Additional file 2

Supplementary tables and figure for: *Subject-aware deep learning for forecasting a causal Mental Health Index from passive smartphone sensing in university students: a prediction model development and temporal validation study*. All tables are generated directly from the result files of the analysis code.

Table S1. Hourly input channels (19 continuous, 7 availability flags, 3 time encodings). Each forecast sample is the 72×29 matrix of these channels in the 72 hours before 04:00 on the target day. Continuous channels were standardised with training statistics (I0) or normalised per participant against that participant's previous hours (I1–I4); flags and time encodings were not rescaled. All imputation uses only earlier hours. Permutation importance (Table S12) permutes the groups shown here, with phone lock and screen inactivity combined, and applications and calls combined.

Group	Channel	Description	Processing	Missing hours	Type
Activity	`act_stationary`	Proportion of activity-inference samples in the hour labelled stationary	count in category / all inference samples in the hour	forward fill $\leq$3 h, then mean of the same hour on previous days	continuous
	`act_walking`	Proportion labelled walking	count in category / all inference samples in the hour	forward fill $\leq$3 h, then mean of the same hour on previous days	continuous
	`act_running`	Proportion labelled running	count in category / all inference samples in the hour	forward fill $\leq$3 h, then mean of the same hour on previous days	continuous
	`act_unknown`	Proportion labelled unknown	count in category / all inference samples in the hour	forward fill $\leq$3 h, then mean of the same hour on previous days	continuous
Audio	`aud_silence`	Proportion of audio-inference samples labelled silence	count in category / all inference samples in the hour	forward fill $\leq$3 h, then mean of the same hour on previous days	continuous
	`aud_voice`	Proportion labelled voice	count in category / all inference samples in the hour	forward fill $\leq$3 h, then mean of the same hour on previous days	continuous
	`aud_noise`	Proportion labelled noise	count in category / all inference samples in the hour	forward fill $\leq$3 h, then mean of the same hour on previous days	continuous
Location (GPS)	`gps_logdist`	log(1 + distance travelled in km)	$\log(1 + \mathrm{km})$	forward fill $\leq$3 h, then mean of the same hour on previous days	continuous
	`gps_entropy`	Location entropy	–	forward fill $\leq$3 h, then mean of the same hour on previous days	continuous
	`gps_logpoints`	log(1 + number of GPS points)	$\log(1 + n)$	forward fill $\leq$3 h, then mean of the same hour on previous days	continuous
Phone lock	`lock_minutes`	Minutes phone locked in the hour (sessions spread forward, max 60)	session duration carried forward to later hours, max 60 min per hour	median of the same hour on previous days	continuous
	`lock_unlocks`	Unlock count	–	median of the same hour on previous days	continuous
Screen inactivity	`inact_minutes`	Minutes screen inactive in the hour (sessions spread forward, max 60)	session duration carried forward to later hours, max 60 min per hour	median of the same hour on previous days	continuous
	`inact_periods`	Number of inactivity periods	–	median of the same hour on previous days	continuous
Applications	`app_count`	Number of running-application records (missing when no device data)	set to missing when no core modality was observed	median of the same hour on previous days	continuous
Calls	`call_in`	Incoming calls	–	0	continuous
	`call_out`	Outgoing calls	–	0	continuous
	`call_missed`	Missed calls	–	0	continuous
	`call_minutes`	Call duration in minutes	–	0	continuous
Availability flags	`flag_activity`	1 if activity data observed in the hour	–	none (always observed)	flag
	`flag_audio`	1 if audio data observed	–	none (always observed)	flag
	`flag_gps`	1 if GPS data observed	–	none (always observed)	flag
	`flag_phonelock`	1 if phone-lock data observed	–	none (always observed)	flag
	`flag_inactivity`	1 if inactivity data observed	–	none (always observed)	flag
	`flag_app`	1 if application data observed	–	none (always observed)	flag
	`flag_device`	1 if any core modality observed	1 if activity, audio, GPS or phone-lock data observed	none (always observed)	flag
Time	`hour_sin`	sin(2*pi*hour/24)	–	none (always observed)	time
	`hour_cos`	cos(2*pi*hour/24)	–	none (always observed)	time
	`weekend`	1 on Saturday/Sunday	–	none (always observed)	time

Table S2. Selected hyperparameters (30 Optuna trials per encoder, condition I3/S0). Search space: dropout {0, 0.1, 0.2, 0.3}; learning rate {3e-4, 1e-3, 3e-3}; weight decay (squared-norm penalty) {0, 1e-4, 1e-3}; batch {16, 32, 64}; embedding dimension {2, 4} with weight decay {1e-4, 1e-3, 1e-2}; head units {8, 16}; augmentation {none, jitter 0.02, jitter 0.05}; plus encoder-specific sizes. LR, learning rate; Emb. (decay), embedding dimension (its weight decay); Aug., online augmentation (Gaussian jitter with the given standard deviation and $\pm 5\%$ scaling). Epochs: median best epoch over the tuning folds (in the main experiment early stopping was applied in each fold; see main text).

Encoder	Architecture	Dropout	LR	Weight decay	Batch	Emb. (decay)	Aug.	Epochs	Val. MAE
GRU	units 24	0.2	0.003	0.001	64	4 (0.01)	none	12	0.513
TCN	filters 12, kernel 2	0.2	0.0003	0.0001	16	4 (0.0001)	jitter 0.05	58	0.511
Transformer	d_model 24, heads 2, layers 2	0.3	0.003	0	32	2 (0.001)	jitter 0.02	2	0.509
TimesFM-lite	d_model 32, heads 4, layers 1, patch 4 h	0	0.003	0	64	4 (0.01)	none	2	0.508

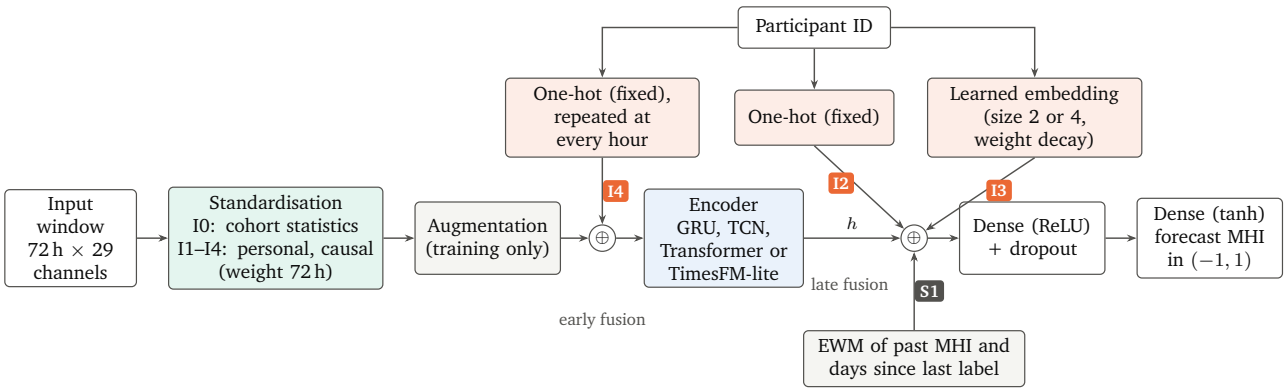

Figure S1. Where subject identity enters the forecasting models. Personal normalisation (I1) is applied to the input before the encoder and is also used by I2–I4; I0 uses cohort statistics. I4 concatenates a fixed one-hot identity vector to the input at every hour (early fusion). I2 (fixed one-hot) and I3 (learned embedding) are concatenated with the encoder output h (late fusion), as are the two static inputs of variant S1. Augmentation is active during training only. The same encoder and head are used for all identity strategies. ReLU, rectified linear unit.

Table S3. Spearman correlations with PHQ-9 totals (bootstrap 95% CI), all participants with at least three labelled days in the window. Calendar windows: 24 March–6 April (pre) and 18–31 May (post); per-participant windows: first and last 14 labelled days.

Analysis	Window	Measure	n	ρ	CI low	CI high
Level, post	Calendar	MHI	22	−0.362	−0.701	0.048
Level, pre	Calendar	MHI	39	−0.305	−0.566	0.010
Change	Calendar	MHI	20	−0.117	−0.572	0.369
Level, post	Calendar	EMA mean (absolute)	22	−0.481	−0.807	−0.021
Level, pre	Calendar	EMA mean (absolute)	39	−0.270	−0.530	0.053
Change	Calendar	EMA mean (absolute)	20	−0.129	−0.628	0.390
Level, post	Calendar	Stress (absolute)	21	−0.374	−0.750	0.135
Level, pre	Calendar	Stress (absolute)	39	−0.270	−0.530	0.053
Change	Calendar	Stress (absolute)	20	−0.024	−0.524	0.468
Level, post	Per participant	MHI	33	−0.294	−0.539	0.019
Level, pre	Per participant	MHI	41	−0.317	−0.575	−0.027
Change	Per participant	MHI	28	0.062	−0.329	0.435
Level, post	Per participant	EMA mean (absolute)	33	−0.439	−0.718	−0.083
Level, pre	Per participant	EMA mean (absolute)	41	−0.313	−0.556	−0.023
Change	Per participant	EMA mean (absolute)	28	−0.032	−0.444	0.361
Level, post	Per participant	Stress (absolute)	32	−0.407	−0.713	−0.032
Level, pre	Per participant	Stress (absolute)	39	−0.270	−0.536	0.052
Change	Per participant	Stress (absolute)	27	0.104	−0.322	0.490

Table S4. Sensitivity of the PHQ-9 validity analysis to the prior weight κ of the causal standardisation ($\kappa = 10$ in the main analysis) and to excluding each participant's first 10 labelled days, when the prior carries at least half of the weight. Spearman correlations with bootstrap 95% CI; – , too few participants.

MHI variant	Window	Analysis	n	ρ	CI low	CI high
MHI, $\kappa = 1$	Calendar	Level, post	22	−0.214	−0.585	0.212
MHI, $\kappa = 1$	Calendar	Level, pre	39	−0.242	−0.517	0.094
MHI, $\kappa = 1$	Calendar	Change	20	−0.014	−0.513	0.459
MHI, $\kappa = 1$	Per participant	Level, post	33	−0.119	−0.404	0.186
MHI, $\kappa = 1$	Per participant	Level, pre	41	−0.289	−0.569	0.024
MHI, $\kappa = 1$	Per participant	Change	28	0.125	−0.263	0.481
MHI, $\kappa = 5$	Calendar	Level, post	22	−0.310	−0.668	0.110
MHI, $\kappa = 5$	Calendar	Level, pre	39	−0.290	−0.561	0.017
MHI, $\kappa = 5$	Calendar	Change	20	−0.071	−0.540	0.408
MHI, $\kappa = 5$	Per participant	Level, post	33	−0.242	−0.495	0.074
MHI, $\kappa = 5$	Per participant	Level, pre	41	−0.308	−0.575	−0.015
MHI, $\kappa = 5$	Per participant	Change	28	0.035	−0.341	0.398
MHI, $\kappa = 10$	Calendar	Level, post	22	−0.362	−0.701	0.048
MHI, $\kappa = 10$	Calendar	Level, pre	39	−0.305	−0.566	0.010
MHI, $\kappa = 10$	Calendar	Change	20	−0.117	−0.572	0.369
MHI, $\kappa = 10$	Per participant	Level, post	33	−0.294	−0.539	0.019
MHI, $\kappa = 10$	Per participant	Level, pre	41	−0.317	−0.575	−0.027
MHI, $\kappa = 10$	Per participant	Change	28	0.062	−0.329	0.435
MHI, $\kappa = 10$, first 10 labelled days excluded	Calendar	Level, post	22	−0.362	−0.701	0.048
MHI, $\kappa = 10$, first 10 labelled days excluded	Calendar	Level, pre	2	–	–	–
MHI, $\kappa = 10$, first 10 labelled days excluded	Calendar	Change	0	–	–	–
MHI, $\kappa = 10$, first 10 labelled days excluded	Per participant	Level, post	29	−0.243	−0.551	0.090
MHI, $\kappa = 10$, first 10 labelled days excluded	Per participant	Level, pre	35	−0.095	−0.455	0.281
MHI, $\kappa = 10$, first 10 labelled days excluded	Per participant	Change	22	−0.270	−0.669	0.183
MHI, $\kappa = 20$	Calendar	Level, post	22	−0.398	−0.734	0.034
MHI, $\kappa = 20$	Calendar	Level, pre	39	−0.289	−0.553	0.028
MHI, $\kappa = 20$	Calendar	Change	20	−0.164	−0.588	0.310
MHI, $\kappa = 20$	Per participant	Level, post	33	−0.378	−0.626	−0.041
MHI, $\kappa = 20$	Per participant	Level, pre	41	−0.322	−0.574	−0.026
MHI, $\kappa = 20$	Per participant	Change	28	0.060	−0.345	0.433

Table S5. Hold-out performance of all configurations (seed ensemble for deep models). CI, 95% cluster bootstrap over participants; seed mean/SD, MAE of the individual seeds.

Model	MAE	CI low	CI high	RMSE	R^2	Within-person r	Seed mean	Seed SD	Seeds
EWM of past MHI	0.481	0.407	0.547	0.573	−0.017	−0.226	0.481	–	1
XGBoost, I2	0.488	0.417	0.553	0.584	−0.055	0.191	0.488	–	1
Transformer, I3, S1	0.491	0.409	0.561	0.581	−0.046	0.006	0.495	0.009	10
XGBoost, I1	0.491	0.425	0.555	0.581	−0.046	0.203	0.491	–	1
Transformer, I0, S1	0.493	0.415	0.558	0.578	−0.035	0.065	0.495	0.012	10
GRU, I3, S1	0.493	0.408	0.564	0.588	−0.071	0.037	0.496	0.005	10
Personal mean	0.496	0.424	0.563	0.584	−0.058	−0.200	0.496	–	1
GRU, I0, S1	0.497	0.416	0.563	0.589	−0.074	0.057	0.502	0.012	10
Transformer, I4, S0	0.498	0.411	0.569	0.591	−0.083	0.015	0.500	0.007	10
Zero forecast	0.499	0.423	0.562	0.576	−0.026	–	0.499	–	1
Transformer, I2, S0	0.499	0.415	0.569	0.591	−0.080	0.112	0.502	0.011	10
TimesFM-lite, I2, S0	0.499	0.412	0.572	0.596	−0.100	0.121	0.504	0.008	10
TimesFM-lite, I0, S1	0.500	0.415	0.572	0.591	−0.083	0.094	0.505	0.012	10
TimesFM-lite, I4, S0	0.500	0.417	0.572	0.594	−0.093	0.075	0.508	0.018	10
Transformer, I1, S0	0.501	0.414	0.573	0.595	−0.097	0.167	0.503	0.009	10
Transformer, I3, S0	0.501	0.417	0.571	0.592	−0.084	0.068	0.502	0.004	10
TimesFM-lite, I0, S0	0.502	0.418	0.573	0.593	−0.089	0.153	0.510	0.015	10
Transformer, I0, S0	0.502	0.421	0.572	0.593	−0.088	0.115	0.504	0.008	10
TCN, I3, S1	0.502	0.417	0.574	0.597	−0.104	−0.035	0.508	0.011	10
Mixed effects	0.503	0.426	0.569	0.597	−0.103	0.066	0.503	–	1
GRU, I2, S0	0.503	0.422	0.576	0.599	−0.112	0.054	0.506	0.007	10
TimesFM-lite, I1, S0	0.503	0.420	0.573	0.595	−0.096	0.046	0.510	0.015	10
TCN, I2, S0	0.504	0.416	0.578	0.604	−0.129	0.042	0.511	0.020	10
TCN, I0, S1	0.505	0.418	0.577	0.604	−0.131	−0.108	0.513	0.020	10
GRU, I3, S0	0.505	0.417	0.582	0.606	−0.137	0.073	0.510	0.005	10
TCN, I3, S0	0.505	0.419	0.579	0.602	−0.124	−0.017	0.511	0.012	10
GRU, I4, S0	0.507	0.431	0.576	0.600	−0.114	0.066	0.509	0.010	10
GRU, I1, S0	0.508	0.420	0.583	0.610	−0.152	0.081	0.510	0.003	10
TimesFM-lite, I3, S0	0.508	0.422	0.579	0.603	−0.125	−0.035	0.515	0.020	10
TimesFM-lite, I3, S1	0.509	0.421	0.580	0.604	−0.131	−0.069	0.517	0.021	10
TCN, I1, S0	0.509	0.419	0.584	0.611	−0.157	−0.093	0.515	0.020	10
TCN, I4, S0	0.509	0.420	0.583	0.609	−0.149	−0.102	0.515	0.017	10
TCN, I0, S0	0.511	0.421	0.586	0.612	−0.161	−0.131	0.517	0.021	10
GRU, I0, S0	0.516	0.429	0.591	0.614	−0.169	0.083	0.519	0.010	10
XGBoost, I0	0.523	0.435	0.599	0.631	−0.232	0.049	0.523	–	1
Persistence	0.535	0.433	0.634	0.731	−0.656	0.042	0.535	–	1
Ridge (n-of-1)	0.576	0.486	0.671	0.744	−0.714	0.032	0.576	–	1

Table S6. Validation (development folds) performance of all configurations.

Model	MAE	CI low	CI high	RMSE	R^2	Within-person r	Seed mean	Seed SD	Seeds
EWM of past MHI	0.499	0.442	0.550	0.606	−0.014	−0.007	0.499	–	1
Transformer, I1, S0	0.510	0.456	0.557	0.603	−0.006	0.010	0.514	0.005	10
Transformer, I2, S0	0.510	0.454	0.558	0.604	−0.010	0.029	0.517	0.006	10
Transformer, I0, S1	0.511	0.456	0.558	0.607	−0.017	0.017	0.516	0.007	10
Transformer, I3, S1	0.513	0.458	0.560	0.608	−0.020	−0.029	0.516	0.006	10
GRU, I0, S1	0.513	0.465	0.555	0.603	−0.006	−0.005	0.517	0.003	10
GRU, I3, S1	0.513	0.462	0.557	0.602	−0.002	−0.002	0.517	0.006	10
TimesFM-lite, I4, S0	0.514	0.460	0.559	0.611	−0.030	0.024	0.525	0.012	10
TimesFM-lite, I0, S1	0.514	0.465	0.557	0.606	−0.014	−0.009	0.523	0.007	10
Transformer, I0, S0	0.514	0.460	0.560	0.610	−0.029	−0.050	0.518	0.007	10
Transformer, I4, S0	0.514	0.458	0.563	0.610	−0.028	−0.013	0.521	0.008	10
Transformer, I3, S0	0.515	0.461	0.562	0.607	−0.020	−0.005	0.519	0.007	10
TimesFM-lite, I0, S0	0.516	0.467	0.559	0.607	−0.017	0.008	0.524	0.010	10
TimesFM-lite, I3, S1	0.516	0.466	0.558	0.609	−0.024	−0.050	0.526	0.007	10
Mixed effects	0.517	0.458	0.565	0.614	−0.042	0.072	0.517	–	1
TCN, I3, S0	0.517	0.467	0.561	0.604	−0.008	−0.001	0.530	0.012	10
Personal mean	0.519	0.463	0.566	0.614	−0.042	−0.138	0.519	–	1
GRU, I3, S0	0.519	0.467	0.562	0.610	−0.028	−0.046	0.524	0.006	10
TCN, I3, S1	0.519	0.469	0.563	0.606	−0.015	−0.055	0.530	0.014	10
GRU, I1, S0	0.519	0.471	0.560	0.610	−0.030	−0.027	0.523	0.007	10
TimesFM-lite, I3, S0	0.520	0.469	0.562	0.614	−0.042	−0.101	0.529	0.011	10
TCN, I0, S1	0.520	0.476	0.560	0.607	−0.018	−0.055	0.530	0.009	10
TCN, I4, S0	0.520	0.470	0.562	0.608	−0.022	−0.040	0.536	0.012	10
TimesFM-lite, I1, S0	0.520	0.473	0.561	0.611	−0.032	−0.101	0.528	0.012	10
TimesFM-lite, I2, S0	0.521	0.470	0.564	0.613	−0.037	−0.143	0.529	0.014	10
GRU, I2, S0	0.522	0.463	0.569	0.616	−0.050	−0.039	0.527	0.007	10
TCN, I1, S0	0.522	0.476	0.563	0.609	−0.024	−0.052	0.539	0.011	10
TCN, I0, S0	0.522	0.479	0.561	0.609	−0.025	−0.056	0.533	0.010	10
GRU, I0, S0	0.522	0.475	0.565	0.613	−0.038	−0.075	0.526	0.004	10
GRU, I4, S0	0.522	0.466	0.568	0.616	−0.048	−0.033	0.527	0.005	10
TCN, I2, S0	0.523	0.476	0.564	0.609	−0.026	−0.050	0.539	0.012	10
Zero forecast	0.537	0.500	0.572	0.606	−0.016	–	0.537	–	1
XGBoost, I0	0.538	0.488	0.582	0.650	−0.167	0.016	0.538	–	1
XGBoost, I1	0.546	0.494	0.592	0.647	−0.157	0.034	0.546	–	1
XGBoost, I2	0.550	0.491	0.602	0.657	−0.192	0.009	0.550	–	1
Persistence	0.557	0.482	0.633	0.750	−0.555	0.120	0.557	–	1
Ridge (n-of-1)	0.641	0.579	0.704	0.800	−0.770	−0.111	0.641	–	1

Table S7. Spread of the hold-out forecasts (seed ensemble for deep models): standard deviation over all hold-out participant-days, mean within-participant standard deviation, standard deviation of participant means, and range. For the observed MHI the corresponding values are SD 0.57 and within-person SD 0.51.

Model	MAE	SD	Within-person SD	Between-person SD	Range
EWM of past MHI	0.481	0.246	0.118	0.217	1.084
XGBoost, I0	0.523	0.204	0.183	0.093	1.076
XGBoost, I1	0.491	0.196	0.196	0.067	1.025
XGBoost, I2	0.488	0.215	0.211	0.079	1.270
Mixed effects	0.503	0.190	0.141	0.117	1.934
Persistence	0.535	0.575	0.514	0.292	1.961
Personal mean	0.496	0.127	0.015	0.132	0.617
Ridge (n-of-1)	0.576	0.553	0.354	0.450	3.845
Zero forecast	0.499	0.000	0.000	0.000	0.000
GRU, I0, S0	0.516	0.146	0.106	0.083	1.239
GRU, I0, S1	0.497	0.144	0.102	0.097	1.226
GRU, I1, S0	0.508	0.112	0.092	0.052	1.021
GRU, I2, S0	0.503	0.120	0.079	0.086	0.969
GRU, I3, S0	0.505	0.114	0.086	0.053	1.067
GRU, I3, S1	0.493	0.119	0.082	0.074	1.017
GRU, I4, S0	0.507	0.134	0.090	0.091	1.022
TCN, I0, S0	0.511	0.099	0.062	0.039	1.416
TCN, I0, S1	0.505	0.101	0.067	0.045	1.293
TCN, I1, S0	0.509	0.075	0.058	0.025	0.641
TCN, I2, S0	0.504	0.076	0.055	0.043	0.649
TCN, I3, S0	0.505	0.077	0.052	0.027	0.805
TCN, I3, S1	0.502	0.076	0.054	0.030	0.773
TCN, I4, S0	0.509	0.084	0.060	0.041	0.723
TimesFM-lite, I0, S0	0.502	0.089	0.064	0.043	1.003
TimesFM-lite, I0, S1	0.500	0.086	0.062	0.044	0.874
TimesFM-lite, I1, S0	0.503	0.060	0.052	0.018	0.421
TimesFM-lite, I2, S0	0.499	0.052	0.038	0.023	0.470
TimesFM-lite, I3, S0	0.508	0.059	0.049	0.028	0.469
TimesFM-lite, I3, S1	0.509	0.054	0.045	0.020	0.510
TimesFM-lite, I4, S0	0.500	0.081	0.057	0.056	0.604
Transformer, I0, S0	0.502	0.028	0.019	0.018	0.174
Transformer, I0, S1	0.493	0.045	0.026	0.033	0.251
Transformer, I1, S0	0.501	0.036	0.023	0.031	0.275
Transformer, I2, S0	0.499	0.036	0.024	0.029	0.244
Transformer, I3, S0	0.501	0.039	0.029	0.033	0.319
Transformer, I3, S1	0.491	0.044	0.031	0.033	0.282
Transformer, I4, S0	0.498	0.030	0.019	0.026	0.214

Table S8. Primary and secondary comparisons of per-participant hold-out MAE (Wilcoxon signed-rank test; Holm correction for the eight primary tests). Δ = first minus second configuration; negative values favour the first. Secondary comparisons are descriptive.

Encoder	Comparison	Type	n	Median Δ	Mean Δ	p	p Holm
GRU	I3 vs I0 (S0)	Primary	19	−0.009	−0.013	0.293	1.000
GRU	I3 vs I1 (S0)	Primary	19	−0.001	−0.005	0.595	1.000
TCN	I3 vs I0 (S0)	Primary	19	−0.001	−0.003	0.984	1.000
TCN	I3 vs I1 (S0)	Primary	19	−0.004	0.000	0.515	1.000
Transformer	I3 vs I0 (S0)	Primary	19	−0.004	−0.003	0.541	1.000
Transformer	I3 vs I1 (S0)	Primary	19	−0.001	0.000	0.984	1.000
TimesFM-lite	I3 vs I0 (S0)	Primary	19	0.006	0.007	0.182	1.000
TimesFM-lite	I3 vs I1 (S0)	Primary	19	0.005	0.004	0.225	1.000
GRU	I4 vs I2 (S0)	Secondary	19	0.003	0.006	0.352	–
GRU	I2 vs I3 (S0)	Secondary	19	0.000	−0.005	0.515	–
TCN	I4 vs I2 (S0)	Secondary	19	0.004	0.005	0.258	–
TCN	I2 vs I3 (S0)	Secondary	19	−0.002	−0.004	0.829	–
Transformer	I4 vs I2 (S0)	Secondary	19	−0.001	−0.002	0.651	–
Transformer	I2 vs I3 (S0)	Secondary	19	−0.002	−0.001	0.623	–
TimesFM-lite	I4 vs I2 (S0)	Secondary	19	0.000	0.002	0.798	–
TimesFM-lite	I2 vs I3 (S0)	Secondary	19	−0.006	−0.011	0.005	–
GRU	S1 vs S0 (I0)	Secondary	19	−0.021	−0.021	0.060	–
GRU	S1 vs S0 (I3)	Secondary	19	−0.007	−0.014	0.087	–
TCN	S1 vs S0 (I0)	Secondary	19	−0.002	−0.005	0.169	–
TCN	S1 vs S0 (I3)	Secondary	19	−0.001	−0.003	0.515	–
Transformer	S1 vs S0 (I0)	Secondary	19	−0.005	−0.008	0.156	–
Transformer	S1 vs S0 (I3)	Secondary	19	−0.008	−0.009	0.032	–
TimesFM-lite	S1 vs S0 (I0)	Secondary	19	0.001	−0.003	0.441	–
TimesFM-lite	S1 vs S0 (I3)	Secondary	19	−0.003	−0.001	0.352	–

Table S9. Equivalence tests (two one-sided t tests on per-participant MAE differences, I3 minus comparator). Margin = 5% of the zero-forecast MAE in the development folds.

Encoder	Comparison	n	Mean Δ	90% CI low	90% CI high	Margin	p TOST	p Holm	Smallest bound	Equivalent
GRU	I3 vs I0 (S0)	19	−0.013	−0.031	0.006	0.027	0.105	0.105	0.031	No
GRU	I3 vs I1 (S0)	19	−0.005	−0.015	0.004	0.027	<0.001	0.002	0.015	Yes
TCN	I3 vs I0 (S0)	19	−0.003	−0.016	0.011	0.027	0.003	0.005	0.016	Yes
TCN	I3 vs I1 (S0)	19	0.000	−0.012	0.012	0.027	<0.001	0.002	0.012	Yes
Transformer	I3 vs I0 (S0)	19	−0.003	−0.010	0.004	0.027	<0.001	<0.001	0.010	Yes
Transformer	I3 vs I1 (S0)	19	0.000	−0.005	0.005	0.027	<0.001	<0.001	0.005	Yes
TimesFM-lite	I3 vs I0 (S0)	19	0.007	−0.003	0.016	0.027	<0.001	0.002	0.016	Yes
TimesFM-lite	I3 vs I1 (S0)	19	0.004	−0.001	0.010	0.027	<0.001	<0.001	0.010	Yes

Table S10. Rank-based check of the equivalence analysis: Hodges–Lehmann (HL) estimate and 90% confidence interval (inverted Wilcoxon signed-rank test) of the per-participant MAE difference, I3 minus comparator.

Encoder	Comparison	n	HL estimate	90% CI low	90% CI high	Margin	Smallest bound	Within margin
GRU	I3 vs I0 (S0)	19	−0.009	−0.024	0.005	0.027	0.024	Yes
GRU	I3 vs I1 (S0)	19	−0.002	−0.009	0.004	0.027	0.009	Yes
TCN	I3 vs I0 (S0)	19	0.000	−0.016	0.010	0.027	0.016	Yes
TCN	I3 vs I1 (S0)	19	−0.004	−0.013	0.007	0.027	0.013	Yes
Transformer	I3 vs I0 (S0)	19	−0.002	−0.008	0.005	0.027	0.008	Yes
Transformer	I3 vs I1 (S0)	19	0.000	−0.005	0.005	0.027	0.005	Yes
TimesFM-lite	I3 vs I0 (S0)	19	0.006	−0.002	0.015	0.027	0.015	Yes
TimesFM-lite	I3 vs I1 (S0)	19	0.004	−0.002	0.011	0.027	0.011	Yes

Table S11. Below-personal-baseline (BPB) days in the hold-out (9 events in 6 participants; prevalence 5.5%). Alert threshold: quantile of validation predictions equal to the validation BPB prevalence; Tier 2: alerts on two consecutive days; /pw, per person-week. No model exceeded the net benefit of alerting all or none at any threshold probability from 0.05 to 0.50. AUROC, area under the receiver operating characteristic curve; AUPRC, area under the precision–recall curve; Sens., sensitivity; PPV, positive predictive value; Cal. slope, calibration slope of the hold-out predictions.

Model	AUROC	AUPRC	Threshold	Sens.	PPV	Tier 1 /pw	Tier 2 /pw	Cal. slope
EWM of past MHI	0.470	0.061	−0.357	0.111	0.091	0.470	0.128	0.507
XGBoost, I0	0.244	0.038	−0.347	0.000	0.000	0.555	0.085	−0.021
XGBoost, I1	0.429	0.050	−0.345	0.000	0.000	0.128	0.000	0.397
XGBoost, I2	0.440	0.052	−0.345	0.000	0.000	0.341	0.043	0.420
Mixed effects	0.496	0.058	−0.253	0.000	0.000	0.598	0.213	0.477
Persistence	0.376	0.050	−0.661	0.111	0.059	0.726	0.043	0.184
Personal mean	0.371	0.045	−0.265	0.000	0.000	0.640	0.384	1.064
Ridge (n-of-1)	0.497	0.066	−0.474	0.111	0.053	0.811	0.043	0.136
Zero forecast	0.500	0.055	0.000	1.000	0.055	7.000	4.780	–
GRU, I0, S0	0.515	0.066	−0.208	0.111	0.056	0.768	0.341	0.112
GRU, I0, S1	0.505	0.064	−0.228	0.000	0.000	0.555	0.085	0.412
GRU, I1, S0	0.480	0.056	−0.209	0.000	0.000	1.024	0.213	0.267
GRU, I2, S0	0.465	0.053	−0.238	0.000	0.000	0.811	0.341	0.636
GRU, I3, S0	0.435	0.051	−0.196	0.000	0.000	0.811	0.256	0.232
GRU, I3, S1	0.493	0.057	−0.198	0.000	0.000	0.768	0.256	0.624
GRU, I4, S0	0.494	0.057	−0.228	0.000	0.000	0.939	0.384	0.419
TCN, I0, S0	0.480	0.069	−0.186	0.222	0.125	0.683	0.128	−0.181
TCN, I0, S1	0.533	0.071	−0.184	0.111	0.059	0.726	0.171	0.146
TCN, I1, S0	0.427	0.056	−0.161	0.000	0.000	0.854	0.128	−0.587
TCN, I2, S0	0.434	0.056	−0.178	0.111	0.048	0.896	0.128	−0.067
TCN, I3, S0	0.455	0.053	−0.183	0.000	0.000	0.299	0.000	−0.039
TCN, I3, S1	0.463	0.054	−0.177	0.000	0.000	0.256	0.000	0.369
TCN, I4, S0	0.449	0.057	−0.187	0.000	0.000	0.726	0.128	−0.128
TimesFM-lite, I0, S0	0.400	0.057	−0.205	0.111	0.100	0.427	0.085	0.991
TimesFM-lite, I0, S1	0.354	0.049	−0.201	0.000	0.000	0.470	0.043	1.191
TimesFM-lite, I1, S0	0.457	0.054	−0.187	0.000	0.000	0.171	0.000	0.808
TimesFM-lite, I2, S0	0.386	0.046	−0.186	0.000	0.000	0.085	0.000	1.929
TimesFM-lite, I3, S0	0.422	0.054	−0.209	0.000	0.000	0.256	0.043	−0.058
TimesFM-lite, I3, S1	0.432	0.052	−0.201	0.000	0.000	0.171	0.000	−0.560
TimesFM-lite, I4, S0	0.509	0.058	−0.227	0.000	0.000	0.512	0.213	1.283
Transformer, I0, S0	0.477	0.077	−0.210	0.000	–	0.000	0.000	0.843
Transformer, I0, S1	0.416	0.074	−0.208	0.000	–	0.000	0.000	1.996
Transformer, I1, S0	0.519	0.090	−0.199	0.000	–	0.000	0.000	2.173
Transformer, I2, S0	0.555	0.104	−0.212	0.000	–	0.000	0.000	2.950
Transformer, I3, S0	0.433	0.073	−0.202	0.000	–	0.000	0.000	1.449
Transformer, I3, S1	0.501	0.164	−0.221	0.000	–	0.000	0.000	3.456
Transformer, I4, S0	0.602	0.110	−0.227	0.000	–	0.000	0.000	2.635

Table S12. Permutation importance in the hold-out for Transformer, I1, S0 (lowest validation MAE among deep models): increase in MAE when one feature group is permuted across samples (10 permutations per seed).

Feature group	Mean $\triangle$MAE	SD	Seeds
Activity	0.001	0.001	10
Applications and calls	0.004	0.003	10
Audio	0.000	0.002	10
Availability flags	0.001	0.001	10
Location (GPS)	0.000	0.004	10
Phone lock and inactivity	0.001	0.002	10
Time encodings	0.000	0.001	10

Table S13. Leakage audit per model (mean of five seeds). App., apparent (pipeline's own cross-validation); Hon., honest (clean hold-out); Dup. val., share of validation rows whose participant-day also occurs in training.

Model	Condition	Seeds	App. R^2	App. R^2 (best seed)	Hon. R^2	App. MAE	Hon. MAE	Dup. val.
ExtraTrees	Clean	5	−0.196	−0.192	−0.058	0.544	0.494	0.000
ExtraTrees	L1 full-history label	5	−0.288	−0.281	−0.125	0.546	0.501	0.000
ExtraTrees	L2 two-sided interpolation	5	−0.186	−0.173	−0.047	0.542	0.491	0.000
ExtraTrees	L3 random early stopping	5	−0.184	−0.154	−0.058	0.542	0.494	0.000
ExtraTrees	L4 augment-then-split	5	0.574	0.580	−0.078	0.247	0.496	0.841
ExtraTrees	L5 all sources	5	0.612	0.621	−0.132	0.249	0.505	0.841
GRU	Clean	5	−0.071	−0.020	−0.168	0.526	0.513	0.000
GRU	L1 full-history label	5	−0.131	−0.111	−0.224	0.518	0.524	0.000
GRU	L2 two-sided interpolation	5	−0.082	−0.050	−0.229	0.527	0.521	0.000
GRU	L3 random early stopping	5	−0.113	−0.067	−0.192	0.530	0.514	0.000
GRU	L4 augment-then-split	5	0.104	0.122	−0.284	0.468	0.527	0.841
GRU	L5 all sources	5	0.402	0.426	−0.652	0.342	0.575	0.841
TCN	Clean	5	−0.067	−0.030	−0.251	0.530	0.524	0.000
TCN	L1 full-history label	5	−0.197	−0.099	−0.243	0.531	0.524	0.000
TCN	L2 two-sided interpolation	5	−0.058	−0.032	−0.270	0.523	0.526	0.000
TCN	L3 random early stopping	5	−0.074	−0.041	−0.245	0.530	0.523	0.000
TCN	L4 augment-then-split	5	−0.025	0.011	−0.254	0.508	0.526	0.841
TCN	L5 all sources	5	0.079	0.193	−0.428	0.465	0.552	0.841
TimesFM-lite	Clean	5	−0.090	−0.051	−0.148	0.529	0.509	0.000
TimesFM-lite	L1 full-history label	5	−0.110	−0.064	−0.173	0.513	0.517	0.000
TimesFM-lite	L2 two-sided interpolation	5	−0.145	−0.039	−0.286	0.534	0.534	0.000
TimesFM-lite	L3 random early stopping	5	−0.124	−0.052	−0.165	0.533	0.508	0.000
TimesFM-lite	L4 augment-then-split	5	0.050	0.111	−0.222	0.487	0.522	0.841
TimesFM-lite	L5 all sources	5	0.449	0.522	−0.814	0.302	0.617	0.841
Transformer	Clean	5	−0.050	−0.032	−0.125	0.513	0.504	0.000
Transformer	L1 full-history label	5	−0.065	−0.043	−0.079	0.503	0.503	0.000
Transformer	L2 two-sided interpolation	5	−0.073	−0.059	−0.133	0.520	0.502	0.000
Transformer	L3 random early stopping	5	−0.101	−0.046	−0.116	0.525	0.506	0.000
Transformer	L4 augment-then-split	5	0.040	0.061	−0.147	0.488	0.507	0.841
Transformer	L5 all sources	5	0.253	0.398	−0.440	0.394	0.557	0.841

Table S14. Additional leakage-audit conditions (all five models, mean of five seeds): conditions without the identity input, and augmentation before splitting followed by a random row-level K-fold split. Abbreviations as in Table S13.

Model	Condition	Seeds	App. R^2	App. R^2 (best seed)	Hon. R^2	App. MAE	Hon. MAE	Dup. val.
ExtraTrees	Clean, no identity	5	−0.127	−0.122	−0.070	0.533	0.491	0.000
GRU	Clean, no identity	5	−0.049	−0.026	−0.196	0.524	0.519	0.000
TCN	Clean, no identity	5	−0.060	−0.027	−0.309	0.528	0.531	0.000
TimesFM-lite	Clean, no identity	5	−0.094	−0.046	−0.162	0.529	0.515	0.000
Transformer	Clean, no identity	5	−0.045	−0.020	−0.104	0.516	0.504	0.000
ExtraTrees	L4 augment-then-split, no identity	5	0.575	0.578	−0.087	0.256	0.493	0.841
GRU	L4 augment-then-split, no identity	5	0.092	0.102	−0.265	0.476	0.523	0.841
TCN	L4 augment-then-split, no identity	5	−0.029	0.003	−0.285	0.511	0.526	0.841
TimesFM-lite	L4 augment-then-split, no identity	5	0.038	0.118	−0.242	0.491	0.524	0.841
Transformer	L4 augment-then-split, no identity	5	0.012	0.019	−0.147	0.499	0.512	0.841
ExtraTrees	L5 all sources, no identity	5	0.607	0.614	−0.109	0.256	0.504	0.841
GRU	L5 all sources, no identity	5	0.400	0.447	−0.695	0.345	0.586	0.841
TCN	L5 all sources, no identity	5	0.034	0.060	−0.403	0.479	0.548	0.841
TimesFM-lite	L5 all sources, no identity	5	0.469	0.568	−0.747	0.295	0.606	0.841
Transformer	L5 all sources, no identity	5	0.214	0.253	−0.285	0.410	0.538	0.841
ExtraTrees	L6 augment, then random K-fold	5	0.735	0.757	−0.081	0.202	0.497	0.936
GRU	L6 augment, then random K-fold	5	0.626	0.664	−0.571	0.250	0.566	0.936
TCN	L6 augment, then random K-fold	5	0.256	0.310	−0.495	0.404	0.554	0.936
TimesFM-lite	L6 augment, then random K-fold	5	0.738	0.762	−0.719	0.148	0.594	0.936
Transformer	L6 augment, then random K-fold	5	0.636	0.687	−0.834	0.224	0.615	0.936
ExtraTrees	L6 augment, then random K-fold, no identity	5	0.722	0.749	−0.087	0.214	0.493	0.936
GRU	L6 augment, then random K-fold, no identity	5	0.600	0.634	−0.640	0.263	0.589	0.936
TCN	L6 augment, then random K-fold, no identity	5	0.154	0.200	−0.422	0.448	0.553	0.936
TimesFM-lite	L6 augment, then random K-fold, no identity	5	0.746	0.777	−0.731	0.157	0.607	0.936
Transformer	L6 augment, then random K-fold, no identity	5	0.593	0.648	−0.988	0.246	0.647	0.936