AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Psychopathology and Clinical Science

ISSN: 2769-7541

2025, Vol. 134, No. 1, 31–40
https://doi.org/10.1037/abn0000933

Use of Passively Collected Actigraphy Data to Detect Individual Depressive Symptoms in a Clinical Subpopulation and a General Population

George D. Price[1, 2], Amanda C. Collins[1, 3], Daniel M. Mackin[1, 3], Michael V. Heinz[1, 4], and Nicholas C. Jacobson[1, 2, 3, 4]

[1] Center for Technology and Behavioral Health, Geisel School of Medicine, Dartmouth College
[2] Quantitative Biomedical Sciences Program, Dartmouth College
[3] Department of Biomedical Data Science, Geisel School of Medicine, Dartmouth College
[4] Department of Psychiatry, Geisel School of Medicine, Dartmouth College

The presentation of major depressive disorder (MDD) can vary widely due to its heterogeneity, including inter- and intraindividual symptom variability, making MDD difficult to diagnose with standard measures in clinical settings. Prior work has demonstrated that passively collected actigraphy can be used to detect MDD at a disorder level; however, given the heterogeneous nature of MDD, comprising multiple distinct symptoms, it is important to measure the degree to which various MDD symptoms may be captured by such passive data. The current study investigated whether individual depressive symptoms could be detected from passively collected actigraphy data in a (a) clinical subpopulation (i.e., moderate depressive symptoms or greater) and (b) general population. Using data from the National Health and Nutrition Examination Survey, a large nationally representative sample ($N = 8{,}378$), we employed a convolutional neural network to determine which depressive symptoms in each population could be detected by wrist-worn, minute-level actigraphy data. Findings indicated a small-moderate correspondence between the predictions and observed outcomes for mood, psychomotor, and suicide items (area under the receiver operating characteristic curve [AUCs] = 0.58–0.61); a moderate-large correspondence for anhedonia (AUC = 0.64); and a large correspondence for fatigue (AUC = 0.74) in the clinical subpopulation ($n = 766$); and a small-moderate correspondence for sleep, appetite, psychomotor, and suicide items (AUCs = 0.56–0.60) in the general population ($n = 8{,}378$). Thus, individual depressive symptoms can be detected in individuals who likely meet the criteria for MDD, suggesting that wrist-worn actigraphy may be suitable for passively assessing these symptoms, providing important clinical implications for the diagnosis and treatment of MDD.

General Scientific Summary

The coupling of deep learning methods with passive monitoring of an individual's naturalistic movement provides a unique opportunity to detect depressive symptoms without the necessity for frequent clinical visits or self-report measures. The present work builds upon previous efforts to evaluate which depressive symptoms are best captured by passively collected physical activity data, and how this differs between individuals in the general population and individuals who meet criteria for depression. Our findings provide insight into which individual depressive symptoms may be best detected by passively collected physical activity data, providing important assessment and treatment implications for depression.

Keywords: major depressive disorder, depressive symptoms, actigraphy, passive sensing, deep learning

Supplemental materials: https://doi.org/10.1037/abn0000933.supp

This article was published Online First October 21, 2024.

Daniel Fulford served as action editor.

George D. Price https://orcid.org/0000-0002-9164-4973

George D. Price and Amanda C. Collins contributed equally as cofirst authors.

The current study was funded by the National Institute on Drug Abuse (Grant T32DA037202) to support the fellowship of Amanda C. Collins. The views expressed are those of the authors and not necessarily those of the National Institute on Drug Abuse. Michael V. Heinz has received consulting fees in areas related to his research. Michael V. Heinz has family ownership in Artisight. Nicholas C. Jacobson receives speaking and consulting fees related to his research. Nicholas C. Jacobson has received a grant from Boehringer Ingelheim.

George D. Price served as lead for conceptualization, formal analysis, methodology, writing–original draft, and writing–review and editing. Amanda C. Collins contributed equally to conceptualization, writing–original draft, and writing–review and editing. Daniel M. Mackin served in a supporting role for writing–original draft and writing–review and editing. Michael V. Heinz served in a supporting role for formal analysis, methodology, and writing–original draft. Nicholas C. Jacobson served as lead for funding acquisition, resources, and supervision and served in a supporting role for conceptualization, writing–original draft, and writing–review and editing.

Correspondence concerning this article should be addressed to George D. Price, Center for Technology and Behavioral Health, Geisel School of Medicine, Dartmouth College, 46 Centerra Parkway Suite 300, Office 336S, Lebanon, NH 03766, United States. Email: gdprice10@gmail.com

Major depressive disorder (MDD) is a highly prevalent psychiatric disorder, with over 20 million adults in the United States endorsing a depressive episode in 2020 (National Institute of Mental Health, 2023). In recent years, three areas of research in the psychopathology field have emerged, focusing on the heterogeneity of disorders (e.g., MDD), the use of passive sensing data, and the application of deep learning methods (Price et al., 2024). Pertaining specifically to MDD, this first area of research has focused on investigating MDD as a heterogeneous construct and at a symptom level (Fried & Nesse, 2015a, 2015b). Although MDD is a widespread and debilitating disorder, it is often difficult to diagnose due to the large degree of heterogeneity in presentation. Specifically, to meet MDD diagnostic criteria, a patient must have at least one of two cardinal symptoms of depression (i.e., depressed mood and/or anhedonia), as well as other associated symptoms (e.g., insomnia, fatigue) which must altogether total at least five of nine possible symptoms. Empirical evidence supports over 1,000 unique symptom profiles, meaning that the way in which an individuals' MDD presents itself can vary significantly from other individuals with MDD (Fried & Nesse, 2015a). This heterogeneity extends to intraindividual diagnosis as well, with certain individuals meeting up to nine unique symptom combinations warranting an MDD diagnosis (Zimmerman et al., 2015). The polythetic nature of the disorder serves as a barrier to effective detection and individualized treatment (Weissman, 1986). This has contributed to the longstanding interest in identifying depression subtypes or profiles characterized by unique subsets of depression symptoms (Beijers et al., 2019; Rush, 2007). However, most methods of MDD detection, including identification of MDD subtypes, rely on the assumption of symptom equivalence despite research showing that interindividual (Harald & Gordon, 2012) and intraindividual (Oquendo et al., 2004) symptom variation in individuals with MDD is prominent and should be considered during diagnosis and treatment (Lux & Kendler, 2010). Conversely, less attention has been given to the identification of predictors of individual symptom domains of depression despite their ability to predict depression course (Moos & Cronkite, 1999) and treatment response (Uher et al., 2012). However, more recent efforts have sought to investigate the heterogeneous nature of MDD, particularly at an individual-symptom level, rather than as a single syndrome (Fried & Nesse, 2015b).

The use of actigraphy has emerged as a commonly used approach to investigate MDD. Actigraphy provides an unobtrusive way to continuously collect data, rather than requiring individuals to complete self-report measures repeatedly throughout a given day. Furthermore, accelerometers are ubiquitous in everyday life as they are built into consumer-grade wrist-worn devices (e.g., FitBit; Godino et al., 2020) and integrated into smartphones (Mohammed et al., 2018), making the burden of adoption for participants low, and accelerometer information is naturally deidentified, unlike other passive streams. Moreover, data collected via actigraphy can provide measures of sleep, movement, and sedentary behaviors, and these objective data are often more reliable than subjective, self-report measures of these behaviors (Patterson et al., 1993). Despite the utility of actigraphy, however, few studies exist that have sought to merge the first two areas of psychopathology research (i.e., heterogeneity of disorders and passive sensing approaches). Consequently, actigraphy may be well essentially suited to investigate the heterogeneity of MDD and its heterogeneous presentation. The existing efforts that have sought to merge these areas of research have indicated that depressed individuals often demonstrate hyper- or insomnia, less movement (e.g., less time spent engaging in physical activity), and more sedentary behaviors (e.g., sitting still for longer periods of time; Gianfredi et al., 2022; Mammen & Faulkner, 2013; Zhai et al., 2015). Prior work has demonstrated that there are unique relationships between sleep and activity with individual depressive symptoms (Difrancesco et al., 2022). When investigating symptom domains (i.e., mood/cognition, somatic/vegetative, and sleep), significant relationships emerged between sleep and activity; specifically, longer sleep duration was associated with the somatic/vegetative domain, and lower sleep efficiency was associated with the sleep symptom dimension. These findings also revealed negative relationships between activity levels and all three symptom domains. Specifically, symptoms representing negative self-views, depressed mood, decreased appetite, and hypersomnia were associated with longer sleep duration, and sad mood and insomnia were associated with lower sleep efficiency. Moreover, negative views of the future, depressed mood, anhedonia, fatigue, weight/appetite, and sleep problems were related to lower activity levels (Difrancesco et al., 2022). Examining how objective measures of sleep and activity levels relate to individual symptoms can reveal unique patterns of MDD heterogeneity that may not otherwise be seen when examining how these constructs relate to MDD as a sum score. Thus, these studies have established a path forward for using passive sensing, particularly the use of actigraphy, for investigating MDD as a heterogeneous disorder.

Extending from previous research, and focusing on the latter two areas (i.e., passive sensing approaches and deep learning methods), a recent study utilized actigraphy to detect depression using machine learning in a nationally representative sample (Price et al., 2024). These findings indicate that actigraphy, coupled with machine learning, was able to moderately detect depression, as measured by the sum score of the Patient Health Questionnaire-9 (PHQ-9; Kroenke & Spitzer, 2002). The prior findings provide promise that depression can be detected using actigraphy data alone when applying deep learning methods.

As outlined above, there are three emerging areas of psychopathology: (a) emphasis of its heterogeneous nature, (b) use of passive sensing approaches (e.g., actigraphy), and (c) application of deep learning methods. It remains unknown, however, how these three areas can be used in combination to better understand MDD. Building off prior work that has demonstrated that MDD as a standalone construct can be detected using actigraphy (Price et al., 2024), the current work sought to investigate whether actigraphy can detect individual symptoms given that MDD is highly heterogeneous and symptom presentation can vary from person to person, even in persons with the same PHQ-9 sum score (Nemesure et al., 2024; Wichers, 2014). In addition, the current work sought to investigate this research question using deep learning models given that they have been successful in detecting MDD as a single syndrome. Indeed, if actigraphy can be coupled with deep learning models to detect both overall MDD (Price et al., 2024) and individual symptoms, then our capability to detect MDD using various approaches will contribute to our efforts as a field in developing personalized detection and prevention models for MDD (Wright & Woods, 2020).

Thus, the current study seeks to bridge three emerging areas of psychopathology research. Specifically, we aim to expand on prior work supporting actigraphy data alone for MDD detection

(Price et al., 2024) by utilizing actigraphy data to detect individual depressive symptoms (given the heterogeneous nature of MDD) using deep learning models with the intention of improving precision medicine for persons with MDD.[1] In this context, the current study will explore the degree to which individual MDD symptoms may be captured by actigraphic data in a large, nationally representative sample while introducing two areas of novelty. First, we will utilize deep learning to provide further insight into how behavioral patterns, as measured by actigraphy, can be used to detect individual MDD symptoms, as measured by the PHQ-9. Our approach aimed at symptom-level MDD detection is unique compared to more commonly explored disorder-level detection in related literature (Tlachac et al., 2023). Second, we will use deep learning and actigraphy to detect MDD symptoms in two populations: a clinical subpopulation (i.e., PHQ-9 $\geq$ 10) and a general population. Contrary to disorder-level analyses, this approach is agnostic to the substantial heterogeneity and symptom overlap present in MDD. Extrapolating from existing literature (Difrancesco et al., 2022), we predict that we will be able to moderately detect the presence of the following individual symptoms in the clinical subpopulation using machine learning and actigraphy data: depressed mood, anhedonia, sleep difficulties, fatigue, and appetite difficulties. Moreover, when investigating the general population, we predict that we will be able to moderately detect the presence of the following individual symptoms: sleep difficulties, fatigue, and appetite difficulties (i.e., not depressed mood and anhedonia as these symptoms are more specific to MDD) symptoms.

Table 1
Baseline Demographics

Characteristic	Participants, no. (%)		
	Clinical subpopulation ($n = 766$)	General population ($n = 8,378$)	p
Age, years	48.61 (16.95)	47.59 (18.54)	.14
Sex			
Female	499 (65.14)	4,290 (51.20)	<.001
Male	267 (34.86)	4,088 (48.80)	
Race and ethnicity			
Non-Hispanic			
Black	174 (22.71)	1,994 (23.80)	<.001
White	327 (42.69)	3,399 (40.57)	
Asian	29 (3.79)	936 (11.17)	
Mexican American	85 (11.10)	993 (11.85)	
Other Hispanic	112 (14.62)	794 (9.48)	
Other race (including multiracial)	39 (5.09)	262 (3.12)	
PHQ-9	14.36 (4.00)	3.24 (4.39)	<.001

Note. Baseline demographic characteristics and depression questionnaire scores of participants in the 2011–2014 National Health and Nutrition Examination Survey for Clinical Subpopulation and General Population groups. Values are listed as mean (standard deviation) for each demographic variable. A Student's t test was performed for age and PHQ-9 total, and the χ^2 test was performed for sex and race and ethnicity demographic variables. PHQ-9 = Patient Health Questionnaire-9.

Method

Participants

The data utilized in the current study is part of the publicly available and nationally representative National Health and Nutrition Examination Survey (NHANES; Centers for Disease Control and Prevention, 2020a, 2020c). For the current study, we combined data from the 2011–2012 and 2013–2014 NHANES cycles (Centers for Disease Control and Prevention, 2005). Participants were only included in the current study if they had a valid week of minute-level actigraphy data and a completed PHQ-9 assessment. Thus, our sample size consisted of 8,378 participants (referred to as the "general population").

To better investigate whether machine learning could detect individual symptoms for those who may meet the criteria for MDD, we created a subset of this sample to reflect a clinical subpopulation. Participants who had a PHQ-9 composite score of 10 or greater, and were not missing any PHQ-9 data, were included in the clinical subpopulation (composite PHQ-9 $\geq$ 10; Manea et al., 2012). See Table 1 for a breakdown of participant demographics for the clinical subpopulation ($N = 766$) and general population ($N = 8,378$). Ethics approval for the NHANES study was received from the National Center for Health Statistics Research Ethics Review Board (Protocol 2011-2017).

Materials

Actigraphy

Participants were provided an Actigraph GT3X+ to wear on their nondominant wrists for at least 1 week during their session in the Mobile Examination Center (MEC), which was programmed to begin data collection at the end of the participant's MEC session (Centers for Disease Control and Prevention, 2022). The Actigraph GT3X+ is a proprietary, triaxial accelerometer that provides a more accurate estimate of densely collected physical activity than consumer-grade counterparts (Dominick et al., 2016; Reid et al., 2017), and has shown utility in collecting and characterizing naturalistic physical activity data for individuals with depression (Peis et al., 2020). In the current study, we transform the minute-level actigraphy data into two-dimensional images using a Gramian angular field (GAF) transformation (Z. Wang & Oates, 2015), a data transformation method that preserves temporal relationships within the data and allows for the implementation of different deep learning architectures. Furthermore, leveraging GAF transformation with minute-level actigraphy data has been shown to be effective in detecting MDD presence. For a more detailed explanation of how the actigraphy data were preprocessed, including how a week of data was determined and standardized across participants, see Price et al. (2024).

PHQ-9

The PHQ-9 is a nine-item self-report measure that is often used to screen for MDD in research studies and clinical practice (e.g., during primary care visits; Kroenke & Spitzer, 2002). Individual items on the PHQ-9 reflect the frequency of each symptom of MDD over

[1] It is important to note that the two groups (i.e., clinical subpopulation and general population) in the current study are different than in Price et al. (2024) as the latter used an MDD group and no-MDD group, although all participants came from the same dataset ($N = 8,378$).

the past 2 weeks: anhedonia, depressed mood, sleep disturbances, fatigue, appetite difficulties, negative self-views, concentration difficulties, motor disturbances, and suicide. Items are scored on a scale from 0 = *not at all* to 3 = *nearly every day*. The PHQ-9 has excellent reliability and validity; greater scores on the PHQ-9 are associated with worse functioning across a range of outcomes (Kroenke et al., 2001). Reliability was also calculated for the PHQ-9 via Cronbach's alpha for the present analyses ($\alpha = .85$; Tavakol & Dennick, 2011). Additionally, prior work has indicated that a cutoff score of 10 demonstrates excellent sensitivity and specificity when classifying MDD (Kroenke et al., 2001) and that this threshold serves as an appropriate threshold for classifying an individual as "depressed" (Manea et al., 2012).

Per the NHANES Interview Setting and Mode of Administration questionnaire protocol, a trained interviewer asked the PHQ-9 questions at the MEC, using the Computer-Assisted Personal Interview system as part of the interview (Centers for Disease Control and Prevention, 2020b, 2020d). As noted above, we created a clinical subpopulation ($n = 766$) from the general population ($n = 8{,}378$) of our NHANES data by calculating the composite scores for each participant. Composite scores are calculated by summing all nine items of the PHQ-9, and higher scores represent more severe depressive symptoms. The PHQ-9 scores of the clinical subpopulation and general population fell within the moderate and minimal range of depressive symptoms, respectively (Kroenke & Spitzer, 2002).

Data Analytic Plan

Data preprocessing and analyses were conducted in Python (V 3.9; Van Rossum & Drake, 2009) with Tensorflow (V 2.12.0; TensorFlow Developers, 2023). We implemented a deep learning approach, which is both well suited to handle dense, longitudinal actigraphy data and outperforms traditional machine learning algorithms, as done in prior work (Price et al., 2024) to investigate whether minute-level, passively collected actigraphy data could be used to detect the nine individual depressive symptoms in a (a) clinical subpopulation and (b) general population. To maintain a parallel modeling approach between the clinical subpopulation and general population data splitting was modified to have similar sample sizes across the training and validation sets between approaches. Specifically, using threefold cross-validation for the clinical subpopulation and 10-fold cross-validation for the general population with stratified sampling (80%) and a completely held-out test set (20%), we used a modified "AlexNet" convolutional neural network (CNN) architecture. Deep learning methods, such as CNNs and recurrent neural networks, offer important advantages over traditional machine learning approaches when processing time series data. One of their distinct strengths is their ability to handle data with minimal preprocessing, with empirically demonstrated performance on time series (Ismail Fawaz et al., 2019). Indeed in our prior work (Price et al., 2024), we found superior performance of a deep CNN model, compared to traditional machine learning models. The AlexNet architecture, a deep CNN, has been effectively and widely applied for two-dimensional image classification problems. While actigraphy data in the raw form could be modeled as unidimensional time series, given the challenges we encountered with such an approach in prior work (Price et al., 2024), we elected for transformation of the actigraphy to two-dimensional image data using GAF. We refer the interested reader to Price et al. (2024) for further information regarding this approach. The model outcome variable was represented by an individual PHQ-9 item, resulting in nine models for each population (18 models total). We binarized each outcome variable as either 0 (endorsed a "0" on the PHQ-9 for that symptom, representing the absence of the symptom) or 1 (endorsed a "1" or greater on the PHQ-9 for that symptom, representing the presence of the symptom). In line with research using actigraphy data with a deep learning model to detect a mental health-related outcome (Heinz et al., 2022; Rahman & Adjeroh, 2019), model performance was reported as the area under the receiver operating characteristic curve (AUC), Sensitivity, Specificity, positive predictive value, and negative predictive value for both the validation set(s) and the completely held-out test set. To assess statistical significance above chance for the held-out test set 95% confidence intervals were included (Hanley & McNeil, 1982). Lastly, to provide qualitative framing for model performance, a conversion of AUC into Cohen's *D* was incorporated. Using this framework, a small-moderate correspondence between the predictions and observed outcomes reflected $0.56 \geq \text{AUC} > 0.64$, a moderate-large correspondence reflected $0.64 \geq \text{AUC} > 0.64$, and a large correspondence reflected $\text{AUC} \geq 0.72$ (Salgado, 2018).

Transparency and Openness

The data utilized for the current study are publicly available on the Center for Disease Control and Prevention website (https://wwwn.cdc.gov/nchs/nhanes/default.aspx). The analysis code is available from the corresponding author upon request. We did not preregister our hypotheses or analytic plan; however, our hypotheses and analytic plan were developed based on previous findings (Difrancesco et al., 2022) and previously implemented analyses using the same NHANES data set (Price et al., 2024). However, it is important to note that the two groups (i.e., clinical subpopulation and general population) are different than in Price et al. (2024) as the latter used an MDD group and no-MDD group, although all participants came from the same data set ($N = 8{,}378$).

Results

Clinical Subpopulation

The modified "AlexNet" model was implemented for all nine PHQ-9 symptoms as outcomes for the clinical subpopulation. The models corresponding to anhedonia (Symptom 1), mood (Symptom 2), fatigue (Symptom 4), motor (Symptom 8), and suicide (Symptom 9) all performed statistically significantly above chance. Specifically, there was a small to moderate correspondence between the predictions and observed outcomes for mood, motor, and suicide ($\text{AUC} = 0.61$, 0.58, and 0.59, respectively); a moderate to large correspondence between the predictions and observed outcomes for anhedonia ($\text{AUC} = 0.64$); and a large correspondence between the predictions and observed outcomes for fatigue ($\text{AUC} = 0.74$; Table 2). Descriptive statistics on the individual items for the PHQ-9 are provided in Table 1 in the online supplemental materials and a correlation matrix of each symptom's model predictions is provided in Figure 1 in the online supplemental materials.

Table 2
Model Performance for the Clinical Subpopulation

Model	Test set						Validation set(s)					
	AUC [95% CI]	Optimal cut point	Sensitivity	Specificity	PPV	NPV	AUC	Optimal cut point	Sensitivity	Specificity	PPV	NPV
Clinical subpopulation												
Symptom 1 (anhedonia)	**0.64[a] [0.59, 0.69]**	0.69	0.66	0.59	0.90	0.23	0.61 ± 0.02	0.71 ± 0.05	0.70 ± 0.21	0.57 ± 0.01	0.90	0.23
Symptom 2 (mood)	**0.61[b] [0.55, 0.67]**	0.57	0.90	0.38	0.94	0.28	0.60 ± 0.01	0.68 ± 0.10	0.61 ± 0.08	0.61 ± 0.13	0.94	0.28
Symptom 3 (sleep)	0.50 [0.43, 0.57]	0.54	0.89	0.27	0.92	0.21	0.59 ± 0.02	0.51 ± 0.19	0.60 ± 0.19	0.74 ± 0.17	0.92	0.21
Symptom 4 (fatigue)	**0.74[c] [0.66, 0.82]**	0.92	0.68	0.75	0.98	0.08	0.72 ± 0.05	0.94 ± 0.08	0.95 ± 0.05	0.52 ± 0.17	0.98	0.08
Symptom 5 (appetite)	0.51 [0.46, 0.56]	0.49	0.35	0.78	0.87	0.23	0.55 ± 0.10	0.40 ± 0.30	0.50 ± 0.00	0.70 ± 0.17	0.87	0.23
Symptom 6 (negative self-views)	0.51 [0.46, 0.56]	0.69	0.83	0.22	0.76	0.31	0.56 ± 0.03	0.45 ± 0.12	0.80 ± 0.21	0.72 ± 0.07	0.76	0.31
Symptom 7 (concentrating)	0.53 [0.48, 0.58]	0.54	0.19	0.90	0.85	0.28	0.57 ± 0.02	0.48 ± 0.01	0.51 ± 0.26	0.71 ± 0.12	0.85	0.28
Symptom 8 (motor)	**0.58[b] [0.54, 0.62]**	0.68	0.34	0.80	0.72	0.45	0.60 ± 0.00	0.78 ± 0.07	0.66 ± 0.23	0.43 ± 0.10	0.72	0.45
Symptom 9 (suicide)	**0.59[b] [0.54, 0.64]**	0.47	0.82	0.39	0.33	0.86	0.59 ± 0.03	0.65 ± 0.05	0.49 ± 0.03	0.53 ± 0.06	0.33	0.86

Note. Model performance for the nine symptom models of the clinical subpopulation for the validation sets and held-out test set, reported as AUC, optimal cut point (Youden's Index), sensitivity, specificity, PPV, and NPV. Models that performed statistically significant above chance are bolded. AUC = area under the receiver operating characteristic curve; CI = confidence interval; PPV = positive predictive value; NPV = negative predictive value.
[a] Moderate-large correspondence. [b] Small-moderate correspondence. [c] Large correspondence between the predictions and observed outcomes.

General Population

Parallel to the clinical subpopulation, the modified "AlexNet" model was implemented for all nine PHQ-9 symptoms as outcomes for the general population. Interestingly, all nine models performed statistically significantly above chance; however, there was only a small to moderate correspondence between the predictions and observed outcomes for sleep (Symptom 3), appetite (Symptom 5), motor (Symptom 8), and suicide (Symptom 9; AUC = 0.56, 0.58, 0.58, and 0.60, respectively; Table 3). Descriptive statistics on the individual items for the PHQ-9 are provided in Table 1 in the online supplemental materials and a correlation matrix of each symptom's model predictions is provided in Figure 2 in the online supplemental materials.

Discussion

The current study demonstrates the utility of combining three blossoming areas of psychopathology to better understand MDD: the heterogeneity of disorders (e.g., MDD), the use of passive sensing approaches, and the application of deep learning methods. In particular, our findings demonstrate that, using a large, nationally representative sample, we are able to combine these three areas of research to moderately detect individual MDD symptoms using machine learning and passively collected data. Furthermore, we investigated whether actigraphy is capable of detecting MDD in two populations: a clinical subpopulation (PHQ-9 $\geq$ 10) and a general population. Specifically, in the clinical subpopulation, the MDD symptoms of anhedonia, depressed mood, fatigue, psychomotor difficulties, and suicide can be detected solely by actigraphy, providing partial support for our hypotheses that actigraphy could detect anhedonia, depressed mood, sleep difficulties, fatigue, and appetite difficulties. Surprisingly, all symptoms performed statistically significantly above chance using actigraphy in the general population, which diverged from our hypothesis that only sleep difficulties, fatigue, and appetite difficulties would be detected.

Clinical Subpopulation

The current results provide further evidence of the heterogeneity of MDD, when assessed via actigraphy, and provide important implications for the detection and treatment of MDD. Specifically, objectively assessed physical activity behaviors can be used to detect the presence of several individual MDD symptoms, including anhedonia, depressed mood, fatigue, psychomotor difficulties, and suicide. These findings are in line with prior research that has investigated the relationships between MDD symptoms and actigraphy (Difrancesco et al., 2022; Price et al., 2024). Anhedonia, in particular, has demonstrated relationships with gross motor activity and sleep midpoint, when assessed by actigraphy, and with moderate physical activity, when assessed by self-report measures (Leventhal, 2012). In particular, higher levels of anhedonia have also been linked to lower overall activity during the day and a later bedtime at night (Difrancesco et al., 2022; Leventhal, 2012). Prior work investigating the association between depressed mood and actigraphy and self-reported physical activity has demonstrated similar relationships; higher levels of feeling down, depressed, or sad have been linked with lower physical activity, lower sleep efficiency, a later sleep midpoint, and a later bedtime at night (Difrancesco et al., 2022; Leventhal, 2012). Moreover, fatigue has been associated

Table 3
Model Performance for the General Population

Model	Test set						Validation set(s)					
	AUC [95% CI]	Optimal cut point	Sensitivity	Specificity	PPV	NPV	AUC	Optimal cut point	Sensitivity	Specificity	PPV	NPV
General population												
Symptom 1 (anhedonia)	**0.55 [0.54, 0.56]**	0.5	0.37	0.72	0.31	0.77	0.60 ± 0.00	0.57 ± 0.08	0.45 ± 0.12	0.61 ± 0.08	0.34 ± 0.01	0.81 ± 0.00
Symptom 2 (mood)	**0.54 [0.53, 0.56]**	0.43	0.74	0.34	0.26	0.81	0.59 ± 0.01	0.53 ± 0.12	0.48 ± 0.03	0.63 ± 0.11	0.29 ± 0.01	0.81 ± 0.01
Symptom 3 (sleep)	**0.56[a] [0.55, 0.57]**	0.51	0.49	0.61	0.43	0.68	0.57 ± 0.02	0.54 ± 0.10	0.49 ± 0.05	0.60 ± 0.09	0.44 ± 0.02	0.70 ± 0.01
Symptom 4 (fatigue)	**0.53 [0.52, 0.54]**	0.48	0.73	0.33	0.51	0.56	0.57 ± 0.02	0.43 ± 0.08	0.49 ± 0.04	0.70 ± 0.08	0.54 ± 0.01	0.59 ± 0.01
Symptom 5 (appetite)	**0.58[a] [0.57, 0.59]**	0.44	0.62	0.50	0.29	0.80	0.59 ± 0.03	0.54 ± 0.12	0.44 ± 0.10	0.62 ± 0.12	0.31 ± 0.03	0.82 ± 0.02
Symptom 6 (negative self-views)	**0.55 [0.53, 0.57]**	0.47	0.46	0.64	0.21	0.85	0.59 ± 0.02	0.44 ± 0.08	0.42 ± 0.06	0.72 ± 0.09	0.22 ± 0.00	0.90 ± 0.02
Symptom 7 (concentrating)	**0.55 [0.53, 0.57]**	0.49	0.44	0.65	0.21	0.85	0.57 ± 0.02	0.57 ± 0.17	0.49 ± 0.03	0.57 ± 0.17	0.22 ± 0.02	0.86 ± 0.01
Symptom 8 (motor)	**0.58[a] [0.56, 0.60]**	0.43	0.53	0.6	0.14	0.91	0.61 ± 0.02	0.52 ± 0.11	0.39 ± 0.08	0.68 ± 0.09	0.15 ± 0.00	0.93 ± 0.00
Symptom 9 (suicide)	**0.60[a] [0.57, 0.64]**	0.49	0.61	0.61	0.05	0.98	0.65 ± 0.02	0.62 ± 0.11	0.51 ± 0.12	0.64 ± 0.12	0.07 ± 0.00	0.98 ± 0.00

Note. Model performance for the nine symptom models of the general population for the validation sets and held-out test set, reported as AUC, optimal cut point (Youden's Index), sensitivity, specificity, PPV, and NPV. Models that performed statistically significant above chance are bolded. AUC = area under the receiver operating characteristic curve; CI = confidence interval; PPV = positive predictive value; NPV = negative predictive value.
[a] Small-moderate correspondence between the predictions and observed outcomes.

with overall lower activity and a later sleep midpoint. Importantly, the current findings further strengthen a growing body of research showing the relationship between different measures of physical activity and certain symptoms of anhedonia, depressed mood, fatigue, psychomotor difficulties, and suicide. However, these results extend upon the current literature by demonstrating that actigraphy is useful for detecting specific depressive symptoms without needing to extract features, making it a viable tool for the early detection of depression.

Existing literature is mixed on whether psychomotor disturbances are related to actigraphy, including physical activity and sleep. Indeed, while differences in activity levels have been identified in depressed individuals with and without psychomotor retardation symptoms (Krane-Gartiser et al., 2015), other recent research indicates that neither psychomotor retardation nor psychomotor agitation are associated with activity levels or sleep disturbances as measured by actigraphy (Difrancesco et al., 2022). Thus, our findings add to the literature and support the prior evidence that psychomotor disturbances are related to actigraphy. However, further investigation to determine if a clearer pattern emerges is warranted to better conclude if psychomotor disturbances can be assessed with actigraphy is warranted.

Prior research has indicated a potential link between suicidal thoughts and behaviors (including ideation and nonsuicidal self-injury) in undergraduate populations (Burke et al., 2022), general populations (Bernert et al., 2017; Difrancesco et al., 2022), and in populations with current suicidality (Bernert et al., 2017; Littlewood et al., 2019) when sleep is measured by actigraphy (Liu et al., 2020). Indeed, prior work using item-response theory has demonstrated that the suicide item provides more information about people with greater levels of trait depression (Cumbe et al., 2020; Nguyen et al., 2014), and that individuals with higher levels of trait depression have the greatest probability of rating the suicide item as present (Barroso et al., 2019; Ma et al., 2021; Reich et al., 2018). Given the preliminary nature of these findings, future work should more thoroughly investigate the association between actigraphy and suicidal thoughts and behaviors.

Although sleep and appetite difficulties have been found to be related to actigraphy previously, the current results indicate that these two symptoms were not detected by actigraphy in our clinical subpopulation. There may be several potential explanations for why our findings diverge from the findings of Difrancesco and colleagues (Difrancesco et al., 2022). First, the current investigation utilized a clinical subpopulation from the NHANES sample, which was conceptualized as having a PHQ-9 sum score $\geq$10. Thus, participants in the clinical subpopulation self-reported moderate to severe levels of MDD symptoms whereas the sample in Difrancesco et al. (2022) self-reported mild levels of MDD symptoms on average. Prior work suggests that the sleep and appetite items on the PHQ-9 demonstrate poor discrimination between levels of depression severity (Ma et al., 2021), so studies that aggregate across the entire dimension of depression severity may be more likely to identify significant associations than those investigating clinical subpopulations and general populations separately. Second, our conceptualization of actigraphy included minute-level measurements, which differed from Difrancesco et al. (2022) as they conceptualized actigraphy as day-level gross motor activity, sleep duration, sleep efficiency, relative amplitude, and sleep midpoint. Third, we used the PHQ-9, which measures sleep disturbances with one item: difficulties falling asleep,

staying asleep, or sleeping too much. Similarly, appetite difficulties are also assessed using one item: decrease or increase in appetite. In contrast, Difrancesco et al. used the Inventory of Depressive Symptomatology (Rush et al., 1996), which measures sleep disturbances with four different items: (a) waking up too early, (b) difficulties falling asleep, (c) difficulties staying asleep, and (d) sleeping too much. Similarly, appetite *and* weight are also assessed using four items in the Inventory for Depressive Symptomatology: (a) increased appetite, (b) decreased appetite, (c) increased weight, and (d) decreased weight. As such, Difrancesco and colleagues were able to investigate the directionality of sleep and appetite symptoms as heterogeneous constructs, and each component of these symptoms was independently related to gross motor activity, sleep duration, sleep efficiency, relative amplitude, and/or sleep midpoint (Difrancesco et al., 2022).

General Population

Regarding the general population (i.e., all NHANES participants), we predicted that our models would detect the presence of sleep difficulties, fatigue, and appetite difficulties, as these are symptoms commonly associated with actigraphy across the literature (Barragán et al., 2021; Sadeh, 2011; Tsai et al., 2023); however, this hypothesis was only partially supported by our findings as the models were able to detect a small correspondence between the predictions and observed outcomes of sleep difficulties and appetite difficulties, as well as psychomotor difficulties and suicidality. Interestingly, further introspection into the latter two findings revealed that, for individuals who endorsed a 1 or greater on either of these items, the PHQ-9 scores fell within the moderate range of severity (i.e., individuals with some degree of psychomotor difficulties in the general population endorsed an overall moderate severity of depressive symptoms). However, this was not the case for any other of the seven symptoms (i.e., individuals with some degree of anhedonia in the general population only endorsed an overall mild severity of depressive symptoms). Thus, it is possible that our models may have found small correspondences for psychomotor difficulties and suicidality given this difference in overall severity if participants endorsed a 1 or greater on these items.

Importantly, the overlap between symptoms of MDD and other psychopathology is the greatest of all *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition (*DSM-5*) disorders. Indeed, the diagnostic criteria for MDD include 10 of the 15 most nonspecific symptoms in the *DSM-5* (Forbes et al., 2023). For example, sleep and appetite difficulties are transdiagnostic symptoms that emerge across several disorders that are comorbid with MDD, including sleep disorders (e.g., insomnia disorder; Carney et al., 2009), eating disorders (e.g., anorexia nervosa; Cooper et al., 2020), and anxiety disorders (e.g., generalized anxiety disorder; Hartwig et al., 2019). Thus, it is possible that our models examining the general population are detecting these symptoms independent of the presence of MDD. Therefore, in the general population models, our findings highlight the ability of actigraphy to capture the presence of these symptoms not only in the context of MDD but transdiagnostically across the spectrum of psychopathology. Even when MDD is present, it more often than not co-occurs with other psychological disorders (Kessler et al., 2003; Zimmerman et al., 2002); therefore, it is possible that actigraphic data are capturing transdiagnostic processes even in the context of MDD. Further investigation is needed to determine whether these findings replicate in another general sample, and in samples that can better account for co-occurring psychopathology.

Clinical Implications

Our findings suggest that actigraphy, coupled with deep learning, is capable of detecting individual MDD symptoms, including fatigue and the cardinal symptoms of MDD: anhedonia and depressed mood. Given that objective measures are less burdensome than repeated self-reported symptom measures that capture symptoms on a similar timescale (i.e., ecological momentary assessment multiple times daily), and that they can detect minute-level changes in activity and sleep patterns, these findings highlight the utility of objective measures that map onto MDD and its individual symptoms. While validated MDD screening measures, such as the PHQ-9 use a simple sum score to assess MDD risk, our approach allows for a more precise identification of those depressive symptoms that are most effectively captured by actigraphy. The added granularity provided by a symptom-level approach is necessary for the alignment of individual depression symptoms with their most suitable passive sensing modalities; it stands to reason, and is supported by our results and those of others (Tlachac et al., 2023), that single passive sensing modalities likely predict some symptoms better than others; thus, the performance on simple sum-score outcomes is potentially diluted by those symptoms not well captured. In the present work, for example, we demonstrate that while actigraphy may partially capture the cardinal symptoms of depressed mood and anhedonia in a clinical subpopulation, this may not be the case for the general population (see Tables 2 and 3). Furthermore, a symptom-level approach has the potential to predict the likelihood of depression by requiring the presence of particular cardinal symptoms (e.g., anhedonia and depressed mood), offering a more nuanced assessment than predicting a simple sum score.

Moreover, objective measures, such as actigraphy, may be used in tandem with digital interventions, including just-in-time adaptive interventions, which provide therapeutic support when needed based on individual symptoms and functioning (Nahum-Shani et al., 2015; L. Wang & Miller, 2020). Prior research has demonstrated that behavioral activation (BA) is an effective treatment in increasing physical activity and mood and decreasing sedentary behaviors in individuals with MDD (Cuijpers et al., 2007). Recent advances in therapeutic interventions have resulted in adaptations of BA for depressed individuals in a digital format, leading to BA being more readily accessible and time sensitive (Huguet et al., 2016). Given the advancements in the digital intervention space, the findings from the current study are timely. Indeed, there remains a significant opportunity to integrate minute-level actigraphy and treatments, such as BA, to rapidly detect and treat depression symptoms at a given time, especially through the delivery of just-in-time adaptive interventions. Future research should aim to investigate whether actigraphy, when paired with machine learning, can detect changes in individual depression symptoms on smaller time scales with more densely sampled outcomes (e.g., via ecological momentary assessment) in order to move toward assessment and delivery of interventions in real time.

Strengths and Limitations

Our work builds upon previous efforts in depression detection to leverage passively collected actigraphy data coupled with a unique

deep-learning framework to detect individual depressive symptoms in a clinical subpopulation and a general population using the nationally representative NHANES 2011–2014 study. Moreover, the current findings provide insight into which individual depressive symptoms may be best detected by passively collected physical activity data, providing important assessment and treatment implications for depression.

Despite the clear strengths of the current work, there are also limitations to be addressed. First, we included participants in the clinical subpopulation based on their sum score of the PHQ-9, which is a self-report measure. Prior research has raised concerns about the use of sum scores to classify MDD given its wide heterogeneity (Fried & Nesse, 2015b). Given that NHANES is a large national database that does not include assessment via clinical interview or review of electronic medical records, we utilized a PHQ-9 cutoff score of 10 for the clinical subpopulation, which previous research has shown to be a reliable bifurcator with adequate sensitivity and specificity (Manea et al., 2012). Second, the manner in which the PHQ-9 assesses the symptoms of sleep and appetite difficulties does not allow for differentiation between maladaptive increases and decreases in these domains, as noted above. Thus, it is likely that insomnia and hypersomnia, and that increased and decreased appetite, are related to distinct patterns in actigraphy, and measuring sleep and appetite symptoms with one item instead of two contributed to our models not being able to detect sleep or appetite difficulties in either subpopulation, which has been associated with actigraphy in prior work (Difrancesco et al., 2022). Future research should investigate whether sleep and appetite difficulties can be detected with actigraphy and machine learning with other measures of MDD symptoms to better understand the relationships between activity and the somatic/vegetative domain of MDD.

Third, MDD is highly comorbid with other disorders, including anxiety disorders; however, the current data did not include other measures of psychopathology. Thus, it is possible that the current methods may extend to psychopathology more generally (i.e., disorders beyond MDD). Given the high degree of overlap in MDD symptoms with symptoms of other disorders in the *DSM-5* (Forbes et al., 2023), it may be that our current models are not detecting MDD symptoms only in the general population, specifically. For example, fatigue is a symptom of both MDD and generalized anxiety disorder. Future research should investigate whether actigraphy and deep learning models, and specifically the methods utilized in the current study, extend to the broader psychopathology and can detect other symptoms or disorders. Lastly, the outcome for each PHQ-9 symptom was updated from a continuous Likert scale (0–3) to a binarized "not endorsed" versus "endorsed" scale, which can be considered a loss of information related to the granularity of the reported outcome; however, the author's implemented this scheme in an effort to increase interpretability of the modeling.

Conclusion

The current work demonstrates that actigraphy, coupled with deep learning, is capable of detecting individual MDD symptoms in a clinical subpopulation, with moderate to severe MDD symptoms. Specifically, our models were able to detect anhedonia, depressed mood, fatigue, psychomotor difficulties, and suicidality in the clinical subpopulation. Moreover, our models detected appetite difficulties, sleep difficulties, psychomotor difficulties, and suicidality in the general population. These findings provide promise for solely using actigraphy to detect MDD individual symptoms to reduce participant burden and improve detection rates of MDD. However, actigraphy alone may not be enough to distinguish these symptoms at the general population level. Future research should aim to investigate whether actigraphy and deep learning models can detect changes in individual symptoms, whether in response to natural changes in symptoms or in direct response to digital interventions.

References

Barragán, R., Zuraikat, F. M., Tam, V., Scaccia, S., Cochran, J., Li, S., Cheng, B., & St-Onge, M.-P. (2021). Actigraphy-derived sleep is associated with eating behavior characteristics. *Nutrients*, *13*(3), Article 852. https://doi.org/10.3390/nu13030852

Barroso, S. M., Melo, A. P. S., da Silva, M. A., & Guimarães, M. D. C. (2019). Evaluation of the Brazilian version of Patient Health Questionnaire (PHQ-9) in Quilombola population using the Item Response Theory. *Salud Mental*, *42*(1), 43–50. https://doi.org/10.17711/sm.0185-3325.2019.006

Beijers, L., Wardenaar, K. J., Van Loo, H. M., & Schoevers, R. A. (2019). Data-driven biological subtypes of depression: Systematic review of biological approaches to depression subtyping. *Molecular Psychiatry*, *24*(6), 888–900. https://doi.org/10.1038/s41380-019-0385-5

Bernert, R. A., Hom, M. A., Iwata, N. G., & Joiner, T. E. (2017). Objectively assessed sleep variability as an acute warning sign of suicidal ideation in a longitudinal evaluation of young adults at high suicide risk. *The Journal of Clinical Psychiatry*, *78*(06), e678–e687. https://doi.org/10.4088/JCP.16m11193

Burke, T. A., Hamilton, J. L., Seigel, D., Kautz, M., Liu, R. T., Alloy, L. B., & Barker, D. H. (2022). Sleep irregularity and nonsuicidal self-injurious urges and behaviors. *Sleep*, *45*(6), Article zsac084. https://doi.org/10.1093/sleep/zsac084

Carney, C. E., Ulmer, C., Edinger, J. D., Krystal, A. D., & Knauss, F. (2009). Assessing depression symptoms in those with insomnia: An examination of the beck depression inventory second edition (BDI-II). *Journal of Psychiatric Research*, *43*(5), 576–582. https://doi.org/10.1016/j.jpsychires.2008.09.002

Centers for Disease Control and Prevention. (2005). *NHANES Analytic and Reporting Guidelines*. https://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/nhanes_analytic_guidelines_dec_2005.pdf

Centers for Disease Control and Prevention. (2020a). *NHANES 2011–2012 Examination Data Overview*. https://Wwwn.Cdc.Gov/Nchs/Nhanes/Search/Datapage.Aspx?Component=Examination&CycleBeginYear=2011

Centers for Disease Control and Prevention. (2020b). *NHANES 2011–2012 Questionnaire Data Overview*. https://Wwwn.Cdc.Gov/Nchs/Nhanes/Search/Datapage.Aspx?Component=Examination&CycleBeginYear=2011

Centers for Disease Control and Prevention. (2020c). *NHANES 2013–2014 Examination Data Overview*. https://Wwwn.Cdc.Gov/Nchs/Nhanes/Continuousnhanes/Overviewexam.Aspx?BeginYear=2013

Centers for Disease Control and Prevention. (2020d). *NHANES 2013–2014 Questionnaire Data Overview*. https://Wwwn.Cdc.Gov/Nchs/Nhanes/Continuousnhanes/Overviewexam.Aspx?BeginYear=2013

Centers for Disease Control and Prevention. (2022). *Data Quality Flag Summary Table for the Physical Activity Monitor (PAM) Data Collected in NHANES 2011–2014 and NNYFS*. https://wwwn.cdc.gov/nchs/nhanes/Pam/Default.aspx

Cooper, A. R., Loeb, K. L., & McGlinchey, E. L. (2020). Sleep and eating disorders: Current research and future directions. *Current Opinion in Psychology*, *34*, 89–94. https://doi.org/10.1016/j.copsyc.2019.11.005

Cuijpers, P., van Straten, A., & Warmerdam, L. (2007). Behavioral activation treatments of depression: A meta-analysis. *Clinical Psychology Review*, *27*(3), 318–326. https://doi.org/10.1016/j.cpr.2006.11.001

Cumbe, V. F. J., Muanido, A., Manaca, M. N., Fumo, H., Chiruca, P., Hicks, L., de Jesus Mari, J., & Wagenaar, B. H. (2020). Validity and item response theory properties of the Patient Health Questionnaire-9 for primary care depression screening in Mozambique (PHQ-9-MZ). *BMC Psychiatry*, *20*(1), Article 382. https://doi.org/10.1186/s12888-020-02772-0

Difrancesco, S., Penninx, B. W. J. H., Riese, H., Giltay, E. J., & Lamers, F. (2022). The role of depressive symptoms and symptom dimensions in actigraphy-assessed sleep, circadian rhythm, and physical activity. *Psychological Medicine*, *52*(13), 2760–2766. https://doi.org/10.1017/S0033291720004870

Dominick, G. M., Winfree, K. N., Pohlig, R. T., & Papas, M. A. (2016). Physical activity assessment between consumer- and research-grade accelerometers: A comparative study in free-living conditions. *JMIR mHealth and uHealth*, *4*(3), Article e110. https://doi.org/10.2196/mhealth.6281

Forbes, M. K., Neo, B., Nezami, O. M., Fried, E. I., Faure, K., Michelsen, B., Twose, M., & Dras, M. (2023). Elemental psychopathology: Distilling constituent symptoms and patterns of repetition in the diagnostic criteria of the DSM-5. *Psychological Medicine*, *54*(5), 886–894. https://doi.org/10.1017/S0033291723002544

Fried, E. I., & Nesse, R. M. (2015a). Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. *Journal of Affective Disorders*, *172*, 96–102. https://doi.org/10.1016/j.jad.2014.10.010

Fried, E. I., & Nesse, R. M. (2015b). Depression sum-scores don't add up: Why analyzing specific depression symptoms is essential. *BMC Medicine*, *13*(1), Article 72. https://doi.org/10.1186/s12916-015-0325-4

Gianfredi, V., Schaper, N. C., Odone, A., Signorelli, C., Amerio, A., Eussen, S. J. P. M., Köhler, S., Savelberg, H. H. C. M., Stehouwer, C. D. A., Dagnelie, P. C., Henry, R. M. A., van der Kallen, C. J. H., van Greevenbroek, M. M. J., Schram, M. T., & Koster, A. (2022). Daily patterns of physical activity, sedentary behavior, and prevalent and incident depression—The Maastricht Study. *Scandinavian Journal of Medicine & Science in Sports*, *32*(12), 1768–1780. https://doi.org/10.1111/sms.14235

Godino, J. G., Wing, D., De Zambotti, M., Baker, F. C., Bagot, K., Inkelis, S., Pautz, C., Higgins, M., Nichols, J., Brumback, T., Chevance, G., Colrain, I. M., Patrick, K., & Tapert, S. F. (2020). Performance of a commercial multi-sensor wearable (Fitbit Charge HR) in measuring physical activity and sleep in healthy children. *PLOS ONE*, *15*(9), Article e0237719. https://doi.org/10.1371/journal.pone.0237719

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29–36. https://doi.org/10.1148/radiology.143.1.7063747

Harald, B., & Gordon, P. (2012). Meta-review of depressive subtyping models. *Journal of Affective Disorders*, *139*(2), 126–140. https://doi.org/10.1016/j.jad.2011.07.015

Hartwig, E. M., Rufino, K. A., Palmer, C. A., Shepard, C., Alfano, C. A., Schanzer, B., Mathew, S. J., & Patriquin, M. A. (2019). Trajectories of self-reported sleep disturbance across inpatient psychiatric treatment predict clinical outcome in comorbid major depressive disorder and generalized anxiety disorder. *Journal of Affective Disorders*, *251*, 248–255. https://doi.org/10.1016/j.jad.2019.03.069

Heinz, M. V., Price, G. D., Ruan, F., Klein, R. J., Nemesure, M., Lopez, A., & Jacobson, N. C. (2022). Association of selective serotonin reuptake inhibitor use with abnormal physical movement patterns as detected using a piezoelectric accelerometer and deep learning in a nationally representative sample of noninstitutionalized persons in the US. *JAMA Network Open*, *5*(4), Article e225403. https://doi.org/10.1001/jamanetworkopen.2022.5403

Huguet, A., Rao, S., McGrath, P. J., Wozney, L., Wheaton, M., Conrod, J., & Rozario, S. (2016). A systematic review of cognitive behavioral therapy and behavioral activation apps for depression. *PLOS ONE*, *11*(5), Article e0154248. https://doi.org/10.1371/journal.pone.0154248

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P.-A. (2019). Deep learning for time series classification: A review. *Data Mining and Knowledge Discovery*, *33*(4), 917–963. https://doi.org/10.1007/s10618-019-00619-1

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., & Rush, A. J. (2003). The epidemiology of major depressive disorder. *JAMA*, *289*(23), Article 3095. https://doi.org/10.1001/jama.289.23.3095

Krane-Gartiser, K., Henriksen, T. E. G., Vaaler, A. E., Fasmer, O. B., & Morken, G. (2015). Actigraphically assessed activity in unipolar depression: A comparison of inpatients with and without motor retardation. *The Journal of Clinical Psychiatry*, *76*(09), 1181–1187. https://doi.org/10.4088/JCP.14m09106

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals*, *32*(9), 509–515. https://doi.org/10.3928/0048-5713-20020901-06

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Leventhal, A. M. (2012). Relations between anhedonia and physical activity. *American Journal of Health Behavior*, *36*(6), 860–872. https://doi.org/10.5993/AJHB.36.6.12

Littlewood, D. L., Kyle, S. D., Carter, L.-A., Peters, S., Pratt, D., & Gooding, P. (2019). Short sleep duration and poor sleep quality predict next-day suicidal ideation: An ecological momentary assessment study. *Psychological Medicine*, *49*(3), 403–411. https://doi.org/10.1017/S0033291718001009

Liu, R. T., Steele, S. J., Hamilton, J. L., Do, Q. B. P., Furbish, K., Burke, T. A., Martinez, A. P., & Gerlus, N. (2020). Sleep and suicide: A systematic review and meta-analysis of longitudinal studies. *Clinical Psychology Review*, *81*, Article 101895. https://doi.org/10.1016/j.cpr.2020.101895

Lux, V., & Kendler, K. S. (2010). Deconstructing major depression: A validation study of the DSM-IV symptomatic criteria. *Psychological Medicine*, *40*(10), 1679–1690. https://doi.org/10.1017/S0033291709992157

Ma, S., Yang, J., Yang, B., Kang, L., Wang, P., Zhang, N., Wang, W., Zong, X., Wang, Y., Bai, H., Guo, Q., Yao, L., Fang, L., & Liu, Z. (2021). The Patient Health Questionnaire-9 versus the Hamilton Rating Scale for depression in assessing major depressive disorder. *Frontiers in Psychiatry*, *12*, Article 747139. https://doi.org/10.3389/fpsyt.2021.747139

Mammen, G., & Faulkner, G. (2013). Physical activity and the prevention of depression: A systematic review of prospective studies. *American Journal of Preventive Medicine*, *45*(5), 649–657. https://doi.org/10.1016/j.amepre.2013.08.001

Manea, L., Gilbody, S., & McMillan, D. (2012). Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): A meta-analysis. *Canadian Medical Association Journal*, *184*(3), E191–E196. https://doi.org/10.1503/cmaj.110829

Mohammed, Z., Elfadel, I., & Rasras, M. (2018). Monolithic Multi Degree of Freedom (MDoF) capacitive MEMS accelerometers. *Micromachines*, *9*(11), Article 602. https://doi.org/10.3390/mi9110602

Moos, R. H., & Cronkite, R. C. (1999). Symptom-based predictors of a 10-year chronic course of treated depression. *The Journal of Nervous & Mental Disease*, *187*(6), 360–368. https://doi.org/10.1097/00005053-199906000-00005

Nahum-Shani, I., Hekler, E. B., & Spruijt-Metz, D. (2015). Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. *Health Psychology*, *34*(Suppl), 1209–1219. https://doi.org/10.1037/hea0000306

National Institute of Mental Health. (2023). *Major Depression*. https://www.nimh.nih.gov/health/statistics/major-depression

Nemesure, M. D., Collins, A. C., Price, G. D., Griffin, T. Z., Pillai, A., Nepal, S., Heinz, M. V., Lekkas, D., Campbell, A. T., & Jacobson, N. C. (2024). Depressive symptoms as a heterogeneous and constantly evolving dynamical system: Idiographic depressive symptom networks of rapid symptom changes among persons with major depressive disorder. *Journal of*

Psychopathology and Clinical Science, *133*(2), 155–166. https://doi.org/10.1037/abn0000884

Nguyen, T. H., Han, H.-R., Kim, M. T., & Chan, K. S. (2014). An introduction to item response theory for patient-reported outcome measurement. *The Patient—Patient-Centered Outcomes Research*, *7*(1), 23–35. https://doi.org/10.1007/s40271-013-0041-0

Oquendo, M. A., Barrera, A., Ellis, S. P., Li, S., Burke, A. K., Grunebaum, M., Endicott, J., & Mann, J. J. (2004). Instability of symptoms in recurrent major depression: A prospective study. *American Journal of Psychiatry*, *161*(2), 255–261. https://doi.org/10.1176/appi.ajp.161.2.255

Patterson, S. M., Krantz, D. S., Montgomery, L. C., Deuster, P. A., Hedges, S. M., & Nebel, L. E. (1993). Automated physical activity monitoring: Validation and comparison with physiological and self-report measures. *Psychophysiology*, *30*(3), 296–305. https://doi.org/10.1111/j.1469-8986.1993.tb03356.x

Peis, I., López-Moríñigo, J.-D., Pérez-Rodríguez, M. M., Barrigón, M.-L., Ruiz-Gómez, M., Artés-Rodríguez, A., & Baca-García, E. (2020). Actigraphic recording of motor activity in depressed inpatients: A novel computational approach to prediction of clinical course and hospital discharge. *Scientific Reports*, *10*(1), Article 17286. https://doi.org/10.1038/s41598-020-74425-x

Price, G. D., Heinz, M. V., Collins, A. C., & Jacobson, N. C. (2024). Detecting major depressive disorder presence using passively-collected wearable movement data in a nationally-representative sample. *Psychiatry Research*, *332*, Article 115693. https://doi.org/10.1016/j.psychres.2023.115693

Rahman, S. A., & Adjeroh, D. A. (2019). Deep learning using convolutional LSTM estimates biological age from physical activity. *Scientific Reports*, *9*(1), Article 11425. https://doi.org/10.1038/s41598-019-46850-0

Reich, H., Rief, W., Brähler, E., & Mewes, R. (2018). Cross-cultural validation of the German and Turkish versions of the PHQ-9: An IRT approach. *BMC Psychology*, *6*(1), Article 26. https://doi.org/10.1186/s40359-018-0238-z

Reid, R. E. R., Insogna, J. A., Carver, T. E., Comptour, A. M., Bewski, N. A., Sciortino, C., & Andersen, R. E. (2017). Validity and reliability of Fitbit activity monitors compared to ActiGraph GT3X+ with female adults in a free-living environment. *Journal of Science and Medicine in Sport*, *20*(6), 578–582. https://doi.org/10.1016/j.jsams.2016.10.015

Rush, A. J. (2007). The varied clinical presentations of major depressive disorder. *The Journal of Clinical Psychiatry*, *68*(2), Article e04. https://doi.org/10.4088/JCP.0207e04

Rush, A. J., Gullion, C. M., Basco, M. R., Jarrett, R. B., & Trivedi, M. H. (1996). The Inventory of Depressive Symptomatology (IDS): Psychometric properties. *Psychological Medicine*, *26*(3), 477–486. https://doi.org/10.1017/S0033291700035558

Sadeh, A. (2011). The role and validity of actigraphy in sleep medicine: An update. *Sleep Medicine Reviews*, *15*(4), 259–267. https://doi.org/10.1016/j.smrv.2010.10.001

Salgado, J. F. (2018). Transforming the area under the normal curve (AUC) into Cohen's *d*, Pearson's *r* pb, odds-ratio, and natural log odds-ratio: Two conversion tables. *The European Journal of Psychology Applied to Legal Context*, *10*(1), 35–47. https://doi.org/10.5093/ejpalc2018a5

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

TensorFlow Developers. (2023). *Tensorflow* (v2.14.0-rc1) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.8306789

Tlachac, M. L., Reisch, M., & Heinz, M. (2023). *Mobile Communication Log Time Series to Detect Depressive Symptoms*. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, 2023 (pp. 1–4). https://doi.org/10.1109/EMBC40787.2023.10341154

Tsai, L., Boyle, E., Buhl, S. F., Kock, G., Brønd, J. C., Visser, M., Mendonça, N., Shiroma, E. J., & Caserotti, P. (2023). Associations between appetite, physical activity and sedentary behaviour from hip- and wrist-worn accelerometers in community-dwelling older adults. *Geriatrics & Gerontology International*, *23*(6), 411–417. https://doi.org/10.1111/ggi.14588

Uher, R., Perlis, R. H., Henigsberg, N., Zobel, A., Rietschel, M., Mors, O., Hauser, J., Dernovsek, M. Z., Souery, D., Bajs, M., Maier, W., Aitchison, K. J., Farmer, A., & McGuffin, P. (2012). Depression symptom dimensions as predictors of antidepressant treatment outcome: Replicable evidence for interest-activity symptoms. *Psychological Medicine*, *42*(5), 967–980. https://doi.org/10.1017/S0033291711001905

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.

Wang, L., & Miller, L. C. (2020). Just-in-the-Moment Adaptive Interventions (JITAI): A meta-analytical review. *Health Communication*, *35*(12), 1531–1544. https://doi.org/10.1080/10410236.2019.1652388

Wang, Z., & Oates, T. (2015). *Imaging time-series to improve classification and imputation* (arXiv:1506.00327). arXiv. https://arxiv.org/abs/1506.00327

Weissman, M. M. (1986). Understanding the clinical heterogeneity of major depression using family data. *Archives of General Psychiatry*, *43*(5), 430–434. https://doi.org/10.1001/archpsyc.1986.01800050028003

Wichers, M. (2014). The dynamic nature of depression: A new micro-level perspective of mental disorder that meets current challenges. *Psychological Medicine*, *44*(7), 1349–1360. https://doi.org/10.1017/S0033291713001979

Wright, A. G. C., & Woods, W. C. (2020). Personalized models of psychopathology. *Annual Review of Clinical Psychology*, *16*(1), 49–74. https://doi.org/10.1146/annurev-clinpsy-102419-125032

Zhai, L., Zhang, Y., & Zhang, D. (2015). Sedentary behaviour and the risk of depression: A meta-analysis. *British Journal of Sports Medicine*, *49*(11), 705–709. https://doi.org/10.1136/bjsports-2014-093613

Zimmerman, M., Chelminski, I., & McDermut, W. (2002). Major depressive disorder and axis I diagnostic comorbidity. *The Journal of Clinical Psychiatry*, *63*(3), 187–193. https://doi.org/10.4088/JCP.v63n0303

Zimmerman, M., Ellison, W., Young, D., Chelminski, I., & Dalrymple, K. (2015). How many different ways do patients meet the diagnostic criteria for major depressive disorder? *Comprehensive Psychiatry*, *56*, 29–34. https://doi.org/10.1016/j.comppsych.2014.09.007

Received October 18, 2023
Revision received May 16, 2024
Accepted May 20, 2024 ■