

Similarity Matrix-Based Anomaly Detection for Clinical Intervention

Ryan D'Mello

Beth Israel Deaconess Medical Center, Harvard Medical School

Jennifer Melcher

Beth Israel Deaconess Medical Center, Harvard Medical School

John Torous (✉ jtorous@bidmc.harvard.edu)

Beth Israel Deaconess Medical Center, Harvard Medical School

Research Article

Keywords: symptomology, psychiatric

Posted Date: November 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1021257/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Similarity Matrix-Based Anomaly Detection for Clinical Intervention

Ryan D'Mello¹, Jennifer Melcher¹, and John Torous²

¹Department of Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA

²Departments of Psychiatry and Clinical Informatics, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA

*jtorous@bidmc.harvard.edu

ABSTRACT

The use of digital phenotyping methods in clinical care has allowed for improved investigation of spatiotemporal behaviors of patients. Moreover, detecting abnormalities in mobile sensor data patterns can be instrumental in identifying potential changes in symptomology. We propose a method that temporally aligns sensor data in order to achieve interpretable measures of similarity. These computed measures can then be used for anomaly detection, baseline routine computation, and trajectory clustering. In addition, we apply this method on a study of 695 college participants, as well as on a patient with worsening anxiety and depression. With varying temporal constraints, we find mild correlations between changes in routine and clinical scores. Furthermore, in our experiment on an individual with elevated depression and anxiety, we are able to cluster GPS trajectories, allowing for improved understanding and visualization of routines with respect to symptomology. In the future, we aim to apply this method on individuals that undergo data collection for longer periods of time, thus allowing for a better understanding of long-term routines and signals for clinical intervention.

Introduction

Digital phenotyping methods have made investigation of spatial and temporal dynamics of mental illness more feasible¹. Using geolocation and accelerometer data derived from smartphones, it is possible to capture mobility and activity patterns from patients. From simple metrics like home time to complex behaviors related to arousal, circadian rhythms, smartphones offer an emerging facet of multimodal data towards understanding mechanisms of psychiatric conditions².

While preliminary research suggests that this mobility and activity data is associated with clinical phenomenology, limitations in current methods preclude further conclusions. For example, a 2015 paper suggested correlations between GPS derived metrics of normalized entropy and location variance with severity of depressive symptoms³. However, the complexity of the relationship between geolocation data and depression is highlighted in a later study by the same team that controlled for multiple comparisons and found less conclusive results⁴. Studies in schizophrenia also present contradictory results with mobility metrics sometimes associated with worsening psychosis^{5,6} and other times not⁷. We propose that a lack of consensus lies not with any study but rather challenges using this data.

Two core challenges involve first transforming the raw smartphone data into clinical features and second decomposing this data into time series metrics. Most studies use custom rules to create features like sleep duration from raw sensor data which often precludes replication. Even Apple often changes its own feature algorithms which makes replication using Apple Health data more complex⁸. While some features must be constructed, such as movement bouts from accelerometer raw data, transforming such data into more advanced metrics like running, sitting, and standing often involve further assumptions. The second challenge of decomposing the series nature of this data into temporally aligned components is critical for utilizing the time series data inherent to smartphone data capture. But such data is often not utilized today, with many papers instead using summary metrics or pre/post changes as data endpoints⁹. While even this data is still valuable, it ignores and loses information of clinical relevance.

Thus, there is a clear need for methods that can 1) utilize raw smartphone data or at least minimally processed features and 2) account for the temporal nature of the data. In this paper, we introduce such a method with the goal of identifying longitudinal anomalies in smartphone sensor data, which may identify clinically relevant "intervals of interest" or change points. The goal of this method is to offer a privacy-persevering model that can work with smartphone raw data from any app (or sensor) and generate a set of time intervals that a patient and clinician can review to better understand how behaviors and actions captured though digital phenotyping may relate to self-reported and clinically-observed symptoms.

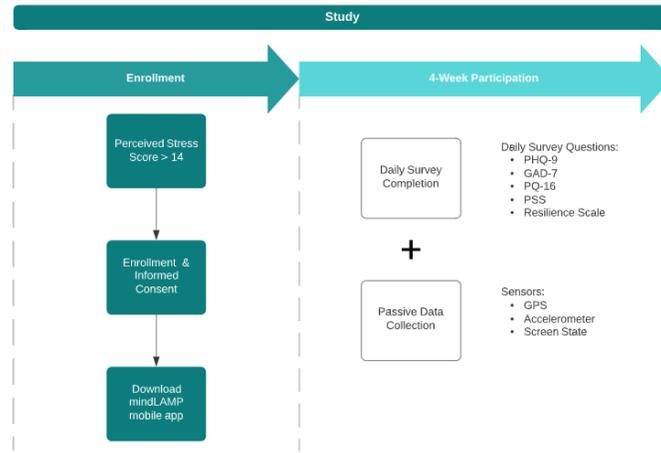


Figure 1. College Study (n = 695)

Table 1. Daily Survey Questions; Response Options: 4-point Likert Scale

Clinical Target	Question
PHQ-9	I felt down today
PHQ-9	Today I feel little interest or pleasure
GAD-7	I felt anxious today
GAD-7	Today I have trouble relaxing
Disorganized (prodromal)	Today my thoughts felt jumbled and confused
Hallucination (prodromal)	Today I heard voices that others apparently couldn't hear
Stress	I felt stressed today
Functioning (prodromal)	I was able to function well today
Resilience	Today I could handle what came my way
Paranoia (prodromal)	Today I felt that others have it in for me
Delusions (prodromal)	I felt that others could read my thoughts today

Results

We label our method "Temporally-aligned Similarity" (TAS) for reference. In Figure 1, we present the outline of our college study in which 695 college students participate. In addition, Table 1 contains all clinical surveys administered daily through the mobile application. Out of the 695 participant in the study, we exclude 220 for having insufficient sensor data. For each participant in the study, we compute TAS scores for the following sensor data: GPS, accelerometer, and screen state. The preprocessing steps for each sensor is different, due to their varying dimensions and domain values. For GPS, we use latitude-longitude pairs without the need of featurizing the raw data. Preprocessing for the accelerometer sensor data involves computing a Jerk value for all points excluding the first point recorded. By performing this transform, we alleviate the problem of dealing with component-wise values and prioritize movements of the mobile device instead. For screen state data, we first determine the bout time spent on and off the phone. Hence, for each "Screen On" recording, we have a time-delta value indicating the amount of time since the mobile screen was previously turned off (either manually or by timeout). We now have screen state trajectories in which we can apply similarity measures, thus allowing for improved spatiotemporal analysis.

For each sensor, we apply three different similarity metrics in order to retrieve similarity scores for the given day, which is partitioned based on a temporal rule and resolution. This approach is detailed in the Methods section. We aim to determine whether there is an association between trajectory similarity scores and survey responses. In Table 2, we have Pearson correlation coefficients of Sensor Trajectory vs. Survey for each similarity measure. We observe a mild positive association between GPS and GAD-7 (Anxiety), measured by DTW ($\rho = 0.1887$, p-value = .0359).

Case Example

In a study on individuals with high anxiety and depression, we invoke this method retrospectively in order to measure the method's ability to detect changes in clinical targets using only the passive sensor data. Patients in the study attended weekly visits with a clinician, completing PHQ-9 and GAD-7 assessments. In addition to completing weekly clinical assessments,

Similarity	Survey	Sensor		
		GPS	Accelerometer	Screen State
DTW	PHQ-9	0.1326	0.0582	0.1049
	GAD-7	0.1887	0.0851	0.1481
	PSS	0.0921	0.0486	0.0984
	Resilience	0.1128	0.0595	0.0752

Table 2. $\tau = 1$ (Hourly Resolution), $R =$ "day of week"; Pearson Correlation Coefficients

patients also completed daily surveys in the MindLAMP application. Despite this patient showing high levels of engagement in both therapy sessions and app usage, their anxiety and depression symptoms continued to worsen. Utilizing the passive data to search for any patterns or relationships expands the scope of clinical work. Specifically, passive data indicated a change in routine for this patient, which was later confirmed to be a result of taking a medical leave from work. In Figure 3, we observe higher daily anxiety and mood scores on many days with elevated GPS similarity scores (days with darker cells). In particular, the participant registered scores of 17 and 13 on the weekly PHQ-9 and GAD-7, respectively (Tuesday, August 3). These were the highest scores registered in the participant’s 8-week study period. In addition to the participant’s weekly scores, their daily mood and anxiety scores on this day (Tuesday, August 3) were the highest in the 8-week period.

In addition to this participant’s daily similarity scores, we also compute an hourly score that measures the periods of the day that are most dissimilar for a given participant. In Figure 4, we have higher dissimilarity during the late afternoon to early evening hours. Based on the GPS data alone, we now have an idea of the mobility routines for this participant, and how mobility patterns may vary across a daily span. Using this hourly distribution as a means of measuring entropy in a participant’s sensor data, clinicians can schedule push notifications at specified times in the day. This allows for improved clinical engagement during periods of the day in which the participant may be more or less vulnerable.

In Figure 2, we perform hierarchical clustering to obtain a dendrogram for average linkage using DTW similarity. This allows us to group mobility trajectories together with a better idea of how each trajectory compares to other trajectories on other days at given periods of the day.

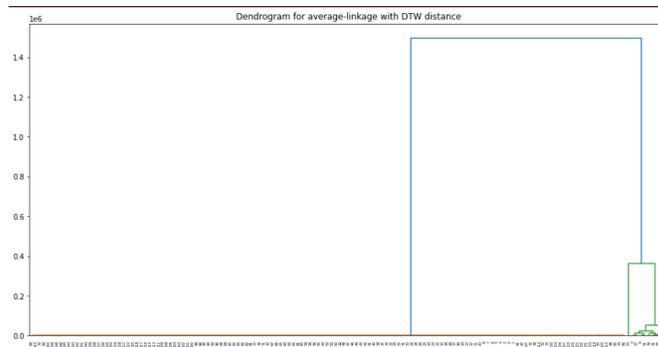


Figure 2. Dendrogram for DTW-based linkage

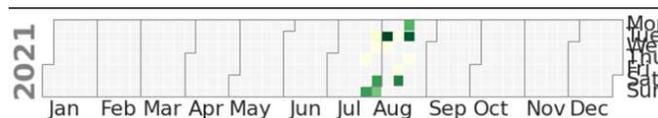


Figure 3. Calendar Heatplot for Participant with elevated anxiety and depression

Discussion

In this analysis, we explored a method of aligning time series data captured from personal smartphones to detect abnormalities in behavior related to mental illness. Our results demonstrate positive associations between changes in mobility and clinical targets, as well as clinically actionable insights seen in the case report, highlighting the potential of this methodology to advance both research and clinical care. Through creating personal indices for an individual’s routines, the ability to utilize hierarchical

clustering allows for flexible analysis that can be customized to the temporal constraint and resolution, ultimately deriving actionable insights from digital phenotyping smartphone data.

Moreover, the temporal alignment and hierarchical clustering that underlie our method can provide clinically interpretable results that are customized to the temporal resolution and sensor(s) of interest. The case report shows we were able to identify changes in mobility patterns at a certain time of day that offered an important target for the therapist to discuss with the patient. Critically, the privacy-persevering nature of this method did not generate the actual location of interest but offered enough detail in terms of hours of the day to allow for the patient and therapist to identify a trigger and create a plan to minimize it. With further validation, this same methodology could also support automated smartphone interventions, often referred to as just-in-time adaptive interventions, although this is beyond the scope of the present paper.

While the need to use digital phenotyping methods for personalized care is well established, methods that enable individual results that account for the temporal nature of this data have been more limited. Our team's prior work has used digital phenotyping to identify relapse risk in patients with psychosis². Unlike that method and others which utilized hand-engineered features from raw data (i.e. transforming GPS into metrics like home time or distance traveled per day), our method is able to utilize that raw data itself, hence avoiding any potential bias related to featurization. This has an additional advantage of working with any smartphone data regardless of the app, and thus is a result that can be adopted and replicated by others.

Our results are in line with those derived from more traditional methods. For example a recent paper from Verily¹⁰ found that smartphone mobility metrics were also correlated with mood. Contrarily, another paper¹¹ with a sample of 255 participants found that there were no significant associations between changes in symptom severity measures and subsequent changes in sensor-derived behavioral features. Based on these different findings in the literature, it is of great importance that we develop personalized model for detecting exacerbation of symptoms through abnormalities in sensor data routines. While we may find features across a population that are significantly associated with symptom severity measures across a sample, these same features may not be generalizable to the population. Thus, the importance of interpretable features, such as our TAS score, can be used to provide improved context to clinicians, along with better visualization of baseline routines and clusters.

Our method does have limitations with regards to the amount of data needed for a given participant. In relying on raw smartphone data, we are limited in analysis on days where less sensor data is collected. We did not perform any interpolation in this analysis, since this could lead to inaccurately predicting an individual's routine. Instead, we can propose that such missing phone data could be itself utilized as a unique signal that may inform engagement with care. While many digital health app-based platforms do suffer from low engagement, our team has demonstrated that with proper support through digital navigator coaching, patients are willing and able to collect high-density sensor data. Another challenge in this work is establishing a baseline for comparison and our use of barycenter computation offers a solution that can establish for personal and population baselines. In summary, methods like those outlined in this paper can help transform digital phenotyping data into actionable targets for both research and care. Through sharing this method and the mobile application used to collect this data as open source code, we hope others will build upon and further expand this work.

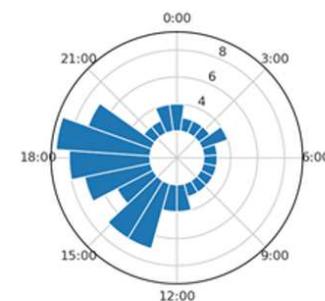


Figure 4. Patient: Hourly TAS score distribution

Methods

Ethics Declaration Statement

The implemented methods and analysis performed were in compliance with pertinent guidelines and protocol. The study was approved by Institutional Review Board at Beth Israel Deaconess Medical Center. Participants were required be at least 18 years old, be fluent in English, own a smartphone, and have a score of at least 14 on the Perceived Stress Scale. Participants

also provided a college email address and student ID card to confirm that they were students prior to completing the informed consent.

Data Collection

GPS, accelerometer, and screen state data were collected from 612 college students attending colleges in the United States. 475 of the 612 students had data that met the requirements for analysis. Data was collected from the students' mobile phones over a 4-week period between November 2020 and May 2021.

Clinical Measures

In addition to passive data collection, students also completed clinical assessments. Upon enrollment, students completed the Perceived Stress Scale. In our analysis, we use the PHQ-9, GAD-7, PSS, and Resilience surveys, which were completed daily on the mindLAMP application. On a bi-weekly basis, students completed an assessment containing full-length versions of the PHQ-9, GAD-7, Perceived Stress Scale (PSS), UCLA Loneliness scale, Prodromal Questionnaire-16 (PQ-16), Pittsburgh Sleep Quality Index (PSQI), and the Digital Working Alliance Inventory (D-WAI). These survey results aren't considered in our analysis due to lower frequency.

Anomaly Detection

We employ a unique approach to identifying anomalous trajectories in the sensor data. First, we aim to group sensor data by some time constraint (e.g. day of the week, weekends vs. weekdays, etc.). The next step involves partitioning each day's sensor data into intervals of equal length, with the purpose of detecting minute changes in patient data within a day (e.g. hourly).

Temporal Alignment of Sensor Data

In order to perform time series analysis on the sensor data, we need to temporally align the data with respect to a given temporal rule, with a specified time resolution.. GPS data contains Latitude, Longitude, and Altitude fields. For our time series representation we will only use Latitude and Longitude pairs.

For GPS data, we perform the following steps:

1. Retrieve the raw GPS data
2. Downsample GPS data to user-specified frequency. For example, a frequency of 30 seconds would return split up raw data into 30 second intervals, returning the mode of all (Lat,Lon) tuples in each interval.
3. Partition the raw data by user-specified temporal constraint
4. Break series into subseries with length equal to user-specified resolution

For accelerometer data, we aim to use an input representation other than the raw data itself. Moreover, we obtain a measure of jerk (i.e. rate of change in accelerometer magnitude). We now have a more interpretable input representation with Jerk, allowing us to observe periods of activity much more easily than the raw representation. We complete steps similar to GPS above for both accelerometer (jerk) and screen state.

Method Overview

For a given participant undergoing passive data collection, consider a set S of k sensors ($Sensor_1, Sensor_2, \dots, Sensor_k$), each sensor potentially collecting data at different points in time. It is worth noting that the dimensionality and range of values for each sensor's data is intrinsically different.

For a given sensor in S , define a time series T to be a sequence of data points (t_1, t_2, \dots, t_n) recorded over a period of time from the sensor, with n being the number of points observed. When necessary, assume T to have already undergone preprocessing steps.

Consider a temporal rule R , which will be used to create partitions of T based on a mutual time characteristic, such as day-of-week or weekday vs. weekend. We aim to partition T into groups of subseries, with each subseries sharing a common "characteristic" with all other subseries in respective groups. These characteristics range from "day of week" to "weekday vs. weekend", with all characteristics being formally specified as a temporal rule R . Furthermore, after applying the rule R to T , we will have a list L of subseries groups, each group not necessarily having same length as other groups. Also, for each group G in L , we have subseries that span a user-specified time resolution τ with $0 < \tau < 24$ being the number of hours. For each group G in L , we compute a matrix that measures the pairwise similarity between all subseries in G . We accomplish this by applying some similarity metric to each distinct pair of subseries. Let l be the length of G . We will have a $l \times l$ symmetric matrix, where the cell $l[i, j]$ corresponds to the pairwise similarity between G_i and G_j . Once all similarity measures are computed, we can

then determine the columns of the matrix with highest dissimilarity, thus signaling that an element of G deviates from the other elements in the group.

In addition to anomaly detection, we can perform hierarchical clustering and use different approaches^{12,13} to compute baseline trajectories. The outline for our method is outlined in Figure 5.

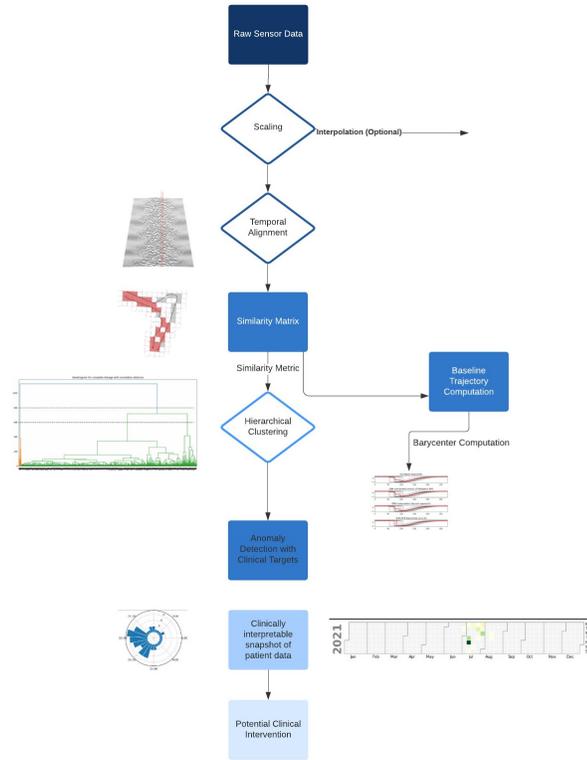


Figure 5. Legend (350 words max). Example legend text.

Similarity Measures

Dynamic Time Warping (DTW)

Dynamic Time Warping (DTW) is an algorithm¹⁴ for computing the distance between two series of points. It has become a popular approach for measuring similarity in GPS trajectories, given the temporal nature of the data. We formulate DTW distance as follows: Given two time series of GPS points, A and B , with lengths m and n ,

$$A = a_1, a_2, \dots, a_m$$

$$B = b_1, b_2, \dots, b_n$$

We define a warp path P :

$$P = p_1, p_2, \dots, p_K$$

where

$$\max(m, n) \leq K < m + n \tag{1}$$

Moreover, $K = |P|$ and the k^{th} element of P is $p_k = (i, j)$, where i and j are indices of A and B , respectively.

Consider the warp path cost matrix D , with the minimum warp distance at index i of A and j of B :

$$D(i, j) = \min[(D(i-1, j), D(i, j-1), D(i-1, j-1))] + d_{euclidean}(i, j) \tag{2}$$

We then have our total cost of D :

$$D_{DTW}(P) = \sum_{k=1}^K d_{euclidean}(i_k, j_k)$$

Longest Common Subsequence (LCSS)

Similar to DTW, Longest Common Subsequence (LCSS)¹⁵ generates a matrix, but instead uses similarity between elements instead of distance. In addition, LCSS does not require all elements to be matched, unlike DTW.

Consider two time series A and B . Let m and n be the lengths of A and B , respectively. We can now generate an LCSS matrix L using the following recurrence:

$$L(i, j) = \begin{cases} 0 & i = 0 \\ 0 & j = 0 \\ 1 + L[i - 1, j - 1] & A_i = B_j \\ \max(L[i - 1, j], L[i, j - 1]) & \text{otherwise} \end{cases}$$

with $1 \leq i \leq n$ and $1 \leq j \leq m$. This recurrence only allows for exact matching between A and B for numerical values. Furthermore, the time complexity of this approach is $\mathcal{O}(mn)$. To allow for non-exact matching between elements, we can add a specified ε argument. We can then use the new recurrence:

$$L(i, j) = \begin{cases} 0 & i = 0 \\ 0 & j = 0 \\ 1 + L[i - 1, j - 1] & |A_i - B_j| < \varepsilon \\ \max(L[i - 1, j], L[i, j - 1]) & \text{otherwise} \end{cases}$$

To check similarity between A and B , we retrieve the value z in cell $L[n, m]$, which would correspond to the length of the longest subsequence between A and B . Moreover, a distance metric¹⁶ can be defined:

$$Dist(A, B) = \frac{m+n+2z}{m+n}$$

Data Availability

The data is not available since it contains identifiable information of participants in the study.

References

1. Torous, J., Kiang, M., Lorme, J. & Onnela, J. P. New tools for new research in psychiatry: A scalable and customizable platform to empower data driven smartphone research. *JMIR mental health* **3**, DOI: <https://doi.org/10.2196/mental.5165> (2016).
2. Henson, P., D’Mello, R., Vaidyam, A. *et al.* Anomaly detection to predict relapse risk in schizophrenia. *Transl Psychiatry* **11**, DOI: <https://doi.org/10.1038/s41398-020-01123-7> (2021).
3. Saeb, S. *et al.* Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study. *J Med Internet Res* **17**, DOI: [10.2196/jmir.4273](https://doi.org/10.2196/jmir.4273) (2015).
4. Saeb, S., Lattie, E., Kording, K. & Mohr, D. Mobile phone detection of semantic location and its relationship to depression and anxiety. *JMIR Mhealth Uhealth* **5**, DOI: [10.2196/mhealth.7297](https://doi.org/10.2196/mhealth.7297) (2017).
5. Wang, R. *et al.* Crosscheck: Toward passive sensing and detection of mental health changes in people with schizophrenia. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp ’16, 886–897, DOI: [10.1145/2971648.2971740](https://doi.org/10.1145/2971648.2971740) (Association for Computing Machinery, New York, NY, USA, 2016).
6. Depp, C., Bashem, J., Moore, R.C. *et al.* Gps mobility as a digital biomarker of negative symptoms in schizophrenia: a case control study. *npj Digit. Med* **2**, DOI: <https://doi.org/10.1038/s41746-019-0182-1> (2019).
7. Fulford, D. *et al.* Smartphone sensing of social interactions in people with and without schizophrenia. *J. Psychiatr. Res.* **137**, 613–620, DOI: <https://doi.org/10.1016/j.jpsychires.2020.11.002> (2021).
8. Fuller, D. *et al.* Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: Systematic review. *JMIR Mhealth Uhealth* **8**, e18694, DOI: [10.2196/18694](https://doi.org/10.2196/18694) (2020).

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, DOI: <https://doi.org/10.1002/wps.20883> (2021). <https://onlinelibrary.wiley.com/doi/pdf/10.1002/wps.20883>.
10. Nickels, S. *et al.* Toward a mobile platform for real-world digital measurement of depression: User-centered design, data quality, and behavioral and clinical modeling. *JMIR Ment Heal.* **8**, e27589, DOI: [10.2196/27589](https://doi.org/10.2196/27589) (2021).
11. 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, DOI: [10.2196/22844](https://doi.org/10.2196/22844) (2021).
12. Marteau, P.-F. Estimating road segments using kernelized averaging of gps trajectories. *Appl. Sci.* **9**, DOI: [10.3390/app9132736](https://doi.org/10.3390/app9132736) (2019).
13. Petitjean, F., Ketterlin, A. & Gançarski, P. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recogn.* **44**, 678–693, DOI: [10.1016/j.patcog.2010.09.013](https://doi.org/10.1016/j.patcog.2010.09.013) (2011).
14. Berndt, D. J. & Clifford, J. Using dynamic time warping to find patterns in time series. In *KDD Workshop* (1994).
15. Vlachos, M., Gunopulos, D. & Kollios, G. Robust similarity measures for mobile object trajectories. In *Proceedings. 13th International Workshop on Database and Expert Systems Applications*, 721–726, DOI: [10.1109/DEXA.2002.1045983](https://doi.org/10.1109/DEXA.2002.1045983) (2002).
16. Cassisi, C., Montalto, P., Aliotta, M., Cannata, A. & Pulvirenti, A. Similarity measures and dimensionality reduction techniques for time series data mining. In Karahoca, A. (ed.) *Advances in Data Mining Knowledge Discovery and Applications*, chap. 3, DOI: [10.5772/49941](https://doi.org/10.5772/49941) (IntechOpen, Rijeka, 2012).

Competing Interests

There are no competing interests of the authors.

Author contributions statement

R.D. composed the methodology, conducted data analysis, and prepared visualization. J.M. outlined and organized the college study. R.D. and J.T. wrote the manuscript text.

Additional information

Code may be found on Github: https://github.com/BIDMCDigitalPsychiatry/LAMP-cortex/tree/dense_anomaly_retrieval