioral class. These summaries describe model feature usage and are not interpreted as causal feature effects.

dataset	outcome	feature_class	top10_count	total	top10_percent
ema	happiness	Activity	55	370	14.865
ema	happiness	CGM	146	370	39.459
ema	happiness	HR/autonomic	143	370	38.649
ema	happiness	Sleep	26	370	7.027
ema	stress	Activity	72	370	19.459
ema	stress	CGM	138	370	37.297
ema	stress	HR/autonomic	129	370	34.865
ema	stress	Sleep	31	370	8.378

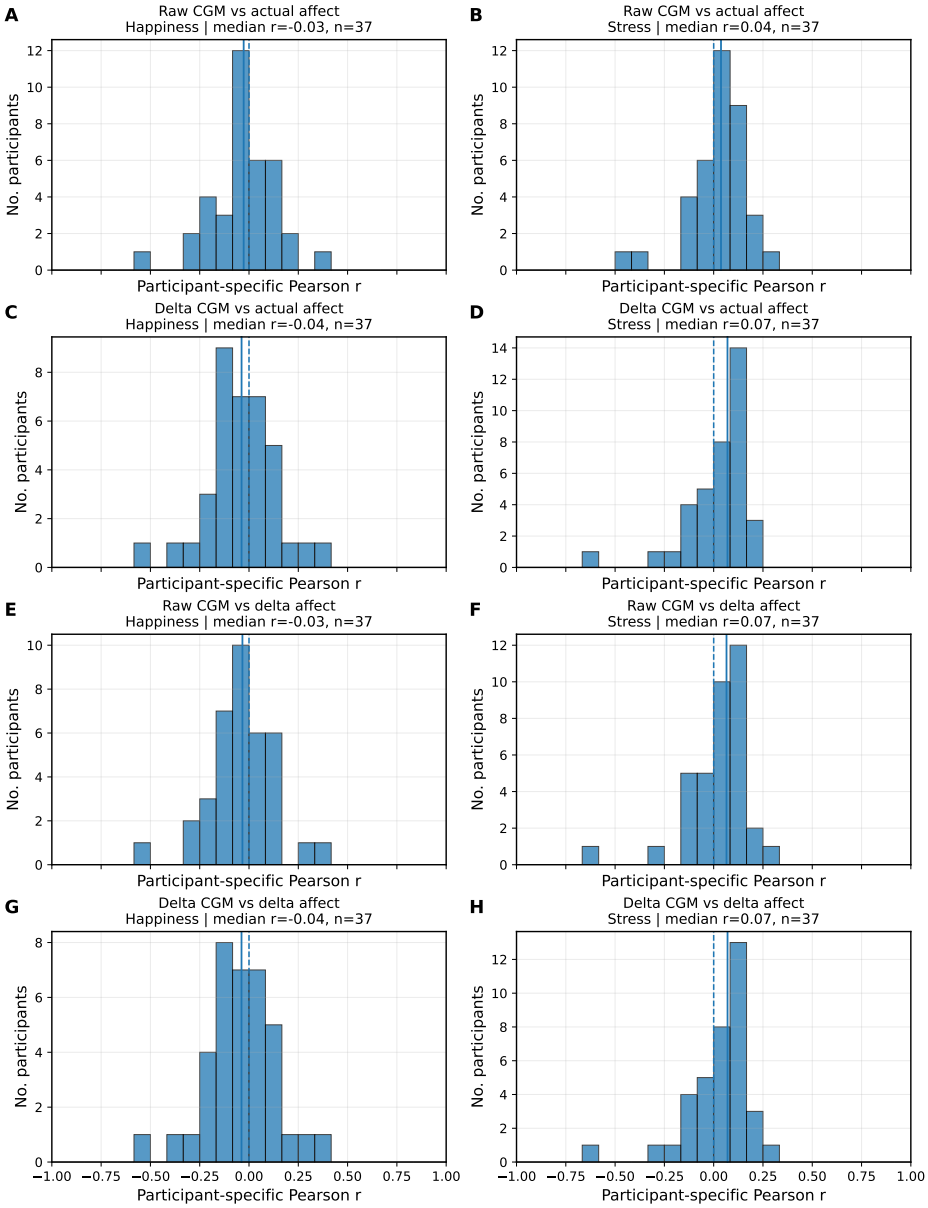

Fig. 1 Participant-specific Pearson correlations between pre-15 min CGM and affect under alternative baseline-adjustment schemes. Histograms show the distribution of participant-specific Pearson correlation coefficients computed using four formulations: raw CGM versus raw affect (A,B), participant-centered (Δ) CGM versus raw affect (C,D), raw CGM versus participant-centered affect (E,F), and participant-centered CGM versus participant-centered affect (G,H). Happiness correlations remained centered near zero across all formulations, whereas stress correlations showed a small positive shift. Median correlation coefficients are indicated in each panel. These analyses demonstrate that baseline adjustment of either CGM or affect variables does not materially alter the observed association structure.

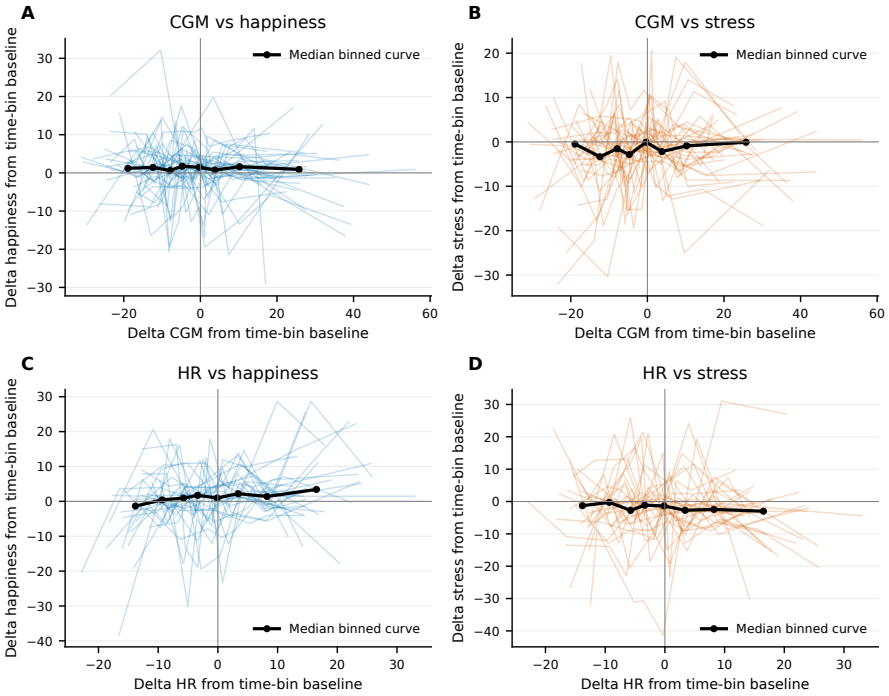

Fig. 2 Within-person binned association curves for CGM, HR, happiness, and stress. **A,** Within-person deviations in CGM from time-bin baseline versus deviations in happiness from time-bin baseline. **B,** Within-person deviations in CGM versus deviations in stress. **C,** Within-person deviations in HR versus deviations in happiness. **D,** Within-person deviations in HR versus deviations in stress. Thin lines show participant-specific binned curves, and black lines show the median binned curve across participants. The absence of a strong monotonic or threshold-like median curve indicates that weak linear sensor–affect associations were not explained by an obvious missed nonlinear relationship.

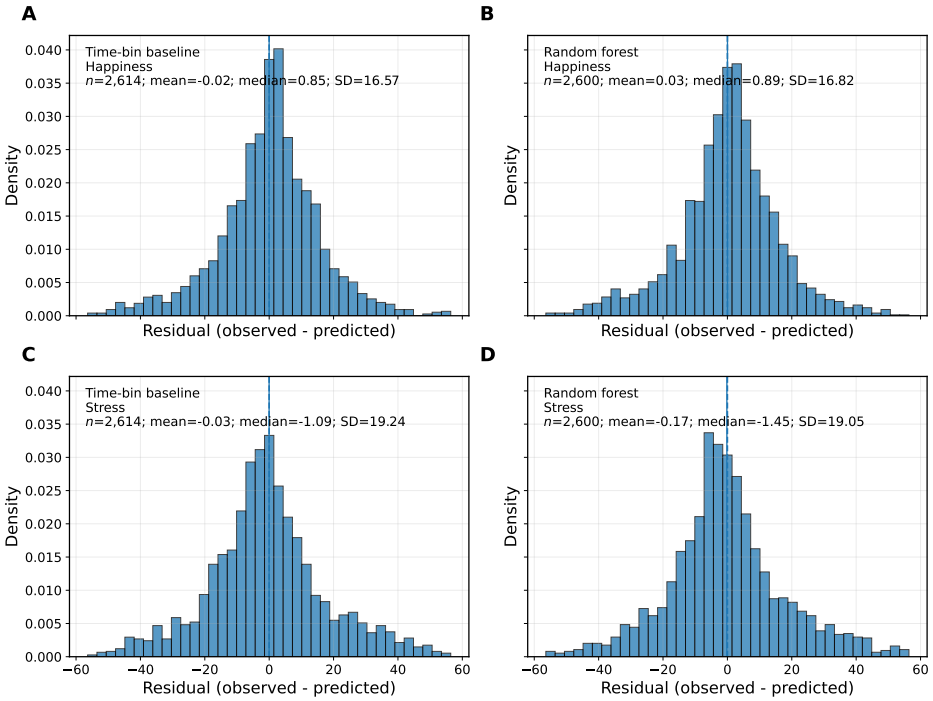

Fig. 3 Pooled residual distributions for held-out affect predictions. A, Residual distribution for time-bin baseline happiness predictions. **B,** Residual distribution for random-forest happiness predictions using pre-context features. **C,** Residual distribution for time-bin baseline stress predictions. **D,** Residual distribution for random-forest stress predictions using pre-context features. Residuals were defined as observed minus predicted affect score. Vertical dashed lines indicate zero residual. Both baseline and random-forest models showed residuals centered near zero but with broad dispersion, indicating limited reduction in momentary prediction error after adding sensor and context features.

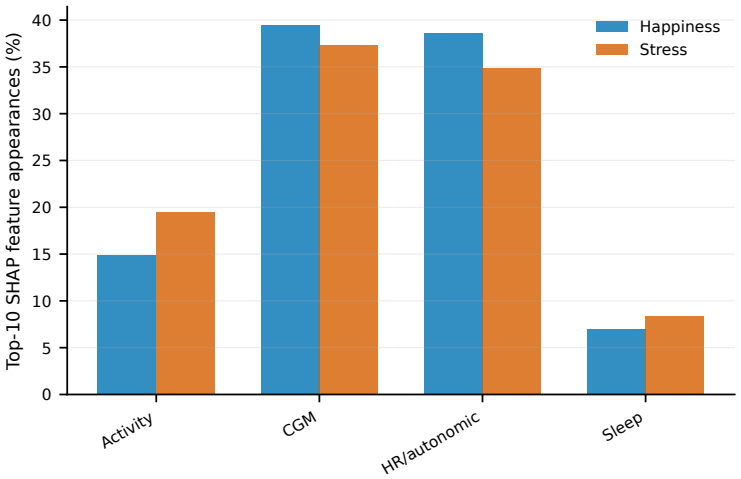

Fig. 4 Grouped SHAP feature-class frequency across personalized random-forest models. Top-10 SHAP-ranked features were grouped into four broader classes: CGM, HR/autonomic, activity, and sleep. Bars show the percentage of top-10 SHAP feature appearances assigned to each class for happiness and stress models. The distribution across feature classes indicates that model explanations were not dominated by glucose alone, but reflected multimodal metabolic, autonomic, activity, and sleep-related context.

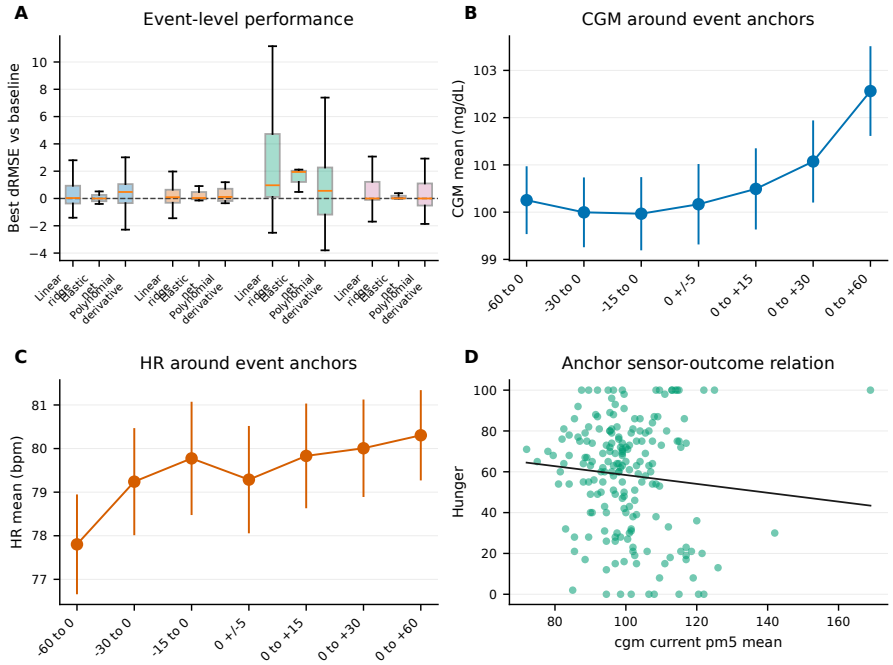

Fig. 5 Fasting event-context analysis. A, Event-level best-model ΔRMSE relative to the time-bin baseline for fasting survey outcomes across linear ridge, elastic-net, and polynomial derivative models. **B,** Mean CGM values across windows before, during, and after fasting event anchors. **C,** Mean HR values across windows before, during, and after fasting event anchors. **D,** Anchor-level relationship between current CGM mean and hunger. These analyses were evaluated using leave-one-anchor-out validation and were interpreted as exploratory event-context analyses rather than day-level prospective prediction.

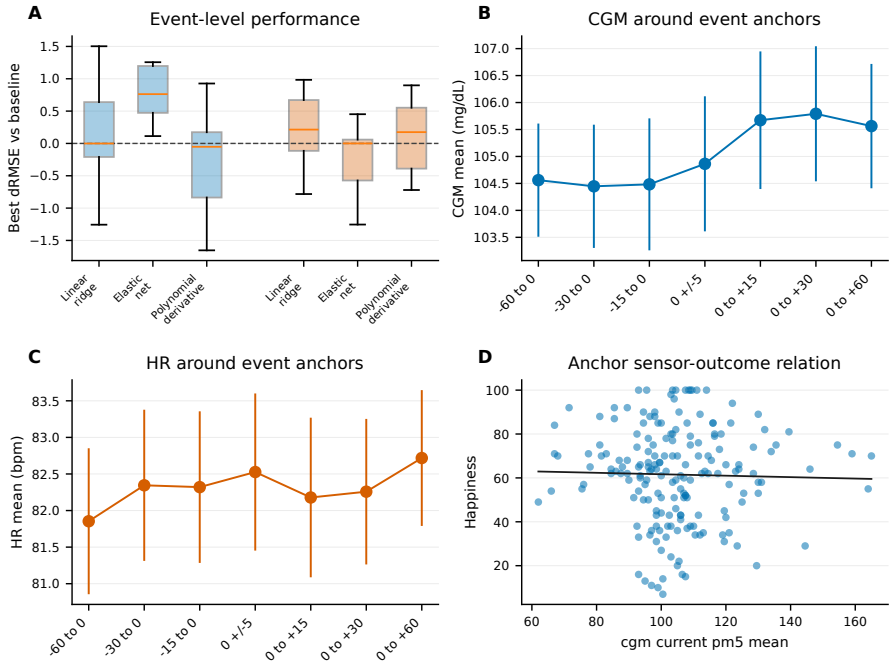

Fig. 6 Post-meal event-context analysis. A, Event-level best-model ΔRMSE relative to the time-bin baseline for post-meal affective outcomes across linear ridge, elastic-net, and polynomial derivative models. **B,** Mean CGM values across windows before, during, and after post-meal event anchors. **C,** Mean HR values across windows before, during, and after post-meal event anchors. **D,** Anchor-level relationship between current CGM mean and happiness. These analyses were exploratory and were used to assess whether metabolically constrained event contexts showed stronger sensor–outcome structure than the primary EMA affect analyses.

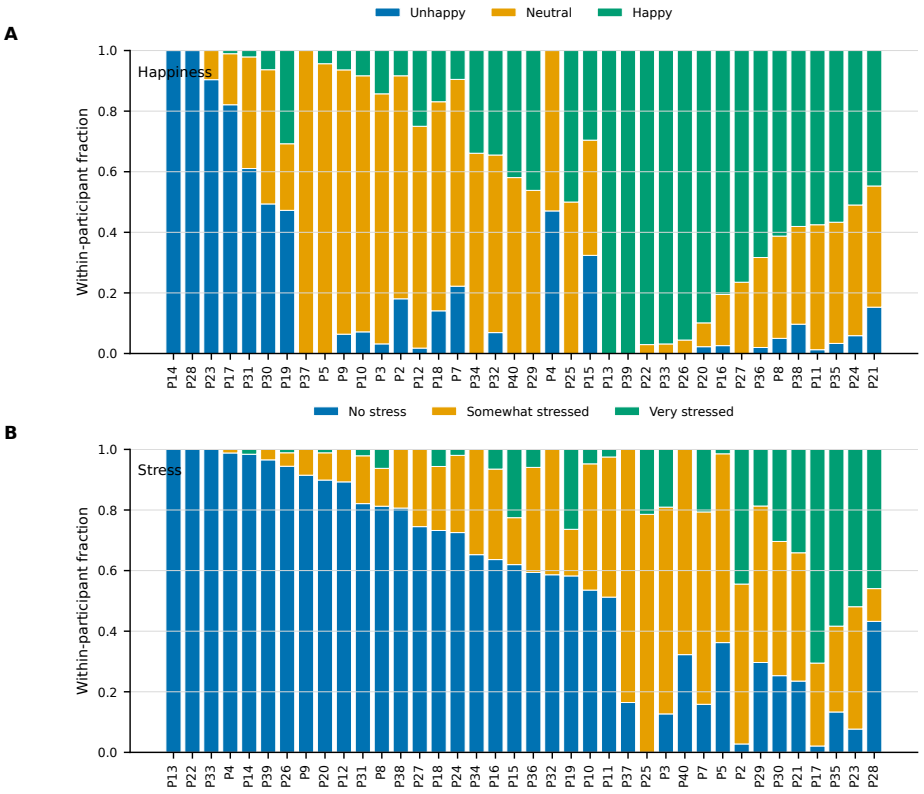

Fig. 7 Participant-level distribution of three-class ordinal affect states. Stacked bars show the fraction of each participant's EMA observations assigned to each three-class ordinal state for happiness and stress. The figure illustrates substantial heterogeneity in participant-level class balance, including participants dominated by a single affective state and others spanning all three ordinal classes. This heterogeneity motivates within-participant validation and comparison against personalized baselines rather than pooled population-level classification alone.

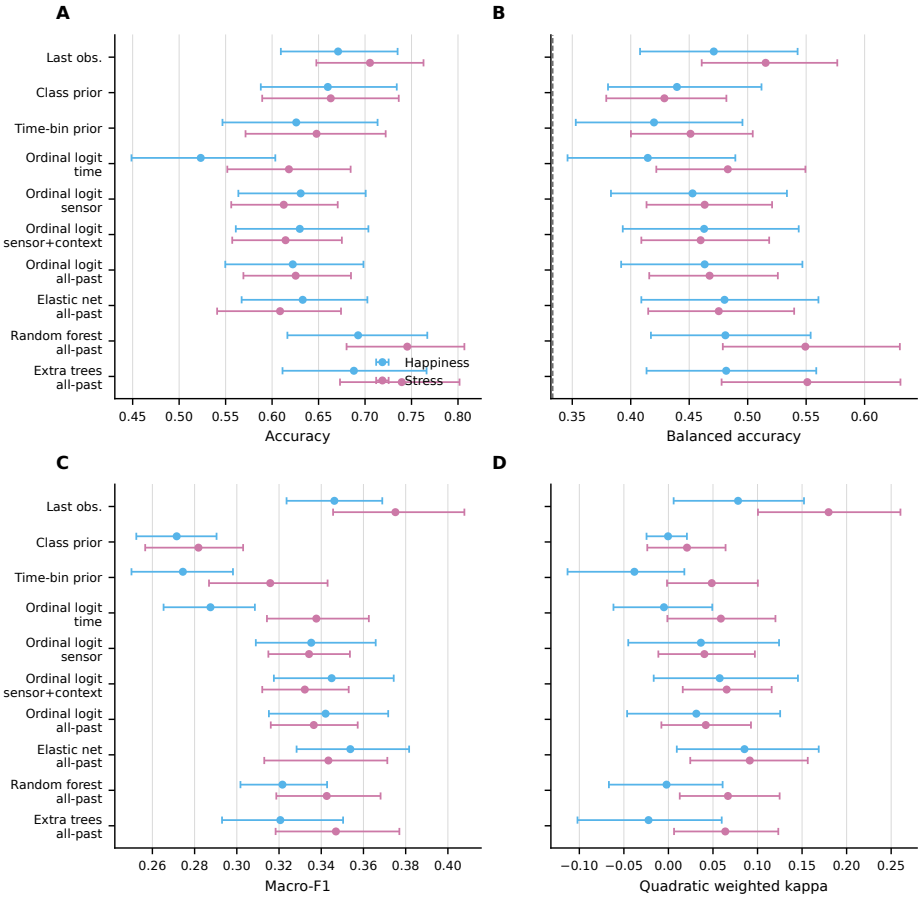

Fig. 8 Full model-level summary of three-class ordinal classification metrics. Summary metrics are shown across baseline and model families for happiness and stress. Metrics include accuracy, balanced accuracy, macro-F1, and quadratic weighted kappa. Accuracy and within-one-class performance were generally higher than macro-F1 and kappa because class distributions were imbalanced and many prediction errors occurred between adjacent ordinal states. These summaries show that improvement over personalized baselines depends on the metric used, with stress showing clearer gains in balanced accuracy and ordinal error than happiness.

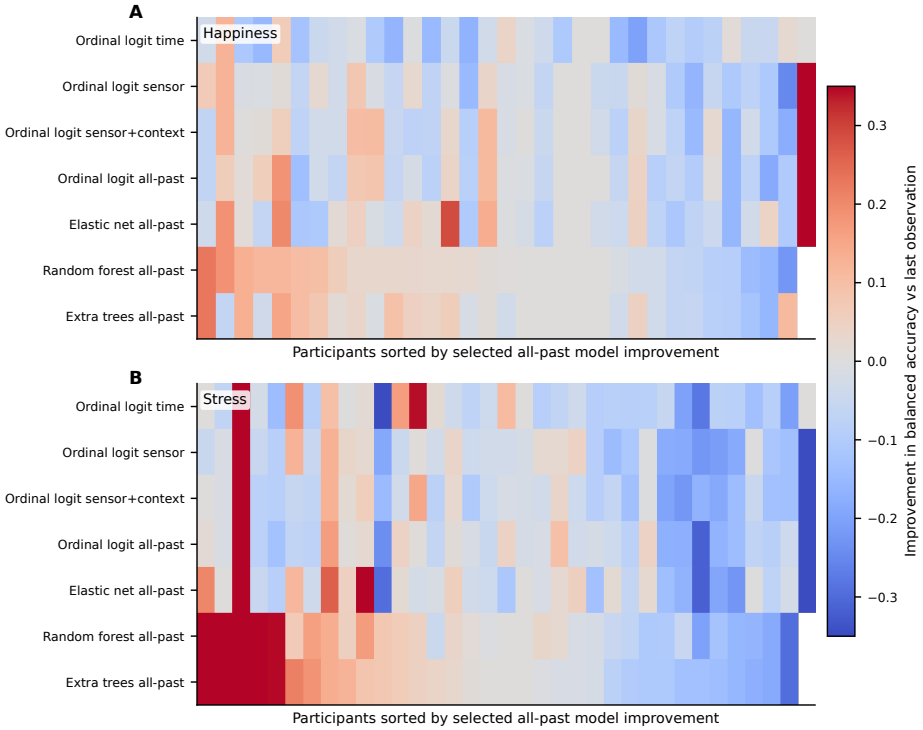

Fig. 9 Participant-level improvement in balanced accuracy across ordinal model families. Heatmap values show the change in balanced accuracy for each participant, outcome, and model family relative to the last-observation baseline. Positive values indicate improvement over the participant's most recent prior affective state, whereas negative values indicate worse performance. Improvements were heterogeneous across participants, consistent with individualized temporal and physiological coupling rather than a uniform sensor–affect relationship across the cohort.

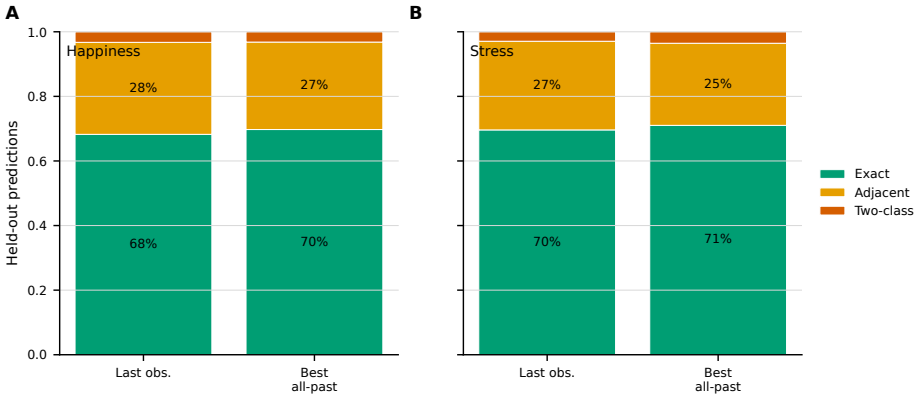

Fig. 10 Exact, adjacent, and far-error structure in three-class ordinal prediction. Classification errors were decomposed into exact predictions, adjacent one-class errors, and far two-class errors. Because the target is ordinal, adjacent errors represent less severe misclassification than errors between opposite affective states. Across outcomes and models, most incorrect predictions were adjacent-class errors, supporting the interpretation that the classifiers often tracked broad affective state even when exact class assignment was imperfect.

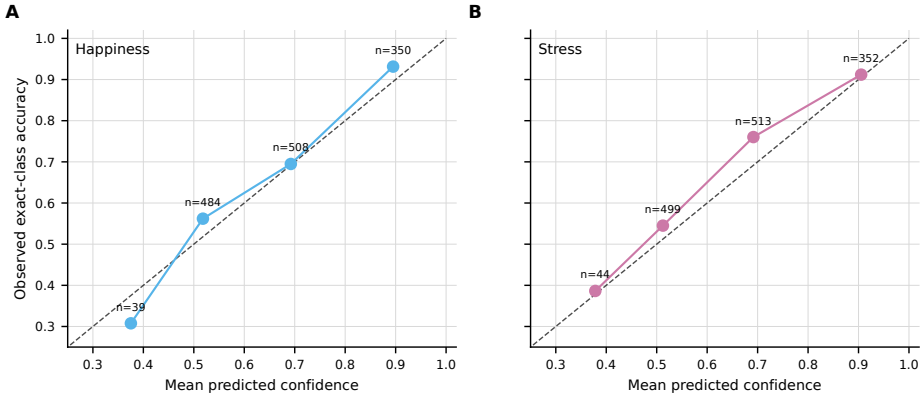

Fig. 11 Probability-confidence diagnostics for three-class ordinal classifiers. Predicted class probabilities were summarized to assess whether model confidence was associated with correct classification. These diagnostics evaluate whether high-confidence predictions were more often correct and whether probability estimates were concentrated or diffuse across classes. The results provide a calibration-oriented supplement to accuracy-based metrics and help distinguish confident state tracking from uncertain class assignment.

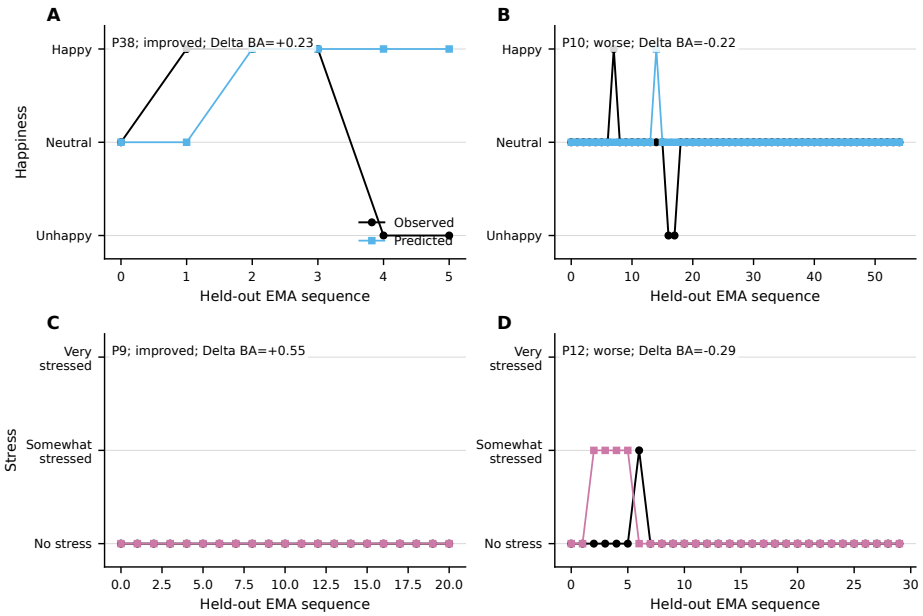

Fig. 12 Representative participant-level ordinal prediction trajectories. Example time series compare observed three-class ordinal affect states with predicted classes from the selected all-past model. These participant-level trajectories illustrate the temporal structure underlying aggregate performance metrics, including periods of correct state tracking, adjacent-class underestimation or overestimation, and missed transitions. The examples emphasize that ordinal model performance is participant-dependent and should be interpreted as personalized state tracking rather than as a universal affect classifier.