

ARTICLE OPEN



Differential temporal utility of passively sensed smartphone features for depression and anxiety symptom prediction: a longitudinal cohort study

Caitlin A. Stamatis^{1,7}✉, Jonah Meyerhoff^{1,7}, Yixuan Meng², Zhi Chong Chris Lin², Young Min Cho³, Tony Liu^{2,4}, Chris J. Karr⁵, Tingting Liu^{3,6}, Brenda L. Curtis⁶, Lyle H. Ungar^{2,3} and David C. Mohr¹

While studies show links between smartphone data and affective symptoms, we lack clarity on the temporal scale, specificity (e.g., to depression vs. anxiety), and person-specific (vs. group-level) nature of these associations. We conducted a large-scale ($n = 1013$) smartphone-based passive sensing study to identify within- and between-person digital markers of depression and anxiety symptoms over time. Participants (74.6% female; M age = 40.9) downloaded the LifeSense app, which facilitated continuous passive data collection (e.g., GPS, app and device use, communication) across 16 weeks. Hierarchical linear regression models tested the within- and between-person associations of 2-week windows of passively sensed data with depression (PHQ-8) or generalized anxiety (GAD-7). We used a shifting window to understand the time scale at which sensed features relate to mental health symptoms, predicting symptoms 2 weeks in the future (distal prediction), 1 week in the future (medial prediction), and 0 weeks in the future (proximal prediction). Spending more time at home relative to one's average was an early signal of PHQ-8 severity (distal $\beta = 0.219$, $p = 0.012$) and continued to relate to PHQ-8 at medial ($\beta = 0.198$, $p = 0.022$) and proximal ($\beta = 0.183$, $p = 0.045$) windows. In contrast, circadian movement was proximally related to ($\beta = -0.131$, $p = 0.035$) but did not predict (distal $\beta = 0.034$, $p = 0.577$; medial $\beta = -0.089$, $p = 0.138$) PHQ-8. Distinct communication features (i.e., call/text or app-based messaging) related to PHQ-8 and GAD-7. Findings have implications for identifying novel treatment targets, personalizing digital mental health interventions, and enhancing traditional patient-provider interactions. Certain features (e.g., circadian movement) may represent correlates but not true prospective indicators of affective symptoms. Conversely, other features like home duration may be such early signals of intra-individual symptom change, indicating the potential utility of prophylactic intervention (e.g., behavioral activation) in response to person-specific increases in these signals.

npj Mental Health Research (2024)3:1 | <https://doi.org/10.1038/s44184-023-00041-y>

INTRODUCTION

Technological advances facilitating personal sensing, or passively collected signals from networked smartphone sensors¹, stand to address critical gaps in measuring and treating affective symptoms. Features assessed using smartphones could signal novel treatment targets; for example, the daily number of calls and texts made may signal changes in social behavior relevant to depression². Similarly, personal pronoun use in text messages has been linked with depression and anxiety symptoms^{3–5}, and reductions in I-pronoun use track broad improvements in therapy⁶. Incorporating sensed data into clinical care may also enhance shared decision-making⁷. For instance, deviations in GPS-location-based features could signal relevant changes to patient depression severity that could trigger a provider notification. Finally, better understanding how personal sensing can be leveraged to reliably signal current or prospective deterioration may address a key question about existing digital mental health interventions^{8,9}, which is how best to optimize the delivery of intervention components so that the right component is received at the right time, while minimizing user burden^{10–13}.

As a foundational step in realizing this potential, studies have evaluated how sensed features relate to affective symptom

severity. Prior work shows that different sensor signals such as the number and type (i.e., incoming or outgoing) of phone calls and text messages relate to affective symptoms^{14,15}. Additional data suggest that the content of text messages predicts mood and anxiety symptoms^{3–5,16}. Even mobile phone keystroke patterns have been associated with mood states¹⁷. Other smartphone signals such as GPS-location-derived features have demonstrated associations with affective symptoms across many different studies^{4,14,18,19}; however, due to challenges with replication and generalizability, there are calls for these findings to be replicated in larger and more heterogeneous samples^{19,20}.

Additional challenges stem from the dearth of studies on how temporal characteristics impact observed relationships between sensed features and symptoms, including the data window (i.e., interval over which sensor data are collapsed) and time lag (i.e., time between of predictor and outcome measurement). Previous studies of mental health outcomes have used 24-h data windows to predict mental health outcomes lagged by short timeframes such as 1 h or 1 day²¹. Other studies have used slightly larger data windows to predict mental health outcomes at lags of 1 or 2 weeks in the future^{3,22,23}. The predictive power of different sensor types may be more or less clinically meaningful depending

¹Department of Preventive Medicine, Center for Behavioral Intervention Technologies (CBITs), Feinberg School of Medicine, Northwestern University, Chicago, IL, USA.

²Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA, USA. ³Positive Psychology Center, University of Pennsylvania, Philadelphia, PA, USA. ⁴Roblox Corporation, San Mateo, CA, USA. ⁵Audacious Software, Chicago, IL, USA. ⁶Technology & Translational Research Unit, National Institute on Drug Abuse (NIDA IRP), National Institutes of Health (NIH), Bethesda, MD, USA. ⁷These authors contributed equally: Caitlin A. Stamatis, Jonah Meyerhoff. ✉email: caitlin.stamatis@northwestern.edu

on the data window and time lag used²². For example, a recent study we conducted of text message language features as they related to depression symptom severity demonstrated that a data window of 4 weeks was the optimal aggregation for prediction⁵. Another example in social media data indicated that using a data window of 2 months to predict depression severity with time lags of between 2 and 4 weeks was the optimal analytic setup²⁴. Understanding how the relationships between sensed features and affective symptoms change depending on data windows and time lags is essential to informing the clinical utility of sensed data for mental health.

Our primary objective for this study was to evaluate smartphone sensor-based markers that prospectively relate to depression and anxiety symptoms. We examined sensed features' prospective relationships to symptom severity for depression and anxiety, as well as their utility as distal or proximal predictors of affective symptom severity, using a shifting 2-week sensor data window across various time lags to predict future affective symptoms.

METHODS

Participants

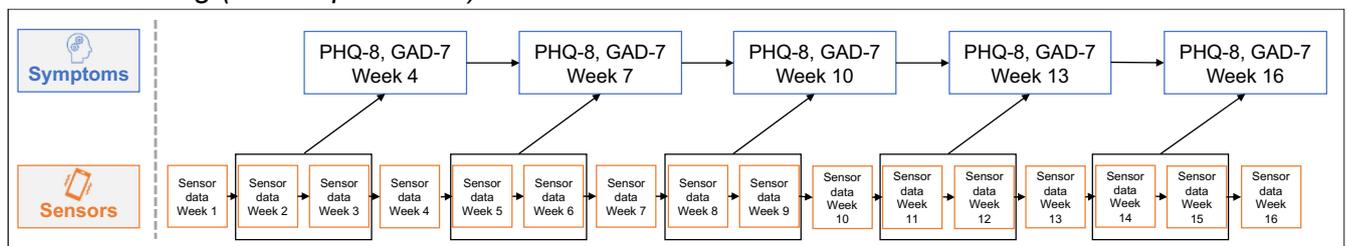
Participants were recruited in 3 waves, with a total of 1,093 enrolled. Participants in wave 1 (July–September 2019) were recruited from the Center for Behavioral Intervention Technologies

(CBITs) Health research registry and ResearchMatch.org, a national health volunteer registry supported by the National Institutes of Health. Participants in wave 2 (February–April 2020) were recruited from the CBITs Health and ResearchMatch.org registries, as well as from Focus Pointe Global, a market research data collection company. Participants in wave 3 (January–April 2021) were recruited from digital advertisements (e.g., posts on Instagram, Facebook, Twitter, craigslist, etc.), the CBITs Health and ResearchMatch.org registries, and Focus Pointe Global.

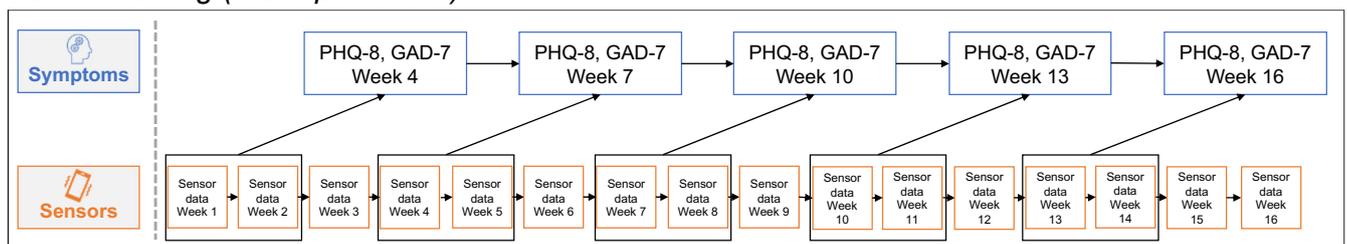
Inclusion and exclusion criteria for waves 1 and 2 did not differ. We conducted stratified sampling based on baseline PHQ-8 scores such that a minimum of 50% experienced at least moderate depression symptoms ($\text{PHQ-8} \geq 10$). In Wave 3, all participants were recruited to have at least moderate depression symptoms ($\text{PHQ-8} \geq 10$). Across all waves, participants were required to be at least 18 years old, a U.S. resident, able to read English, and own an Android smartphone with an active data and text messaging plan. Participants were excluded if they self-reported a diagnostic history of bipolar disorder, manic, or hypomanic episode, schizophrenia, or other psychotic disorder.

Participants were compensated up to \$142 for completion of assessments, as well as bonuses delivered at the end of each assessment week for participants who were running the latest version of the app and had transmitted sensed data within the past 2 days.

1a. 1-week lag (medial prediction)



1b. 2-week lag (distal prediction)



1c. 0-week lag (proximal prediction)

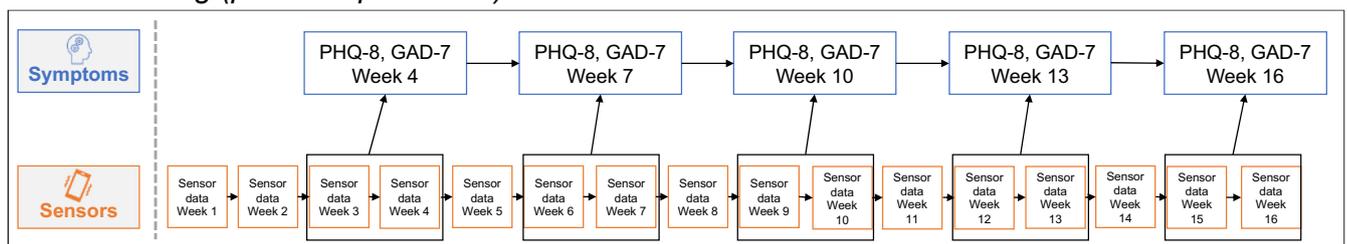


Fig. 1 Timing of associations between sensed data and affective symptoms. Testing the influence of past 2 week sensor data on subsequent week depression and anxiety symptoms (**1a**, medial prediction, 1-week lag), as well as the effect of shifting the sensor data time window on symptom prediction (**1b**, distal prediction, 2-week lag; and **1c**, proximal prediction, 0-week lag). The orange boxes in each panel depict the sliding sensor window across various lag times.

Table 1. Demographic data.				
Variable	All waves (<i>n</i> = 1013)	Wave 1 (<i>n</i> = 265)	Wave 2 (<i>n</i> = 332)	Wave 3 (<i>n</i> = 416)
Age in years, mean (sd)	40.91 (12.69)	38.89 (12.01)	43.06 (12.26)	40.48 (13.22)
Sex at birth, <i>n</i> (%)				
Female	756 (74.63%)	207 (78.11%)	241 (72.59%)	308 (74.04%)
Male	257 (25.37%)	58 (21.89%)	91 (27.41%)	108 (25.96%)
Gender identity, <i>n</i> (%)				
Woman	732 (72.26%)	197 (74.34%)	240 (72.29%)	295 (70.91%)
Man	253 (24.98%)	58 (21.89%)	91 (27.41%)	104 (25.00%)
Non-binary	15 (1.48%)	6 (2.26%)	1 (0.30%)	8 (1.92%)
Transgender	7 (0.69%)	0 (0%)	0 (0%)	7 (1.68%)
Unknown	6 (0.59%)	4 (1.51%)	0 (0%)	2 (0.48%)
Race, <i>n</i> (%)				
White	804 (79.37%)	211 (79.62%)	254 (76.51%)	339 (81.49%)
Black/African American	107 (10.56%)	20 (7.55%)	52 (15.66%)	35 (8.41%)
Asian	33 (3.26%)	9 (3.40%)	6 (1.81%)	18 (4.33%)
Native American/Alaska Native	10 (0.99%)	3 (1.13%)	4 (1.20%)	3 (0.72%)
More than one Race	53 (5.23%)	20 (7.55%)	15 (4.52%)	18 (4.33%)
Unknown	6 (0.59%)	2 (0.75%)	1 (0.30%)	3 (0.72%)
Ethnicity, <i>n</i> (%)				
Hispanic/Latinx	78 (7.7%)	25 (9.43%)	19 (5.72%)	34 (8.17%)
Non-Hispanic/Non- Latinx	932 (92.0%)	238 (89.81%)	313 (94.28%)	381 (91.59%)
Unknown	3 (0.30%)	2 (0.75%)	0 (0%)	1 (0.24%)
Highest level education completed, <i>n</i> (%)				
Some high school, no diploma	14 (1.38%)	3 (1.13%)	4 (1.20%)	7 (1.68%)
High school/GED	77 (7.60%)	12 (4.53%)	28 (8.43%)	37 (8.89%)
Some college, no degree	233 (23.0%)	42 (15.85%)	72 (21.69%)	119 (28.61%)
Associate's degree	163 (16.09%)	37 (13.96%)	68 (20.48%)	58 (13.94%)
Bachelor's degree	312 (30.80%)	94 (35.47%)	99 (29.82%)	119 (28.61%)
Master's Degree	174 (17.18%)	59 (22.26%)	53 (15.96%)	62 (14.90%)
Professional Degree	19 (1.88%)	7 (2.64%)	5 (1.51%)	7 (1.68%)
Doctoral Degree	19 (1.88%)	11 (4.15%)	2 (0.60%)	6 (1.44%)
Unknown	1 (0.10%)	0 (0%)	1 (0.30%)	0 (0%)
Marital status, <i>n</i> (%)				
Single/never married	338 (33.37%)	95 (35.85%)	106 (31.93%)	137 (32.93%)
Domestic partnership	158 (15.60%)	45 (16.98%)	51 (15.36%)	62 (14.90%)
Married	335 (33.07%)	88 (33.21%)	107 (32.23%)	140 (33.65%)
Separated	32 (3.16%)	5 (1.89%)	12 (3.61%)	15 (3.61%)
Divorced	143 (14.12%)	29 (10.94%)	55 (16.57%)	59 (14.18%)
Unknown	7 (0.69%)	3 (1.13%)	1 (0.30%)	3 (0.72%)
Household income, <i>n</i> (%)				
<\$10,000	67 (6.61%)	12 (4.53%)	23 (6.93%)	32 (7.69%)
\$10,000–19,999	90 (8.88%)	19 (7.17%)	33 (9.94%)	38 (9.13%)
\$20,000–39,999	212 (20.93%)	40 (15.09%)	66 (19.88%)	106 (25.48%)
\$40,000–59,999	206 (20.34%)	55 (20.75%)	69 (20.78%)	82 (19.71%)
\$60,000–99,999	242 (23.89%)	83 (31.32%)	78 (23.49%)	81 (19.47%)
>\$100,000	169 (16.68%)	48 (18.11%)	58 (17.47%)	63 (15.14%)
Unknown	27 (2.67%)	8 (3.02%)	5 (1.51%)	14 (3.37%)
Employment, <i>n</i> (%)				
Employed	643 (63.47%)	206 (77.74%)	211 (63.55%)	226 (54.33%)
Unemployed	135 (13.33%)	21 (7.92%)	42 (12.65%)	72 (17.31%)
Disability	104 (10.27%)	15 (5.66%)	33 (9.94%)	56 (13.46%)
Retired	49 (4.84%)	10 (3.77%)	16 (4.82%)	23 (5.53%)
Other	78 (7.70%)	12 (4.53%)	29 (8.73%)	37 (8.89%)
Unknown	4 (0.39%)	1 (0.38%)	1 (0.30%)	2 (0.48%)

Table 1 continued

Variable	All waves (<i>n</i> = 1013)	Wave 1 (<i>n</i> = 265)	Wave 2 (<i>n</i> = 332)	Wave 3 (<i>n</i> = 416)
Baseline PHQ-8				
Minimal (0–4), <i>n</i> (%)	140 (13.82%)	54 (20.38%)	86 (25.9%)	0 (0%)
Mild (5–9), <i>n</i> (%)	129 (12.73%)	45 (16.98%)	84 (25.3%)	0 (0%)
Moderate (10–14), <i>n</i> (%)	330 (32.58%)	89 (33.58%)	68 (20.48%)	173 (41.59%)
Moderate-Severe (15–19), <i>n</i> (%)	289 (28.53%)	57 (21.51%)	56 (16.87%)	176 (42.31%)
Severe (20–24), <i>n</i> (%)	125 (12.34%)	20 (7.55%)	38 (11.45%)	67 (16.11%)
Baseline GAD-7				
Minimal (0–4), <i>n</i> (%)	216 (21.32%)	74 (27.92%)	106 (31.93%)	36 (8.65%)
Mild (5–9), <i>n</i> (%)	267 (26.36%)	85 (32.08%)	83 (25.00%)	99 (23.8%)
Moderate (10–14), <i>n</i> (%)	267 (26.36%)	61 (23.02%)	69 (20.78%)	137 (32.93%)
Severe (15–21), <i>n</i> (%)	257 (25.37%)	44 (16.6%)	72 (21.69%)	141 (33.89%)

Procedure

After providing written informed consent, participants enrolled in the study for 16 weeks. All participants downloaded the LifeSense app²⁵, which automatically collected GPS-based sensor data, app, and device use data, and communication data from participants' smartphones (see Supplementary Table S1 for a list of sensors used and frequency acquired, consistent with Saeb et al., 2015). Participants responded to web-based surveys (e.g., GAD-7)²⁶ through the REDCap platform at baseline and every 3 weeks thereafter (i.e., weeks 1, 4, 7, 10, 13, 16)^{27,28}. Participants also completed PHQ-8 surveys via the LifeSense app at the beginning and end of every third week in the study²⁹. Because of this cadence, PHQ-8 instructions were modified to ask participants about their symptoms over the past week rather than past two weeks. All procedures were approved by the Northwestern University Institutional Review Board.

Analytic methods

Multilevel regression models were tested in R using the *lmerTest* package with maximum likelihood estimation³⁰. Specifically, we evaluated the associations of clustered sensor features aggregated over a 2 week window (see Supplementary Table S2 for details on clustering) with subsequent depression and anxiety symptoms. The 2 week window was selected for three reasons: to permit sufficient density of sensor data, to align with gold-standard assessments of depression and anxiety symptoms that ask about the past 2 weeks^{26,29}, and to be consistent with prior sensing studies^{4,31,32}. The prediction window was shifted such that three different models were tested for each outcome: (1) medial prediction is at a 1-week lag (Fig. 1a), (2) distal prediction is shifted back 1 week for a 2-week lag (Fig. 1b), and (3) proximal prediction is shifted forward 1 week for a 0-week lag (Fig. 1c). While there was no overlap between the sensor window and symptom reporting for distal or medial prediction, proximal prediction involved taking sensor data from the week immediately before and the week concurrent with symptom reporting (e.g., weeks 3 and 4 of sensor data predicting the week 4 symptom assessment). Sensor predictors were person-mean centered, and for each sensor predictor, both a person mean term and a within-person deviation term were included in the model. Additional model terms included time (week; centered around zero), the random intercept, and the demographic covariates of age (centered), gender, and urbanicity/rurality. See Supplementary Materials for more detail on modeling.

RESULTS

Data aggregation and demographics

Data were available from 1013 participants (74.6% female; mean age = 40.9 years [*SD* = 12.7]), including a total of 4731 PHQ-8 scores (of 5065 possible; 6.59% missing) and 4649 GAD-7 scores (of 5065 possible; 8.21% missing). Table 1 contains complete demographic data.

Primary results

Table 2 (PHQ-8) and 3 (GAD-7) present results for all within-person and between-person effects of sensor data on symptoms over time; for parsimony, only features with at least some significant relationships to outcomes are described below in the text.

Location features. Spending more time at home relative to one's own average (i.e., within-person) was associated with increased future PHQ-8 severity across prediction windows (distal $\beta = 0.219$, $p = 0.012$; medial $\beta = 0.198$, $p = 0.022$; proximal $\beta = 0.183$, $p = 0.045$). Within-person time spent at home was not significantly associated with GAD-7 severity across any of the time windows (Table 2). We observed no evidence that between-person effects for time spent at home were related to PHQ-8 or GAD-7 severity. People with greater GPS variability and mobility less severe next-week PHQ-8 (medial $\beta = -0.503$, $p = 0.046$), but this signal was absent for distal ($\beta = -0.464$, $p = 0.073$) and proximal ($\beta = -0.424$, $p = 0.093$) associations. Table 3.

Two other sensed location features were reflective of near- or medial-term PHQ-8 severity but did not predict PHQ-8 severity far in the future. First, people spending time in more frequently visited venues relative to their own average were likely to have lower impending or concurrent PHQ-8 scores (medial $\beta = -0.185$, $p = 0.003$; proximal $\beta = -0.168$, $p = 0.007$); however, going to more frequently visited venues did not prospectively predict PHQ-8 severity in the more distant future (distal $\beta = -0.064$, $p = 0.308$). Second, people who showed more circadian movement (i.e., regularity in 24-h movement patterns) relative to their own average just before and at the time of reporting depression symptoms had less severe PHQ-8 scores than those who showed less circadian movement (proximal $\beta = -0.131$, $p = 0.035$); however, circadian movement did not prospectively predict PHQ-8 severity (distal $\beta = 0.034$, $p = 0.577$; medial $\beta = -0.089$, $p = 0.138$).

Communication features. People spending more time on messaging apps relative to their own average reported more severe impending or concurrent PHQ-8 symptoms (proximal $\beta = 0.162$, $p = 0.015$), but this effect was non-significant for distal ($\beta = 0.059$, $p = 0.385$) and medial ($\beta = 0.115$, $p = 0.083$) prediction. While we did not see a significant association

Table 2. Multilevel model results predicting PHQ-8 from sensing data across shifting prediction windows.

Predictor	Sensing predicting PHQ-8 with 2-week lag ($R^2 = 0.049$)			Sensing predicting PHQ-8 with 1-week lag ($R^2 = 0.048$)			Sensing predicting PHQ-8 with 0-week lag ($R^2 = 0.053$)		
	Estimate	SE	p -value	Estimate	SE	p -value	Estimate	SE	p -value
Home duration—B	0.089	0.206	0.666	0.112	0.203	0.582	0.113	0.201	0.575
Home duration - W	0.219	0.087	0.012*	0.198	0.087	0.022*	0.183	0.091	0.045*
Circadian movement - B	-0.323	0.266	0.226	-0.189	0.273	0.490	-0.375	0.266	0.159
Circadian movement - W	0.034	0.062	0.577	-0.089	0.060	0.138	-0.131	0.062	0.035*
Location variability - B	-0.030	0.227	0.893	-0.149	0.230	0.516	-0.043	0.226	0.848
Location variability - W	-0.133	0.130	0.306	-0.129	0.129	0.316	-0.110	0.131	0.400
More frequent venues - B	-0.116	0.232	0.619	-0.092	0.232	0.693	-0.073	0.237	0.758
More frequent venues - W	-0.064	0.063	0.308	-0.185	0.062	0.003**	-0.168	0.063	0.007**
Less frequent venues - B	-0.345	0.213	0.106	-0.244	0.215	0.256	-0.295	0.220	0.179
Less frequent venues - W	-0.022	0.060	0.714	-0.069	0.058	0.234	-0.019	0.056	0.729
GPS variability and mobility - B	-0.464	0.258	0.073	-0.503	0.252	0.046*	-0.424	0.252	0.093
GPS variability and mobility - W	-0.052	0.050	0.302	-0.086	0.050	0.083	-0.037	0.050	0.458
Call and text communication - B	-0.079	0.206	0.702	-0.181	0.211	0.391	-0.197	0.211	0.350
Call and text communication - W	-0.077	0.081	0.340	-0.043	0.081	0.592	-0.049	0.079	0.534
App-based messaging - B	0.317	0.235	0.178	0.373	0.238	0.117	0.408	0.238	0.087
App-based messaging - W	0.059	0.067	0.385	0.115	0.066	0.083	0.162	0.066	0.015*
Social media - B	-0.289	0.211	0.171	-0.157	0.210	0.454	-0.241	0.206	0.243
Social media - W	-0.008	0.070	0.904	0.022	0.069	0.753	0.019	0.068	0.785
Screen-on time - B	0.503	0.208	0.016*	0.272	0.211	0.196	0.541	0.214	0.012*
Screen-on time - W	-0.009	0.050	0.857	0.000	0.047	0.995	0.037	0.047	0.424
Browser - B	0.219	0.206	0.287	0.308	0.206	0.135	0.367	0.205	0.075
Browser - W	-0.052	0.065	0.430	-0.032	0.064	0.617	0.024	0.064	0.709
Email - B	-0.107	0.199	0.592	-0.132	0.203	0.517	-0.102	0.200	0.611
Email - W	0.076	0.068	0.269	0.026	0.064	0.689	0.030	0.065	0.647
Game - B	0.006	0.205	0.978	-0.007	0.203	0.974	-0.044	0.203	0.826
Game - W	0.017	0.058	0.775	0.024	0.059	0.681	-0.028	0.059	0.634
Launcher - B	-0.596	0.223	0.008**	-0.525	0.222	0.018*	-0.653	0.224	0.004**
Launcher - W	-0.040	0.069	0.561	-0.111	0.070	0.115	-0.161	0.071	0.023*
Age	-0.603	0.174	0.001**	-0.573	0.174	0.001**	-0.581	0.173	0.001**
Male (vs0. female)	-0.521	0.169	0.002**	-0.542	0.170	0.001**	-0.563	0.169	0.001**
Urban (vs0. rural)	-0.311	0.168	0.065	-0.320	0.168	0.058	-0.307	0.168	0.068
Study week	-0.119	0.026	<0.001***	-0.107	0.026	<0.001***	-0.112	0.025	<0.001***
(Intercept)	9.458	0.163	0.000	9.467	0.163	0.000	9.460	0.163	0.000

Features highlighted in bold have at least one significant relationship to the outcome.

B between, W within.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

between within-person app-based messaging and GAD-7 at any of the time points, people engaging in more app-based messaging at the between-person level were more likely to report higher distal ($\beta = 0.486$, $p = 0.041$) and medial ($\beta = 0.481$, $p = 0.046$) GAD-7 severity; however, the association of between-person app-based messaging and GAD-7 severity was non-significant for proximal prediction ($\beta = 0.466$, $p = 0.053$). Additionally, calling and texting more relative to one's own average was associated with GAD-7 severity across all prediction windows (distal $\beta = 0.279$, $p = 0.005$; medial $\beta = 0.386$, $p < 0.001$; proximal $\beta = 0.293$, $p = 0.003$). There were no significant associations between PHQ-8 and call/text-based communication at either the within-person or between-person level.

Other phone use features. People who used the launcher more on average had lower PHQ-8 scores across time windows (distal $\beta = -0.596$, $p = 0.008$; medial $\beta = -0.525$, $p = 0.018$; proximal

$\beta = -0.653$, $p = 0.004$). When people used the launcher more relative to their own average, they reported lower impending or concurrent PHQ-8 scores (proximal $\beta = -0.161$, $p = 0.023$). Launcher use was not found to be associated with GAD-7 severity at the within or between person level. People who on average had more screen-on time tended to have greater distal ($\beta = 0.503$, $p = 0.016$) and proximal ($\beta = 0.541$, $p = 0.012$) PHQ-8 severity; however, this association was non-significant for next-week prediction (medial $\beta = 0.272$, $p = 0.196$).

Demographic effects. Higher PHQ-8 and GAD-7 severity were found for younger people (β : [0.573–1.163], p : [<0.001 –0.001]) and women (β : [0.360–0.563], p : [0.001–0.036]). People living in rural areas reported higher GAD-7 (β : [0.520–0.532], p : [0.002–0.002]), but not PHQ-8 (β : [0.307–0.320], p : [0.058–0.068]).

Time effects. There was a significant fixed effect of time, such that people reported decreasing PHQ-8 and GAD-7 severity over

Table 3. Multilevel model results predicting GAD-7 from sensing data across shifting prediction windows.

Predictor	Sensing predicting GAD-7 with 2-week lag ($R^2 = 0.058$)			Sensing predicting GAD-7 with 1-week lag ($R^2 = 0.056$)			Sensing predicting GAD-7 with 0-week lag ($R^2 = 0.057$)		
	Estimate	SE	p -value	Estimate	SE	p -value	Estimate	SE	p -value
Home duration - B	-0.154	0.209	0.460	-0.089	0.206	0.665	-0.115	0.204	0.575
Home duration - W	-0.009	0.107	0.932	-0.066	0.106	0.537	-0.126	0.112	0.261
Circadian movement - B	-0.335	0.275	0.223	-0.142	0.281	0.614	-0.331	0.273	0.226
Circadian movement - W	0.015	0.076	0.847	0.014	0.074	0.847	-0.021	0.077	0.783
Location variability - B	0.111	0.232	0.633	-0.026	0.235	0.913	0.064	0.231	0.783
Location variability - W	-0.263	0.159	0.099	-0.165	0.158	0.297	-0.146	0.161	0.366
More frequent venues - B	-0.277	0.238	0.245	-0.182	0.233	0.435	-0.169	0.236	0.472
More frequent venues - W	0.129	0.077	0.092	-0.052	0.076	0.493	-0.073	0.078	0.352
Less frequent venues - B	-0.151	0.216	0.485	-0.077	0.218	0.722	-0.106	0.224	0.635
Less frequent venues - W	0.109	0.074	0.138	0.010	0.072	0.887	-0.006	0.069	0.935
GPS variability and mobility - B	-0.109	0.262	0.679	-0.235	0.256	0.358	-0.128	0.257	0.617
GPS variability and mobility - W	-0.107	0.061	0.080	-0.089	0.061	0.149	0.061	0.062	0.324
Call and text communication - B	-0.028	0.209	0.894	-0.074	0.214	0.729	-0.076	0.215	0.725
Call and text communication - W	0.279	0.099	0.005**	0.386	0.100	<0.001***	0.293	0.098	0.003**
App-based messaging - B	0.486	0.237	0.041*	0.481	0.240	0.046*	0.466	0.240	0.053
App-based messaging - W	0.137	0.083	0.097	0.121	0.082	0.142	0.067	0.082	0.414
Social media - B	-0.237	0.212	0.264	-0.178	0.212	0.402	-0.257	0.211	0.222
Social media - W	0.012	0.088	0.887	0.082	0.085	0.339	0.066	0.085	0.437
Screen-on time - B	0.282	0.184	0.125	0.247	0.241	0.305	0.272	0.246	0.268
Screen-on time - W	-0.051	0.063	0.422	-0.041	0.056	0.469	0.013	0.057	0.823
Browser - B	0.077	0.208	0.710	0.132	0.208	0.526	0.217	0.209	0.299
Browser - W	0.005	0.080	0.953	-0.035	0.078	0.653	0.004	0.078	0.963
Email - B	0.023	0.202	0.909	-0.008	0.206	0.968	0.010	0.204	0.960
Email - W	0.084	0.084	0.316	-0.004	0.080	0.963	0.019	0.081	0.818
Game - B	0.013	0.206	0.949	-0.016	0.205	0.938	-0.010	0.205	0.961
Game - W	0.077	0.072	0.284	0.046	0.073	0.528	0.017	0.073	0.814
Launcher - B	-0.219	0.226	0.333	-0.216	0.230	0.348	-0.260	0.233	0.264
Launcher - W	-0.121	0.085	0.153	-0.130	0.086	0.131	-0.117	0.087	0.180
Age	-1.163	0.177	<0.001***	-1.135	0.177	<0.001***	-1.148	0.176	<0.001***
Male (vs. female)	-0.360	0.172	0.036*	-0.378	0.173	0.029*	-0.388	0.172	0.024*
Urban (vs. rural)	-0.520	0.170	0.002**	-0.532	0.170	0.002**	-0.530	0.170	0.002**
Study week	-0.183	0.032	<0.001***	-0.171	0.032	<0.001***	-0.167	0.031	<0.001***
(Intercept)	8.807	0.166	0.000	8.811	0.166	0.000	8.813	0.166	0.000

Features highlighted in bold have at least one significant relationship to the outcome.

B between, W within.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

the course of the study (β : [-0.107 to -0.183], p : [<0.001 to <0.001]).

Overall variability explained. The models explained a modest amount of overall variability in PHQ-8 (distal $R^2 = 0.049$; medial $R^2 = 0.048$; proximal $R^2 = 0.053$) and GAD-7 (distal $R^2 = 0.058$; medial $R^2 = 0.056$; proximal $R^2 = 0.057$) symptom severity.

DISCUSSION

In the present study, we aimed to identify passively sensed digital markers that relate to future depression and anxiety symptoms at both the within-person and between-person levels, and across multiple time windows. Location features were more strongly linked with depression symptoms, whereas communication features related to both depression and anxiety. Results highlighted the importance of the prediction lag in understanding

personally sensed signals of affective symptoms: certain features (e.g., time spent at home) were consistent predictors of symptom severity across more distal and more proximal prediction windows, whereas others (e.g., circadian movement) were only associated with next-week or current symptoms.

Overall, location features—and time spent at home in particular—were more strongly linked with depression symptoms than anxiety symptoms. The most robust predictor of depression symptoms was spending more time at home relative to one's own average, which signaled that a participant was likely to report increases in depressive symptoms 1–3 weeks later. This aligns with meta-analytic evidence indicating that greater time spent at home is one of the sensed features that most consistently relates to depression¹⁴. Broadly, spending more time at home may be reflective of reductions in motivation or hedonic capacity³³; if this is the case, the finding that increases in time spent at home relate

to future depression symptoms would align with the notion of anhedonia as an endophenotype of depression³⁴.

In contrast to location features, communication features related to both depression and anxiety symptoms, with a dissociation for communication type: messaging apps signaled impending depression, and both messaging apps and calling/texting signaled future anxiety. Social media messaging apps are feature-rich³⁵, such that their usage may reflect a range of different behaviors related to depression (e.g., “doomscrolling”; engaging in social comparison; ruminating; checking to see why others didn’t respond to a message), and they tend to involve indirect conversations about a shared visual stimulus. Conversely, calling and texting are feature-poor and primarily facilitate direct communication with others³⁵; in the context of anxiety, within-person increases in these forms of communication may signal greater activation or reassurance seeking. In general, there were more consistent associations of communication data with anxiety symptoms than depression symptoms across prediction windows and communication modalities, suggesting that changes in communication—like changes in home duration for depression—may be an especially useful signal for understanding anxiety. While studies have linked changes in calling and texting with depression symptoms in bipolar disorder^{36,37}, the absence of an association with depression in our study aligns with prior research reporting null findings around communication changes in unipolar depression^{31,38}. Continued replication of these null findings may suggest that changes in call and text based communication are not a useful proxy for the social withdrawal and decreased motivational processes that characterize depression symptoms³⁹.

By using multilevel models to disaggregate within- and between-person effects over time, we identified differential relationships of sensed features with affective symptoms across time windows that have implications for identifying novel treatment targets, personalizing digital mental health interventions, and enhancing traditional patient-provider interactions¹². One of the predominant hypothesized methods for bringing personalized digital mental health interventions to fruition is understanding how personal sensing can be leveraged to reliably signal current or prospective worsening symptoms^{8,9}. Our findings underscore that the sensing context and timing (i.e., prediction lag) are critical factors impacting the utility of sensed features as a marker of affective symptoms. For example, prior studies have shown a broad correlation between circadian movement and depression symptoms^{31,32}. Given that within-person changes in circadian movement occur immediately before and contemporaneously with depression rather than predicting symptoms further in the future, interventions in response to decreased circadian movement may benefit from strategies focused on more immediate or impending depression symptoms. Conversely, in light of the prospective, within-person relationships between time at home and depression severity, developers may consider deploying prophylactic depression-focused content (e.g., behavioral activation) in response to person-specific increases in these signals. Finally, features that are significantly related to symptoms primarily at the between-person level (e.g., launcher use with PHQ-8 or app-based messaging with GAD-7) are unlikely to be helpful signals for individualized intervention or as signals of deterioration.

It is important to consider these implications in the context of the low overall amount of variance explained (approximately 5–6% across the different outcomes and lags), as compared to the larger effect sizes seen in early sensing studies, generally in small samples^{4,31,32}. While we opted to use multilevel models for explainability, future studies may consider machine learning models to optimize variance explained in light of the high dimensionality of sensor data^{40,41}; these models may also provide greater insight into prediction accuracy metrics (e.g., rates of false positives and false negatives) to inform algorithms designed to prospectively predict clinical symptoms. Additionally, although we lagged sensors and

symptom assessments, these data are still correlational and should not be interpreted as implying causality. To the best of our knowledge, there has been no research to date that has attempted to change these sensed constructs through targeted interventions, which would provide stronger evidence of potential causality. It will also be important for future studies to vary the sensor data window—which we kept consistent at 2 weeks—along with the lag to determine impacts on predictive power, and to better understand the impact of missing data over time on observed relationships. Further, the declaration of a national emergency due to COVID-19 in March 2020 occurred partway through our second wave of data collection. We did not see differences across waves substantial enough to warrant separate analysis by wave. However, the variability in the environment since the onset of COVID-19 may have tempered some of the associations between certain features (e.g., geographic location) and symptoms due to changing routines. Additional limitations are the differences in delivery mechanism and timeframe of reporting instructions for the GAD-7 (REDCap; past 2 weeks) and PHQ-8 (in-app; past week), which may have influenced responses. Finally, given the relative lack of demographic diversity in our sample, it will be important for future studies to test whether these findings generalize across more diverse populations.

Overall, findings from this large-scale mobile sensing study point to location features as important in predicting depression symptoms, and communication features in predicting both depression and anxiety symptoms. The multilevel, longitudinal approach allowed us to identify that features such as home duration were true prospective markers of intraindividual change in depression symptoms, whereas others, such as circadian movement, may be more indicative of impending or concurrent depression symptoms.

DATA AVAILABILITY

De-identified self-report data (PHQ-8 and GAD-7) will be made available through the NIMH Data Archive at the conclusion of the study. Passively collected data are not publicly available due to potentially identifying information that could compromise participant privacy.

CODE AVAILABILITY

Code for all modeling is available from the authorship team upon request.

Received: 27 April 2023; Accepted: 19 October 2023;

Published online: 04 January 2024

REFERENCES

- Mohr, D. C., Shilton, K. & Hotopf, M. Digital phenotyping, behavioral sensing, or personal sensing: names and transparency in the digital age. *npj Digit. Med.* **3**, 1–2 (2020).
- Pratap, A. et al. The accuracy of passive phone sensors in predicting daily mood. *Depress. Anxiety* **36**, 72–81 (2019).
- Stamatis, C. A. et al. Prospective associations of text-message-based sentiment with symptoms of depression, generalized anxiety, and social anxiety. *Depress. Anxiety* **39**, 794–804 (2022).
- Meyerhoff, J. et al. Evaluation of changes in depression, anxiety, and social anxiety using smartphone sensor features: Longitudinal cohort study. *J. Med. Internet Res.* **23**, e22844 (2021).
- Liu T., et al. The relationship between text message sentiment and self-reported depression. *J. Affect. Disord.* *Proc. Natl Acad. Sci.* **119**, e2114737119 (2022). <https://doi.org/10.1016/j.jad.2021.12.048>
- Nook, E. C., Hull, T. D., Nock, M. K. & Somerville, L. H. Linguistic measures of psychological distance track symptom levels and treatment outcomes in a large set of psychotherapy transcripts. *Proc. Natl Acad. Sci.* **119**, e2114737119 (2022).
- Hsin, H., Torous, J. & Roberts, L. An adjuvant role for mobile health in psychiatry. *JAMA Psychiatry* **73**, 103–104 (2016).
- Onnela, J. P. & Rauch, S. L. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology* **41**, 1691–1696 (2016).

9. Torous, J. et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry* **20**, 318–335 (2021).
10. Bidargaddi, N., Schrader, G., Klasanja, P., Licinio, J. & Murphy, S. Designing m-Health interventions for precision mental health support. *Transl. Psychiatry* **10**, 222 (2020).
11. Radhakrishnan, K. et al. The potential of digital phenotyping to advance the contributions of mobile health to self-management science. *Nurs. Outlook* **68**, 548–559 (2020).
12. Wang, L. & Miller, L. C. Just-in-the-moment adaptive interventions (JITAI): a meta-analytical review. *Health Commun.* **35**, 1531–1544 (2020).
13. Nahum-Shani, I. et al. Just-in-time adaptive interventions (JITAI) in mobile health: key components and design principles for ongoing health behavior support. *Ann. Behav. Med.* **52**, 446–462 (2018).
14. De Angel, V. et al. Digital health tools for the passive monitoring of depression: a systematic review of methods. *NPJ Digit. Med.* **5**, 3 (2022).
15. Jacobson, N. C., Summers, B. & Wilhelm, S. Digital biomarkers of social anxiety severity: digital phenotyping using passive smartphone sensors. *J. Med. Internet Res.* **22**, e16875 (2020).
16. Razavi, R., Gharipour, A. & Gharipour, M. Depression screening using mobile phone usage metadata: a machine learning approach. *J. Am. Med. Inform. Assoc.* **27**, 522–530 (2020).
17. Zulueta, J. et al. Predicting mood disturbance severity with mobile phone keystroke metadata: a biaffect digital phenotyping study. *J. Med. Internet Res.* **20**, e241 (2018).
18. Mullick, T., Radovic, A., Shaaban, S. & Doryab, A. Predicting depression in adolescents using mobile and wearable sensors: multimodal machine learning-based exploratory study. *JMIR Form. Res.* **6**, e35807 (2022).
19. Currey, D. & Torous, J. Digital phenotyping correlations in larger mental health samples: analysis and replication. *BJPsych Open* **8**, e106 (2022).
20. Müller, S. R., Chen, X., Peters, H., Chaintreau, A. & Matz, S. C. Depression predictions from GPS-based mobility do not generalize well to large demographically heterogeneous samples. *Sci. Rep.* **11**, 14007 (2021).
21. Jacobson, N. C. & Chung, Y. J. Passive sensing of prediction of moment-to-moment depressed mood among undergraduates with clinical levels of depression sample using smartphones. *Sensors* **20**, 3572 (2020).
22. Sano A., et al. Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: observational study. *J. Med. Internet Res.* **20**, e210 (2018).
23. Opoku Asare, K. et al. Predicting depression from smartphone behavioral markers using machine learning methods, hyperparameter optimization, and feature importance analysis: exploratory study. *JMIR Mhealth Uhealth* **9**, e26540 (2021).
24. Hu, Q., Li, A., Heng, F., Li, J. & Zhu, T. Predicting depression of social media user on different observation windows. *2015 IEEE/WIC/ACM Int. Conf. Web Intell. Intell. Agent Technol. (WI-IAT)* **1**, 361–364 (2015).
25. Audacious Software. Passive Data Kit. Published online. <https://passivedatakit.org/> (2018).
26. Spitzer, R. L., Kroenke, K., Williams, J. B. & Löwe, B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch. Internal Med.* **166**, 1092–1097 (2006).
27. Harris, P. A. et al. The REDCap consortium: building an international community of software platform partners. *J. Biomed. Inform.* **95**, 103208 (2019).
28. Harris, P. A. et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* **42**, 377–381 (2009).
29. Kroenke, K. et al. The PHQ-8 as a measure of current depression in the general population. *J. Affect Disord.* **114**, 163–173 (2009).
30. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest Package: tests in linear mixed effects models. *J. Stat. Soft* **82**, 1–26 (2017).
31. Saeb, S. et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. *J. Med. Internet Res.* **17**, e175 (2015).
32. Saeb, S., Lattie, E. G., Schueller, S. M., Kording, K. P. & Mohr, D. C. The relationship between mobile phone location sensor data and depressive symptom severity. *PeerJ* **4**, e2537 (2016).
33. Treadway, M. T. & Zald, D. H. Reconsidering anhedonia in depression: Lessons from translational neuroscience. *Neurosci. Biobehav. Rev.* **35**, 537–555 (2011).
34. Pizzagalli, D. A. Depression, stress, and anhedonia: toward a synthesis and integrated model. *Annu. Rev. Clin. Psychol.* **10**, 393–423 (2014).
35. Zhang, R., N. Bazarova, N. Reddy, M. Distress disclosure across social media platforms during the COVID-19 pandemic: untangling the effects of platforms, affordances, and audiences. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21. Association for Computing Machinery. <https://doi.org/10.1145/3411764.3445134> (2021).
36. Beiwinkel, T. et al. Using smartphones to monitor bipolar disorder symptoms: a pilot study. *JMIR Mental Health* **3**, e4560 (2016).
37. Faurholt-Jepsen, M. et al. Behavioral activities collected through smartphones and the association with illness activity in bipolar disorder. *Int. J. Methods Psychiatr. Res.* **25**, 309–323 (2016).
38. Pedrelli, P. et al. Monitoring changes in depression severity using wearable and mobile sensors. *Front. Psychiatry* **11**, 584711 (2020).
39. Kupferberg, A., Bicks, L. & Hasler, G. Social functioning in major depressive disorder. *Neurosci. Biobehav. Rev.* **69**, 313–332 (2016).
40. Bishop, C. M., Nasrabadi, N. M. *Pattern Recognition and Machine Learning*. Vol 4. Springer. Accessed March 2, 2023. <https://link.springer.com/book/9780387310732> (2006).
41. Mohr, D. C., Zhang, M. & Schueller, S. M. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. *Annu. Rev. Clin. Psychol.* **13**, 23–47 (2017).

ACKNOWLEDGEMENTS

We acknowledge support from the National Institute of Mental Health (NIMH) [Grants: R01MH111610, T32MH115882, R34MH124960, K08MH128640], the National Institute on Alcohol Abuse and Alcoholism (NIAAA) [Grant: 1R01AA028032], and the Intramural Research Program of the National Institutes of Health (NIH), National Institute on Drug Abuse (NIDA).

AUTHOR CONTRIBUTIONS

C.A.S.: conceptualization, methodology, writing—original draft, writing—review & editing, visualization; J.M.: conceptualization, methodology, writing—original draft, writing—review & editing; Y.M.: formal analysis, writing—original draft, writing—review & editing; Z.C.C.L. and Y.M.C.: formal analysis; Tony Liu: formal analysis, writing—review & editing; C.J.K.: software, data curation; writing—review & editing; Tingting Liu: conceptualization, writing—review & editing; B.L.C.: conceptualization, writing—review & editing; L.H.U.: conceptualization, methodology, writing—review & editing, supervision, funding acquisition; D.C.M.: conceptualization, methodology, writing—review & editing, supervision, funding acquisition.

COMPETING INTERESTS

J.M. has accepted consulting fees from Boehringer Ingelheim. C.A.S. has received salary and equity support from Google and Akili Interactive Labs. D.C.M. has accepted honoraria and consulting fees from Otsuka Pharmaceuticals, Optum Behavioral Health, Centerstone Research Institute, and the One Mind Foundation, royalties from Oxford Press, and has an ownership interest in Adaptive Health, Inc.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s44184-023-00041-y>.

Correspondence and requests for materials should be addressed to Caitlin A. Stamatidis.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.