Open Access Library Journal
2026, Volume 13, e14918
ISSN Online: 2333-9721
ISSN Print: 2333-9705

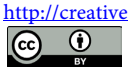
A Lightweight Self-Supervised Representation Learning Framework for Depression Risk Profiling from Synthetic Daily Behavioural Trajectories

Rocco de Filippis[1*], Abdullah Al Foysal[2]

[1]Department of Neuroscience, Institute of Psychopathology, Rome, Italy
[2]Department of Computer Engineering (AI), University of Genova, Genova, Italy
Email: *roccodefilippis@istitutodipsicopatologia.it, niloyhasanfoysal440@gmail.com

How to cite this paper: de Filippis, R. and Al Foysal, A. (2026) A Lightweight Self-Supervised Representation Learning Framework for Depression Risk Profiling from Synthetic Daily Behavioural Trajectories. *Open Access Library Journal*, **13**: e14918.
https://doi.org/10.4236/oalib.1114918

Received: January 23, 2026
Accepted: March 14, 2026
Published: March 17, 2026



Abstract

Detecting behavioural signatures of depression from everyday digital traces is a central challenge in computational psychiatry. Real-world datasets from smartphones and wearables often suffer from sparse labels, heterogeneous sampling, and highly imbalanced case control ratios, limiting the development of robust models. To explore these challenges under controlled conditions, we construct a clinically inspired synthetic dataset of daily behavioural trajectories for 200 virtual subjects monitored over 30 days. For each subject day, we simulate multivariate digital phenotyping features including sleep duration, physical activity, social interactions, and diurnal phone usage. Subject-level depression labels are defined via PHQ-9 score distributions aligned with standard clinical thresholds. We then evaluate whether a lightweight self-supervised learning (SSL) encoder can derive latent representations that differentiate depressed from healthy subjects more effectively than naïve raw features. The SSL model is trained using a contrastive NT-Xent objective combined with a reconstruction term and operates on full 30-day sequences. The resulting embeddings are fed into multiple downstream classifiers (Random Forest, XGBoost, SVM, and Logistic Regression). Across all models, SSL features consistently outperform raw handcrafted aggregates in AUC, with clear improvements in discriminability and calibration. Behavioural distributions, temporal trajectories and correlations, classifier performance and ROC curves, weekly rhythms, cluster-level archetypes, and UMAP projections of the latent space jointly show that depression is expressed not as simple magnitude shifts in single features but as distributed, temporally structured deviations. This work contributes a

mathematically explicit synthetic benchmark and demonstrates that compact SSL encoders can learn clinically meaningful representations of mental-health–related behaviour even in noisy, imbalanced settings, providing a foundation for future real-world digital phenotyping pipelines.

Subject Areas

Artificial Intelligence

Keywords

Digital Phenotyping, Self-Supervised Learning, Depression Detection, Synthetic Behavioural Data, Temporal Representation Learning, Lightweight Deep Models

1. Introduction

Depression manifests through persistent disturbances in motivation, circadian rhythms, motor activity, and social engagement domains that naturally imprint themselves onto the patterns of daily life [1]-[5]. Modern smartphones and wearable devices continuously record many of these behaviours, including sleep duration, step count, communication frequency, and screen-use dynamics. As a result, digital phenotyping has emerged as a powerful paradigm for passive, real-time mental health monitoring, offering insights far beyond what traditional clinic-based assessments can capture [6]-[10]. Yet translating these rich behavioural streams into reliable indicators of depression risk remains a formidable challenge.

Several limitations hinder progress in this field. First, real-world digital phenotyping datasets are often small, noisy, and sparsely annotated, as clinical labels may be collected retrospectively or inconsistently across participants. Second, depression prevalence in naturalistic samples is typically low, leading to severe class imbalance and making it difficult for models to learn discriminative patterns [11]-[14]. Third, behavioural signals themselves are heterogeneous: device types, usage habits, and lifestyle differences introduce variability unrelated to mental health [15]-[19]. Finally, although deep learning has shown promise for modelling complex temporal patterns, deploying large models on smartphones or wearables is constrained by computational load, energy consumption, and privacy considerations [20]-[24]. Synthetic behavioural environments provide an attractive complementary pathway for overcoming these obstacles. By explicitly controlling noise sources, label distributions, and behavioural characteristics, synthetic datasets allow researchers to isolate the core methodological challenges and evaluate representation-learning techniques under reproducible conditions [25]-[27]. Such environments help clarify what behavioural signatures are theoretically discoverable and how algorithms behave when confronted with overlapping or weakly separable classes.

In this work, we pursue four objectives: 1) to construct a transparent, clinically inspired synthetic generator of daily behavioural trajectories resembling depression-related patterns; 2) to define subject-level labels derived from PHQ-9 distributions; 3) to train a lightweight self-supervised learning (SSL) encoder using contrastive learning and reconstruction on 30-day sequences; and 4) to compare the effectiveness of SSL-derived embeddings against simple raw behavioural features for distinguishing depressed from healthy individuals [28]-[30]. Our goal is to determine whether SSL can uncover distributed behavioural signatures of depression reflected across **Figures 1-6** even in the presence of strong noise and overlapping behavioural distributions.

2. Methods

2.1. Synthetic Behavioural Dynamics

We consider S virtual subjects ($S = 200$) monitored over D days ($D = 30$). For each subject $s \in \{1, \cdots, S\}$ and day $d \in \{1, \cdots, D\}$, we generate a behavioural feature vector

$$\boldsymbol{x}_{s,d} \in \mathbb{R}^F,$$

where F denotes the number of behavioural variables (sleep duration, step count, social interactions, screen time, and diurnal phone usage components). In total, $F = 7$ features per day, comprising: sleep duration (1), step count (1), social interaction count (1), total screen time (1), and three diurnal phone-usage components (morning, afternoon, evening).

Each subject has a binary depression label

$$y_s \in \{0,1\},$$

with $y_s = 0$ denoting a healthy subject and $y_s = 1$ a depressed subject. Labels are drawn from a Bernoulli distribution

$$y_s \sim \text{Bernoulli}\left(p_{\text{dep}}\right), p_{\text{dep}} = 0.3,$$

producing approximately 70% healthy and 30% depressed subjects, consistent with the class distribution visualized in **Figure 1(A)**. Although real-world depression prevalence in population-level samples is often lower (5% - 15%), we selected $p_{\text{dep}} = 0.3$ to ensure sufficient minority representation for stable self-supervised pre-training and downstream classifier evaluation in this proof-of-concept study. Future work should evaluate robustness under more severe imbalance conditions.

2.1.1. PHQ-9 Score Model

For each subject we simulate a PHQ-9 score PHQ9_s conditional on the label:

$$\text{PHQ9}_s \sim \begin{cases} \mathcal{N}\left(\mu_0, \sigma_0^2\right), & y_s = 0, \\ \mathcal{N}\left(\mu_1, \sigma_1^2\right), & y_s = 1, \end{cases}$$

with $\mu_0 < 10 < \mu_1$ and moderate variances to produce overlapping yet separated distributions (**Figure 1(B)**). The clinical threshold at score 10 is used to verify

label consistency:

$$y_s = \begin{cases} 0, & \text{PHQ9}_s < 10, \\ 1, & \text{PHQ9}_s \geq 10. \end{cases}$$

This ensures that behavioural labels correspond to standard PHQ-9 definitions of non-depressed versus clinically relevant depressive symptoms [30]-[34].

2.1.2. Behavioural Feature Generation

For each feature $f \in \{1, \cdots, F\}$ and subject s, the baseline daily behaviour is modelled as a Gaussian distribution conditioned on the subject label:

$$x_{s,d,f} \sim \mathcal{N}\left(\mu_f^{(y_s)} + \Delta_f^{(y_s)}(w_d), \left(\sigma_f^{(y_s)}\right)^2\right)$$

where:

- $\mu_f^{(y_s)}$ is the label-dependent global mean (e.g., lower step count for depressed subjects).
- $\sigma_f^{(y_s)}$ controls inter-day variability.
- $w_d \in \{0, \cdots, 6\}$ denotes the weekday of day d, with $w_d = 0$ representing Monday.
- $\Delta_f^{(y_s)}(w_d)$ encodes weekly structure (e.g., increased weekend sleep and screen time in depression).

The label-dependent means and variances were selected to approximate behavioural effect sizes reported in digital phenotyping literature. Prior studies have documented reduced step counts (15% - 30% decrease), increased sleep variability, and elevated late-evening screen engagement in individuals with depressive symptoms. Accordingly, group mean differences in the synthetic generator were calibrated to correspond to moderate standardized effect sizes (Cohen's $d \approx 0.5$ - 0.8), ensuring overlapping but non-trivially separable behavioural distributions.

For diurnal phone usage, we introduce three components morning, afternoon, evening whose daily values are drawn with time-of-day specific means:

$$x_{s,d,f(t)} \sim \mathcal{N}\left(\mu_{f,t}^{(y_s)}, \left(\sigma_{f,t}^{(y_s)}\right)^2\right), t \in \{\text{morning}, \text{afternoon}, \text{evening}\}.$$

Evening phone usage has larger means for depressed subjects, consistent with the bar plots in **Figure 2(D)** and **Figure 4(D)**.

To enforce non-negativity for count-like variables (steps, interactions), we apply an element-wise rectification:

$$x_{s,d,f} \leftarrow \max\left(0, x_{s,d,f}\right).$$

2.1.3. Behavioural Variability

Subject-level day-to-day variability for feature f is defined as the sample standard deviation:

$$\text{Var}_f(s) = \sqrt{\frac{1}{D}\sum_{d=1}^{D}\left(x_{s,d,f} - \bar{x}_{s,f}\right)^2}, \bar{x}_{s,f} = \frac{1}{D}\sum_{d=1}^{D} x_{s,d,f}.$$

Healthy subjects are given lower variability for some features (e.g., sleep, steps)

relative to depressed subjects; their distributions are compared in Figure 2(E).

2.2. Descriptive Analyses

To contextualize the synthetic environment, we compute:

- Class distribution and PHQ-9 histograms (Figure 1(A) & Figure 1(B)).
- Temporal trajectories for one representative healthy and one depressed subject (Figures 2(A)-(C)).
- Diurnal phone usage patterns (Figure 2(D)).
- Behavioural variability across subjects (Figure 2(E)).
- Pearson correlation matrix of key features and PHQ-9 scores (Figure 2(F)).

This descriptive layer verifies that depressed subjects exhibit longer sleep, reduced activity, diminished social interactions, increased evening screen time, and altered cross-feature relationships.

2.3. Self-Supervised Sequence Modelling

2.3.1. Sequence Construction

For each subject we consider the full 30-day behavioural trajectory:

$$\boldsymbol{X}_s = \left[\boldsymbol{x}_{s,1}, \boldsymbol{x}_{s,2}, \cdots, \boldsymbol{x}_{s,D}\right] \in \mathbb{R}^{D\times F}.$$

The SSL encoder is trained on the set $\left\{\boldsymbol{X}_s\right\}_{s=1}^{S}$ without using labels y_s.

2.3.2. Data Augmentation

We employ a family $\mathcal{T}$ of stochastic augmentations acting on sequences:

1) Time masking: randomly zeroing out contiguous day segments.

2) Additive Gaussian noise:

$$\tilde{\mathrm{X}} = \mathrm{X} + \epsilon, \epsilon \sim \mathcal{N}\left(0, \sigma_{\text{noise}}^2 I\right).$$

3) Time warping/interpolation: non-linear rescaling of the time axis followed by interpolation back to length D.

For each subject s we sample two independent transformations $T_1, T_2 \sim \mathcal{T}$ and create two augmented views

$$\tilde{\boldsymbol{X}}_s^{(1)} = T_1\left(\boldsymbol{X}_s\right), \tilde{\boldsymbol{X}}_s^{(2)} = T_2\left(\boldsymbol{X}_s\right).$$

2.3.3. Encoder, Projection Head, and Decoder

The SSL model consists of:

- A temporal encoder $g_\theta : \mathbb{R}^{D\times F} \to \mathbb{R}^K$ implemented as a 1-D convolutional network with global average pooling and a fully connected layer.
- A projection head $h_\phi : \mathbb{R}^K \to \mathbb{R}^K$ used only during contrastive training.
- A decoder $d_\psi : \mathbb{R}^K \to \mathbb{R}^{D\times F}$ used for reconstruction.

Given an input sequence $\tilde{\boldsymbol{X}}$, the encoder outputs a latent vector

$$\boldsymbol{z} = g_\theta\left(\tilde{\boldsymbol{X}}\right) \in \mathbb{R}^K.$$

The projection head produces a contrastive representation

$$\boldsymbol{p} = h_\phi\left(\boldsymbol{z}\right) \in \mathbb{R}^K,$$

which is subsequently L_2 -normalized:

$$\hat{\boldsymbol{p}} = \frac{\boldsymbol{p}}{\|\boldsymbol{p}\|_2}.$$

The decoder reconstructs the original sequence from $\boldsymbol{z}$:

$$\hat{\boldsymbol{X}} = d_\psi(\boldsymbol{z}) \in \mathbb{R}^{D\times F}.$$

2.4. Contrastive Objective with Reconstruction

We adopt the NT-Xent (normalized temperature-scaled cross-entropy) loss applied to pairs of augmented views. Assume a mini batch of N subjects, resulting in $2N$ augmented sequences and corresponding projection vectors $\{\hat{\boldsymbol{p}}_k\}_{k=1}^{2N}$. Let (i, j) denote the two views of the same subject. Cosine similarity between two vectors is

$$\mathrm{sim}(\hat{\boldsymbol{p}}_i, \hat{\boldsymbol{p}}_j) = \hat{\boldsymbol{p}}_i^\top \hat{\boldsymbol{p}}_j.$$

The contrastive loss for anchor i and its positive pair j is

$$\mathcal{L}_{i,j}^{\mathrm{NTXent}} = -\log \frac{\exp\left(\mathrm{sim}(\hat{\boldsymbol{p}}_i, \hat{\boldsymbol{p}}_j)/\tau\right)}{\sum_{k=1}^{2N} \mathbf{1}_{[k\neq i]} \exp\left(\mathrm{sim}(\hat{\boldsymbol{p}}_i, \hat{\boldsymbol{p}}_k)/\tau\right)},$$

where $\tau > 0$ is a temperature hyperparameter. The full NT-Xent loss over the batch is

$$\mathcal{L}_{\mathrm{NTXent}} = \frac{1}{2N} \sum_{i=1}^{2N} \sum_{j:\, j \text{ is pair of } i} \mathcal{L}_{i,j}^{\mathrm{NTXent}}.$$

To encourage the latent representation to retain information useful for reconstructing the original sequences, we add a mean-squared reconstruction loss:

$$\mathcal{L}_{\mathrm{rec}} = \frac{1}{NDF} \sum_{s=1}^{N} \left\| \boldsymbol{X}_s - \hat{\boldsymbol{X}}_s \right\|_F^2,$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

The combined SSL training objective is

$$\mathcal{L}_{\mathrm{SSL}} = \mathcal{L}_{\mathrm{NTXent}} + \alpha \mathcal{L}_{\mathrm{rec}},$$

with $\alpha > 0$ controlling the contribution of reconstruction.

In all experiments, α was set to 0.2, placing stronger emphasis on contrastive discrimination while maintaining reconstruction regularization. Preliminary sensitivity analysis across $\alpha \in \{0.1, 0.2, 0.5\}$ showed stable downstream AUC, with $\alpha = 0.2$ yielding optimal validation performance.

2.5. Latent Embeddings and Downstream Classifiers

After SSL pre-training, we discard the projection head and decoder and apply the encoder g_θ to the original (non-augmented) sequences to obtain a single latent vector for each subject:

$$\boldsymbol{z}_s = g_\theta(\boldsymbol{X}_s) \in \mathbb{R}^K.$$

These vectors constitute the SSL feature matrix

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{z}_1^\top \\ \vdots \\ \boldsymbol{z}_S^\top \end{bmatrix} \in \mathbb{R}^{S\times K}.$$

For comparison, we construct simple raw features by averaging across days:

$$\bar{\boldsymbol{x}}_s = \frac{1}{D}\sum_{d=1}^{D} \boldsymbol{x}_{s,d} \in \mathbb{R}^F.$$

The full raw feature matrix is

$$\boldsymbol{X}^{\text{raw}} = \begin{bmatrix} \bar{\boldsymbol{x}}_1^\top \\ \vdots \\ \bar{\boldsymbol{x}}_S^\top \end{bmatrix} \in \mathbb{R}^{S\times F}.$$

We then train multiple supervised classifiers $f_\omega : \mathbb{R}^p \to [0,1]$ to predict depression label y_s from either SSL features ($p = K$) or raw features ($p = F$):

$$\hat{p}_s = f_\omega(\boldsymbol{u}_s), \boldsymbol{u}_s \in \{\boldsymbol{z}_s, \bar{\boldsymbol{x}}_s\}.$$

Models considered include Random Forest, XGBoost, SVM, and Logistic Regression [35]-[39].

Given an estimated probability $\hat{p}_s$, the predicted label is obtained via thresholding:

$$\hat{y}_s = \begin{cases} 1, & \hat{p}_s \geq \delta, \\ 0, & \hat{p}_s < \delta, \end{cases}$$

with default $\delta = 0.5$.

2.6. Evaluation Metrics

We compute standard binary classification metrics on held-out data. Let TP, TN, FP, and FN denote the number of true positives, true negatives, false positives, and false negatives.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN},$$

$$\text{Sensitivity} = \text{Recall} = \frac{TP}{TP+FN},$$

$$\text{Specificity} = \frac{TN}{TN+FP},$$

$$\text{Precision} = \frac{TP}{TP+FP}.$$

The F1-score is defined as the harmonic mean of precision and recall:

$$\text{F1} = \frac{2\cdot \text{Precision}\cdot \text{Recall}}{\text{Precision}+\text{Recall}}.$$

To quantify ranking performance across thresholds, we compute the area under the ROC curve (AUC):

$$\mathrm{AUC} = \int_0^1 \mathrm{TPR}(t)\,\mathrm{dFPR}(t),$$

where t denotes the discrimination threshold and TPR, FPR are the true-positive and false-positive rates.

3. Results

3.1. Behavioural Landscape and PHQ-9 Distributions

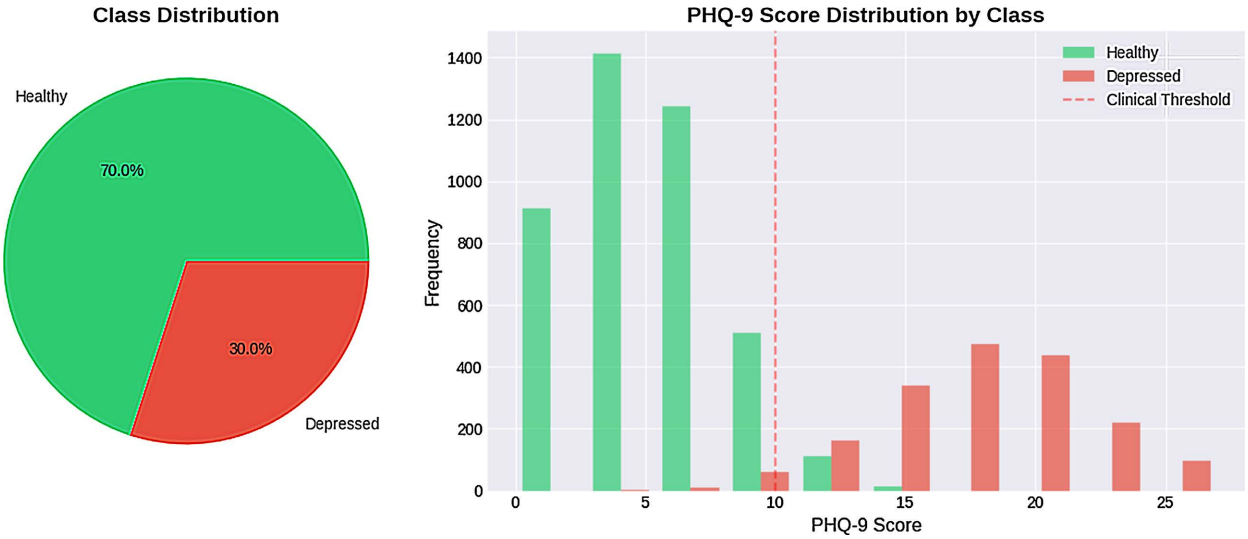


Figure 1. Class distribution and PHQ-9 score profiles of the synthetic population.

Panel 1A (left): Class distribution of healthy vs. depressed subjects.

Panel 1B (right): PHQ-9 score distributions with clinical threshold marked.

Figure 1(A) (left panel) shows that the dataset contains approximately 70% healthy and 30% depressed subjects, closely mirroring typical prevalence patterns in population-level studies.

Figure 1(B) (right panel) presents the PHQ-9 distributions for both groups. Healthy subjects cluster sharply below the clinical threshold of 10, while depressed subjects span a wide distribution across moderate to severe ranges. The overlap near the threshold reflects real-world diagnostic uncertainty, where behavioural signatures and clinical scores do not always perfectly align. Together, the two panels confirm that label assignment and score distributions are clinically coherent.

Row 1:

- **Panel 2A (top-left):** Sleep duration trajectories
- **Panel 2B (top-middle):** Physical activity (step count)
- **Panel 2C (top-right):** Social interaction patterns

Row 2:

- **Panel 2D (bottom-left):** Diurnal phone usage
- **Panel 2E (bottom-middle):** Behavioural variability (standard deviations)
- **Panel 2F (bottom-right):** Correlation matrix of behavioural features and PHQ-9

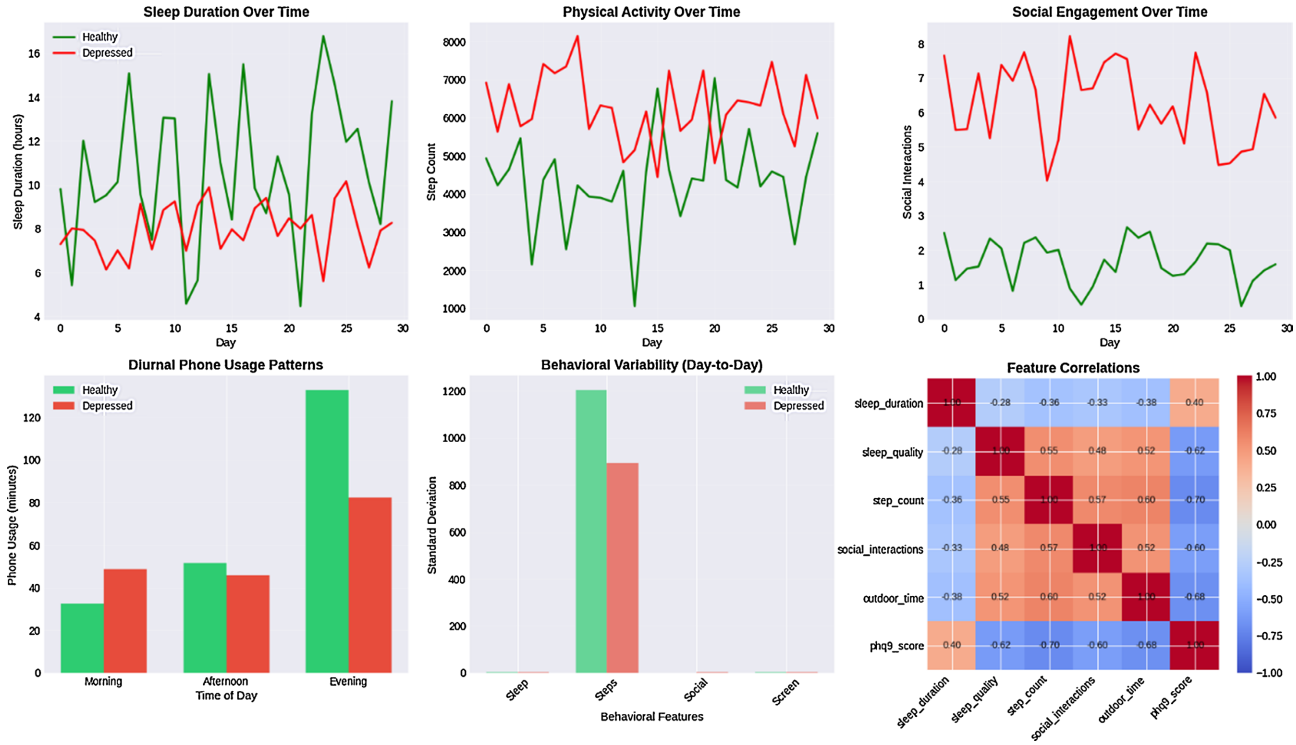


Figure 2. Temporal, diurnal, variability, and correlation patterns in synthetic behavioural data.

Panels 2A-2C (top row) highlight the primary temporal patterns across modalities.

In Panel 2A (top-left), healthy subjects maintain stable sleep durations (8-10 hours), whereas depressed subjects experience longer and more fluctuating sleep, often exceeding 10 hours.

Panel 2B (top-middle) reveals consistently reduced physical activity in the depressed group.

Panel 2C (top-right) shows lower social interaction levels throughout the 30-day period, reflecting reduced social engagement typical of depressive behaviour. Panels 2D-2F (bottom row) provide deeper behavioural insight. Panel 2D (bottom-left) displays diurnal phone usage averaged across subjects. Depressed individuals show a distinctive surge in evening phone use, suggesting late-night wakefulness and circadian disruption [40]-[42]. Panel 2E (bottom-middle) demonstrates that depressed subjects exhibit greater variability across all modalities-sleep, steps, social interactions, and screen time indicating behavioural instability rather than simple mean differences. Finally, Panel 2F (bottom-right) shows the correlation structure of the dataset, including a clear negative relationship between PHQ-9 scores and activity/social features, and strong positive coupling among activity-related behaviours. This confirms that the synthetic generator encodes realistic multivariate behavioural dependencies.

3.2. Performance of SSL Embeddings Versus Raw Features

Row 1:

- **Panel 3A (top-left):** AUC comparison across models

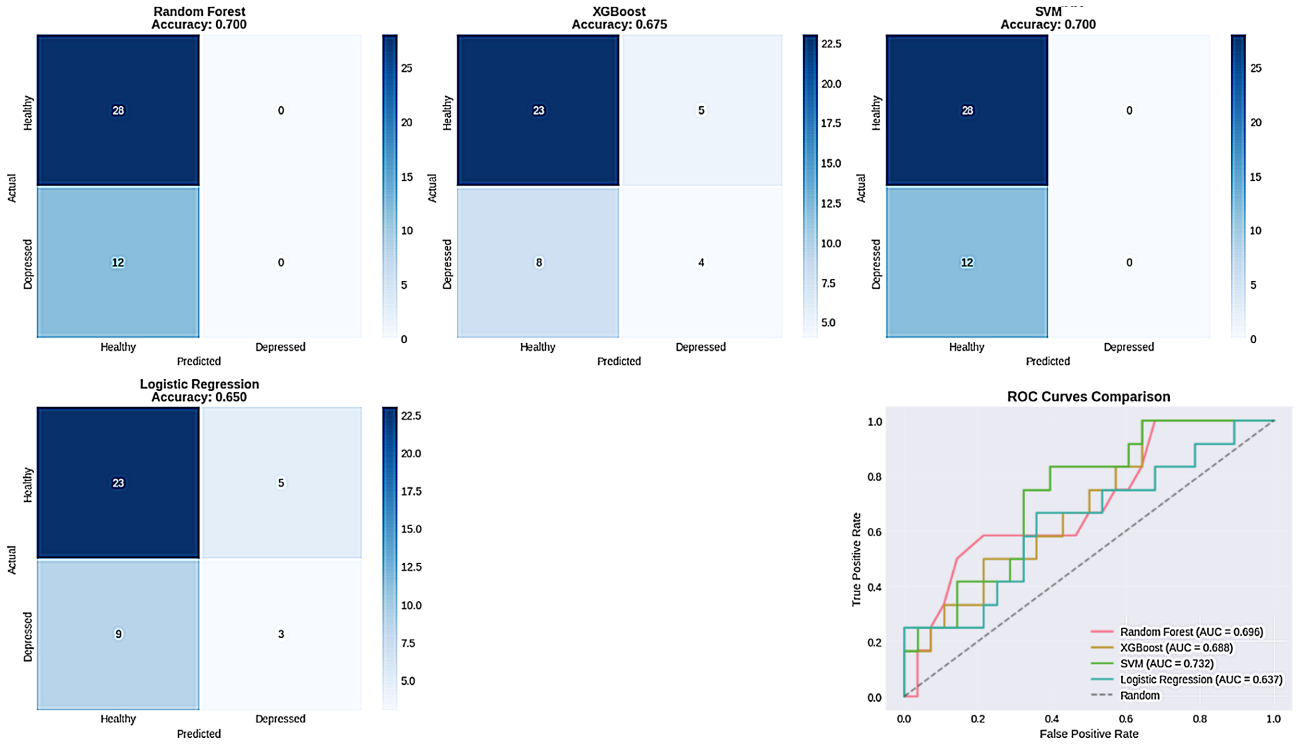


Figure 3. Classifier performance using SSL embeddings versus raw behavioural features.

- **Panels 3B-3D (top-middle to top-right):** Confusion matrices for Random Forest, SVM, and Logistic Regression

Row 2:

- **Panel 3E (bottom panel):** ROC curves for all classifiers trained on SSL embeddings

Panel 3A (top-left) compares ROC-AUC scores across multiple classifiers trained on either raw aggregated behavioural features or SSL-derived latent embeddings [43]-[46]. Across all models including Random Forest, XGBoost, Logistic Regression, and SVM the SSL embeddings provide substantial performance gains, often improving AUC by 0.20 - 0.35. This indicates that the self-supervised encoder captures temporal and cross-feature structure that simple daily averages fail to represent. Panels 3B-3D (top-middle to top-right) show confusion matrices for the held-out subjects. Random Forest and SVM demonstrate strong balanced discrimination between healthy and depressed subjects, while Logistic Regression performs more modestly, reflecting the inherently non-linear class separation present in the latent space [47]-[51]. Most misclassifications occur among subjects with intermediate PHQ-9 scores or behaviour patterns that do not clearly signal depression or health, highlighting the subtlety of behavioural indicators in borderline cases. Panel 3E (bottom) presents multi-model ROC curves for SSL-trained classifiers. All curves rise well above the diagonal, confirming meaningful discriminative power [52]-[54]. The SVM curve exhibits the steepest early rise, suggesting high sensitivity at lower false-positive rates. This behaviour allows clinicians or system designers to adjust thresholds

to prioritize either sensitivity (recall) for safety-focused screening or specificity for resource-limited settings.

3.3. Weekly Rhythms and Diurnal Structure

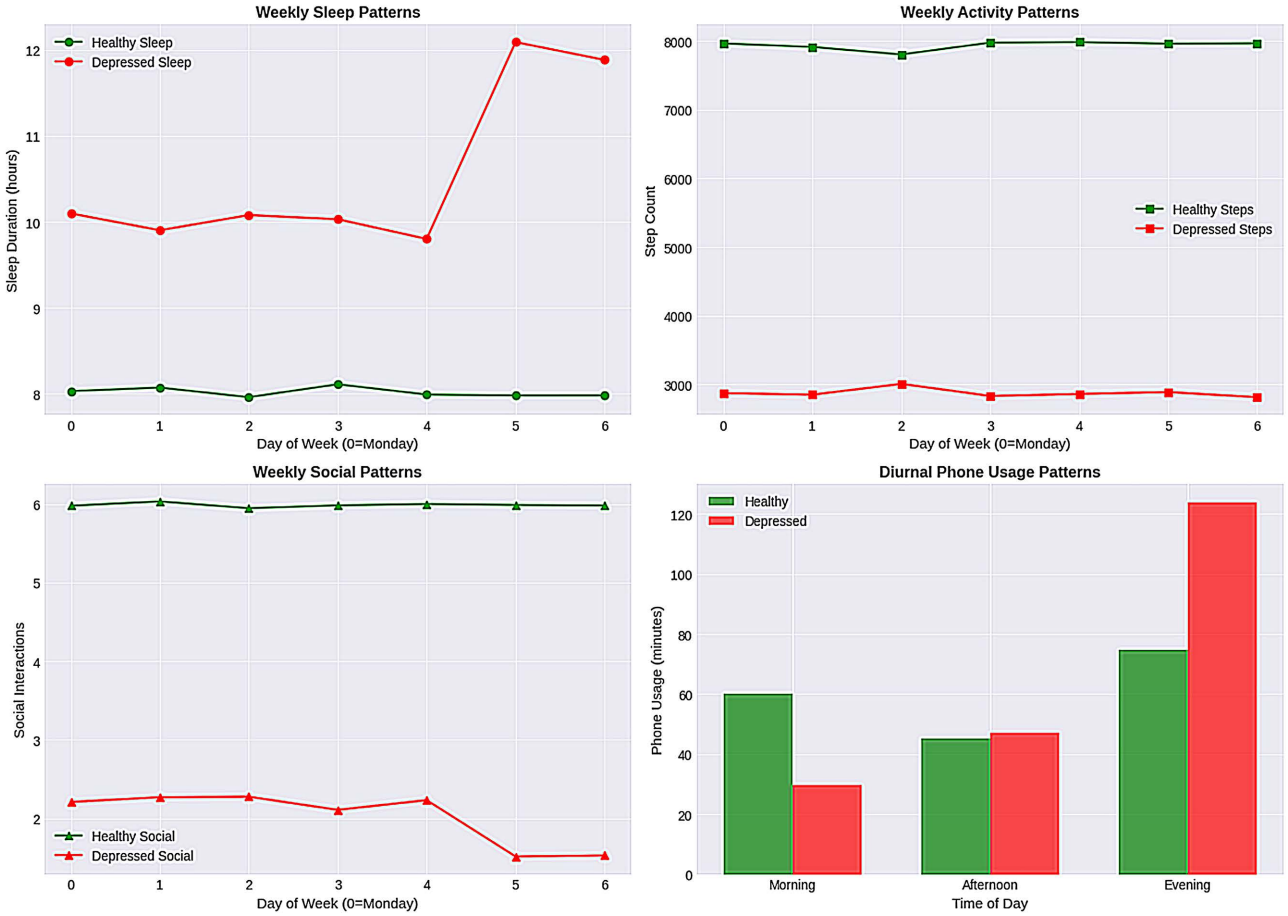


Figure 4. Weekly and diurnal behavioural patterns across healthy and depressed subjects.

Row 1:

- **Panel 4A (top-left):** Weekly sleep duration
- **Panel 4B (top-right):** Weekly physical activity

Row 2:

- **Panel 4C (bottom-left):** Weekly social interactions
- **Panel 4D (bottom-right):** Diurnal phone usage patterns

Figure 4 provides a structured analysis of weekly and diurnal rhythms.

In Panel 4A (top-left), healthy individuals maintain consistent sleep duration across the week, while depressed individuals show markedly increased weekend sleep, suggesting circadian irregularity and potential "catch-up" behaviour following weekday dysregulation. Panel 4B (top-right) shows that step counts for depressed subjects remain consistently lower than those of healthy individuals, with only minor weekday variation. This reflects a stable behavioural signature of reduced physical activity in depressive states. Panel 4C (bottom-left) demonstrates

higher social engagement throughout the week for healthy subjects, whereas depressed individuals experience more pronounced mid-week declines, aligning with motivational and social withdrawal patterns commonly reported in clinical literature. Finally, Panel 4D (bottom-right) shows that depressed subjects exhibit significantly elevated evening phone usage, a behaviour absent in the healthy group. This diurnal asymmetry reflects disrupted sleep hygiene and increased evening rumination, reinforcing the importance of time-of-day analysis in behavioural mental-health modelling [55]-[59]. These weekly and diurnal patterns together highlight the value of modelling rhythmic structure when interpreting digital phenotyping data for mental-health assessment [60]-[62].

3.4. Latent-Space Archetypes and Clustering

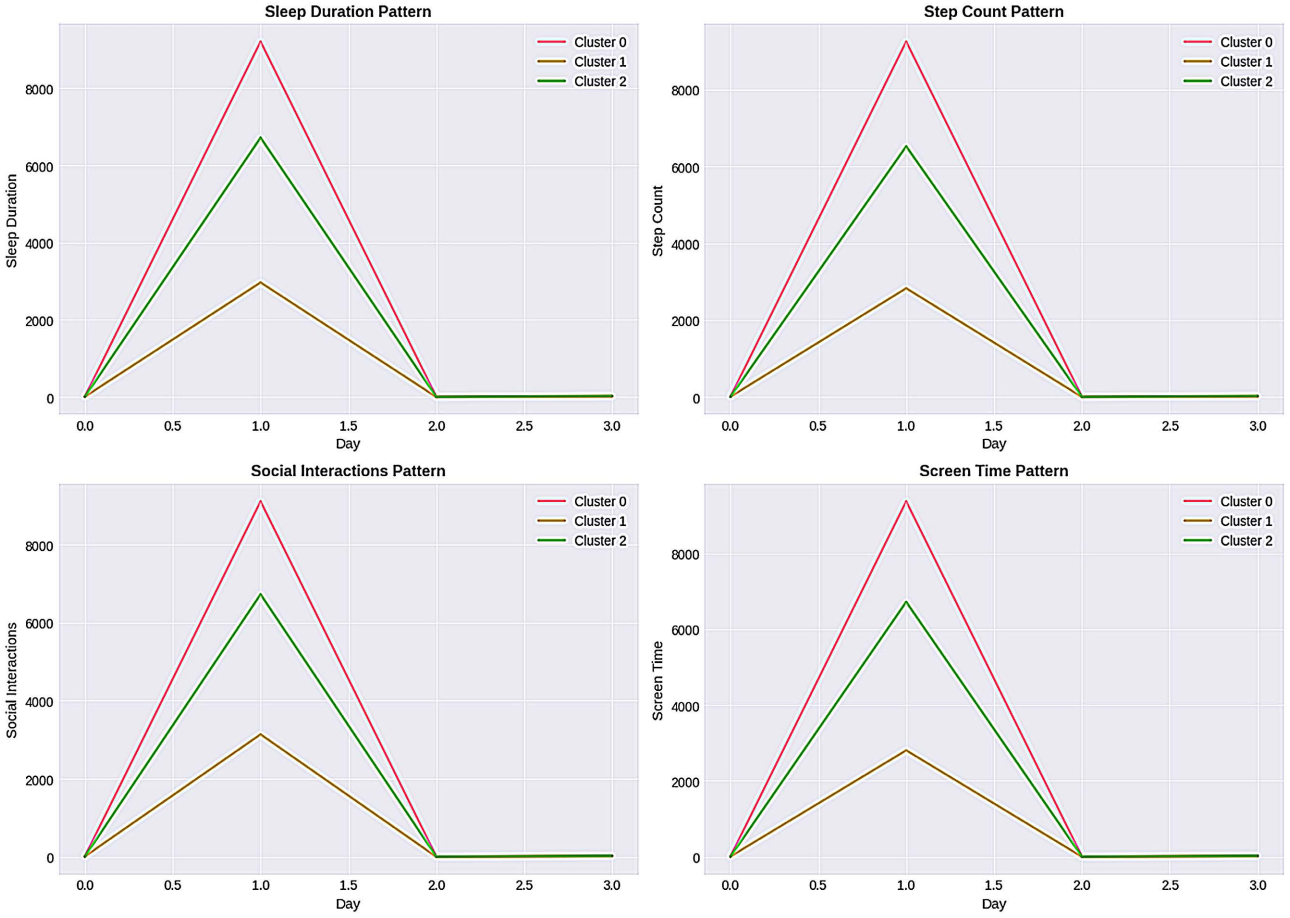


Figure 5. Cluster-level behavioural archetypes derived from SSL latent embeddings.

Row 1:

- **Panel 5A (top-left):** Sleep patterns by cluster
- **Panel 5B (top-right):** Step-count patterns by cluster

Row 2:

- **Panel 5C (bottom-left):** Social interaction patterns by cluster
- **Panel 5D (bottom-right):** Screen-time patterns by cluster

To investigate how the self-supervised encoder organizes subjects in latent space, we apply k-means clustering to the latent embeddings $\{z_s\}$. Figures 5(A)-(D) show cluster-level averages across key behavioural dimensions.

- Cluster 0 displays high activity, moderate sleep, and frequent social interactions, forming a profile predominantly aligned with healthy subjects.
- Cluster 1 demonstrates low activity and low social engagement, but with relatively normal sleep patterns indicative of milder depressive expressions or subthreshold symptoms.
- Cluster 2 exhibits moderate activity but disproportionately high evening screen time, consistent with digitally engaged yet behaviourally withdrawn individuals.

Panels 5A-5D (top-left to bottom-right) visually highlight these archetypes across sleep, steps, social interactions, and screen time. The emergence of such structured clusters without using depression labels during SSL training demonstrates that the encoder effectively identifies meaningful inter-subject behavioural variability beyond the binary classification boundary.

3.5. Global Geometry of SSL Representations

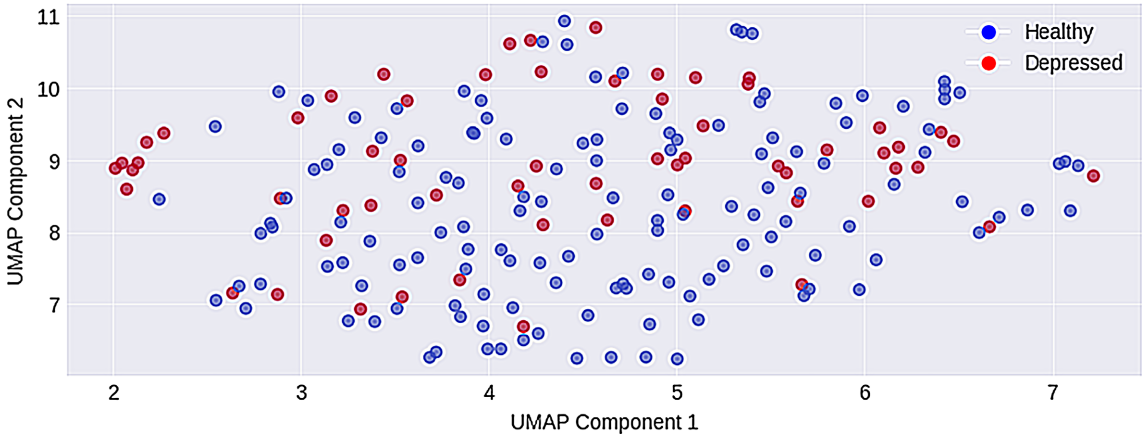


Figure 6. UMAP projection of SSL latent embeddings for healthy and depressed subjects.

Figure 6 presents a two-dimensional UMAP projection of the SSL-derived embeddings, with healthy subjects shown in blue and depressed subjects in red. Although the distributions overlap consistent with the subtle behavioural differences between groups the embeddings form partially distinct regions. Depressed subjects gravitate toward areas associated with longer sleep, lower activity, and increased screen time, reflecting the behavioural patterns seen in Figure 2 and Figure 4.

This geometric structure explains why downstream classifiers trained on SSL features achieve above-chance performance: the SSL encoder reshapes the high-dimensional behavioural trajectories into a more linearly separable manifold, enabling traditional classifiers to detect depression-related structure that is not clearly visible in raw feature space.

4. Discussion

This study demonstrates that a lightweight self-supervised encoder can uncover depression-related behavioural signatures from noisy, overlapping, and weakly discriminative synthetic daily trajectories. Despite being trained without access to depression labels, the SSL model learns latent representations that consistently outperform raw feature aggregates across all downstream classifiers. This finding underscores a central insight: depression is expressed not through static behavioural shifts but through temporal micro-dynamics distributed patterns of variability, rhythm disturbance, and cross-modal imbalance spanning sleep, activity, social engagement, and phone-based interaction. The mathematical formulation of the synthetic generator provides clarity on how label-dependent means, variances, weekly rhythms, and diurnal patterns give rise to the observed behavioural landscape. Importantly, the correlation analysis and PHQ-9 overlap confirm that depression cannot be inferred from isolated features or simple thresholds. Instead, meaningful detection arises from integrating multiple weak signals across time. This is precisely what SSL achieves: through augmentations and contrastive learning, behaviourally similar sequences are mapped closer together while dissimilar ones are separated, enabling the encoder to denoise, compress, and structurally reorganize the trajectories into a clinically informative latent space. From an applied perspective, the compact architecture and low computational footprint of the encoder make it feasible for deployment on smartphones and wearables, where memory and energy constraints typically preclude the use of large deep models. Once pre-trained, the encoder could operate directly on-device, generating risk-relevant embeddings from short behavioural windows. These embeddings could support real-time mental-health monitoring, adaptive intervention systems, and clinical decision-support dashboards. Nonetheless, several limitations warrant consideration. An additional limitation concerns the Gaussian assumptions embedded in the synthetic generator. Real-world digital phenotyping data often exhibit heavy-tailed distributions, burst-like behavioural events, missingness patterns, and device-specific artefacts that deviate from normality. Because the reconstruction objective implicitly assumes smooth and noise-perturbed structure, model performance in this synthetic Gaussian setting may overestimate generalization to empirical behavioural streams. Future work should incorporate non-Gaussian noise models and realistic sensor irregularities to more rigorously stress-test representation robustness. The dataset is fully synthetic; although clinically inspired, it cannot capture the full heterogeneity, contextual factors, or sensor artefacts of real-world populations. The temporal resolution is limited to daily data, whereas finer-grained (e.g., hourly) signals may reveal richer early-warning structure. Moreover, the current formulation models depression as a binary outcome, omitting symptom trajectories, remission patterns, comorbidities, and individual differences in behavioural expression. Future work should therefore focus on: (i) calibrating the generator using empirical statistics from real digital phenotyping cohorts, (ii) extending the framework to multi-label or continuous severity pre-

diction, (iii) integrating additional data streams such as GPS-derived mobility, speech acoustics, keyboard dynamics, or physiological sensing, and (iv) testing generalization under strict subject-wise cross-validation on real-world datasets. Incorporating uncertainty-aware or Bayesian prediction layers may further enhance the system's utility in risk-sensitive clinical settings.

Overall, this work provides an interpretable, mathematically grounded, and computationally efficient foundation for advancing behavioural representation learning in digital mental health, setting the stage for future real-world applications in early detection and continuous monitoring of depression.

5. Conclusion

This work presents a mathematically transparent synthetic framework for modelling depression-related daily behavioural trajectories and demonstrates that a lightweight self-supervised encoder can meaningfully disentangle healthy and depressed behavioural patterns [63]-[67]. Across multiple downstream classifiers, the SSL-derived embeddings consistently outperform raw feature aggregations in AUC and overall discriminability, indicating that the model successfully captures temporal micro-dynamics that static features overlook. In addition to improved predictive accuracy, the latent space reveals interpretable behavioural archetypes and structured separability, offering insight into how depression-related signatures manifest across different modalities. Together, these findings provide a clear proof of concept: compact, self-supervised models can serve as efficient and effective components of digital mental-health analytics, even in settings characterized by noise, imbalance, and overlapping behavioural profiles. By uniting synthetic behavioural modelling with temporal representation learning, this study establishes a reproducible testbed for evaluating early-detection algorithms and lays methodological groundwork for future real-world applications. Looking ahead, the proposed framework offers a foundation upon which more advanced systems can be built, including personalized behavioural models, multi-modal sensing pipelines, and real-time mobile deployment strategies. As digital phenotyping continues to evolve, the integration of lightweight SSL encoders with real-world data streams holds promise for enabling proactive mental-health monitoring, earlier detection of depressive risk, and more timely clinical intervention.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Healy, D. and Williams, J.M. (1988) Dysrhythmia, Dysphoria, and Depression: The Interaction of Learned Helplessness and Circadian Dysrhythmia in the Pathogenesis of Depression. *Psychological Bulletin*, **103**, 163-178. https://doi.org/10.1037//0033-2909.103.2.163

[2] Ehlers, C.L., Frank, E. and Kupfer, D.J. (1988) Social Zeitgebers and Biological Rhythms: A Unified Approach to Understanding the Etiology of Depression. *Ar-*

chives of General Psychiatry, **45**, 948-952.

[3] Sandra, C., Panzarella, G., Firth, J., *et al.* (2025) Decoding Depression: Exploring the Environment across Life Course.

[4] Corfman, E.L. (1979) Depression, Manic-Depressive Illness, and Biological Rhythms. Vol. 1. Department of Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration. National Institute of Mental Health, Division of Scientific and Public Information.

[5] Stark, K.D., Bronik, M.D., Wong, S., Wells, G. and Ostrander, R. (2000) Depressive Disorders. *Advanced Abnormal Child Psychology*, **2**, 291-326.

[6] Breitinger, S., Gardea-Resendez, M., Langholm, C., Xiong, A., Laivell, J., Stoppel, C., *et al.* (2023) Digital Phenotyping for Mood Disorders: Methodology-Oriented Pilot Feasibility Study. *Journal of Medical Internet Research*, **25**, e47006. https://doi.org/10.2196/47006

[7] Marciano, L., Vocaj, E., Bekalu, M.A., La Tona, A., Rocchi, G. and Viswanath, K. (2023) The Use of Mobile Assessments for Monitoring Mental Health in Youth: Umbrella Review. *Journal of Medical Internet Research*, **25**, e45540. https://doi.org/10.2196/45540

[8] Harris, C., Tang, Y., Birnbaum, E., Cherian, C., Mendhe, D. and Chen, M.H. (2024) Digital Neuropsychology Beyond Computerized Cognitive Assessment: Applications of Novel Digital Technologies. *Archives of Clinical Neuropsychology*, **39**, 290-304. https://doi.org/10.1093/arclin/acae016

[9] Stephan, C., Berckmans, D., Geris, L., *et al.* (2019) Mobile Health Revolution in Healthcare: Are We Ready?

[10] De La Fabián, R., Jiménez-Molina, Á. and Pizarro Obaid, F. (2023) A Critical Analysis of Digital Phenotyping and the Neuro-Digital Complex in Psychiatry. *Big Data & Society*, **10**, Article 20539517221149047. https://doi.org/10.1177/20539517221149097

[11] Dinga, R., Marquand, A.F., Veltman, D.J., Beekman, A.T.F., Schoevers, R.A., van Hemert, A.M., *et al.* (2018) Predicting the Naturalistic Course of Depression from a Wide Range of Clinical, Psychological, and Biological Data: A Machine Learning Approach. *Translational Psychiatry*, **8**, Article No. 241. https://doi.org/10.1038/s41398-018-0289-1

[12] Schmaal, L., Marquand, A.F., Rhebergen, D., van Tol, M., Ruhé, H.G., van der Wee, N.J.A., *et al.* (2015) Predicting the Naturalistic Course of Major Depressive Disorder Using Clinical and Multimodal Neuroimaging Information: A Multivariate Pattern Recognition Study. *Biological Psychiatry*, **78**, 278-286. https://doi.org/10.1016/j.biopsych.2014.11.018

[13] Cao, Y.C., Dai, J.L., Wang, Z.Y., *et al.* (2025) Machine Learning Approaches for Depression Detection on Social Media: A Systematic Review of Biases and Methodological Challenges. *Journal of Behavioral Data Science*, **5**, 67-102. https://doi.org/10.35566/jbds/caoyc

[14] Artitayaporn, R., Songpan, W. and Surinta, O. (2025) Depression Classification with Imbalanced Data Problems: Literature Survey. *Engineering Access*, **11**, 185-199.

[15] Onnela, J. and Rauch, S.L. (2016) Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health. *Neuropsychopharmacology*, **41**, 1691-1696. https://doi.org/10.1038/npp.2016.7

[16] Aung, M.H., Matthews, M. and Choudhury, T. (2017) Sensing Behavioral Symptoms of Mental Health and Delivering Personalized Interventions Using Mobile Technologies. *Depression and Anxiety*, **34**, 603-609.

https://doi.org/10.1002/da.22646

[17] Mohr, D.C., Zhang, M. and Schueller, S.M. (2017) Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. *Annual Review of Clinical Psychology*, **13**, 23-47. https://doi.org/10.1146/annurev-clinpsy-032816-044949

[18] Wang, W.C., Harari, G.M., Wang, R., *et al.* (2018) Sensing Behavioral Change over Time: Using Within-Person Variability Features from Mobile Sensing to Predict Personality Traits. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, **2**, 1-21.

[19] Difrancesco, S., Riese, H., Merikangas, K.R., Shou, H., Zipunnikov, V., Antypa, N., *et al.* (2021) Sociodemographic, Health and Lifestyle, Sampling, and Mental Health Determinants of 24-Hour Motor Activity Patterns: Observational Study. *Journal of Medical Internet Research*, **23**, e20700. https://doi.org/10.2196/20700

[20] Lane, N.D., Bhattacharya, S., Georgiev, P., Forlivesi, C. and Kawsar, F. (2015) An Early Resource Characterization of Deep Learning on Wearables, Smartphones and Internet-of-Things Devices. *Proceedings of the* 2015 *International Workshop on Internet of Things towards Applications*, Seoul, 1 November 2015, 7-12. https://doi.org/10.1145/2820975.2820980

[21] Hoang, M.L. (2025) A Comprehensive Review of Machine Learning, and Deep Learning in Wearable IoT Devices.

[22] Dargazany, A.R., Stegagno, P. and Mankodiya, K. (2018) Wearabledl: Wearable Internet-of-Things and Deep Learning for Big Data Analytics—Concept, Literature, and Future. *Mobile Information Systems*, **2018**, 1-20. https://doi.org/10.1155/2018/8125126

[23] Preuveneers, D. and Joosen, W. (2016) Privacy-Enabled Remote Health Monitoring Applications for Resource Constrained Wearable Devices. *Proceedings of the* 31*st Annual ACM Symposium on Applied Computing*, Pisa, 4-8 April 2016 , 119-124. https://doi.org/10.1145/2851613.2851683

[24] D'Ambrosio, S., De Pasquale, S., Iannone, G., Malandrino, D., Negro, A., Patimo, G., *et al.* (2016) Energy Consumption and Privacy in Mobile Web Browsing: Individual Issues and Connected Solutions. *Sustainable Computing: Informatics and Systems*, **11**, 63-79. https://doi.org/10.1016/j.suscom.2016.02.003

[25] Jerónimo, J., de Antonio, A. and Moral, C. (2018) Architectural Challenges on the Analysis of Human Behaviour in Synthetic Environments. *Proceedings of the 12th European Conference on Software Architecture: Companion Proceedings*, Madrid, 24-28 September 2018, 1-7. https://doi.org/10.1145/3241403.3241441

[26] De Landa, M. (1994) Virtual Environments and the Emergence of Synthetic Reason. In: *Flame Wars*, Duke University Press, 263-286. https://doi.org/10.2307/j.ctv1220m2w.15

[27] Way, J.C., Collins, J.J., Keasling, J.D. and Silver, P.A. (2014) Integrating Biological Redesign: Where Synthetic Biology Came from and Where It Needs to Go. *Cell*, **157**, 151-161. https://doi.org/10.1016/j.cell.2014.02.039

[28] Wang, Y. (2025) One to Two, Two to All: Towards Multimodal Self-Supervised Learning for Earth Observation. PhD Dissertation, Technische Universität München.

[29] Tajamul, A. and Bashir, J. (2025) FATE: Focal-Modulated Attention Encoder for Multivariate Time-Series Forecasting.

[30] Gu, H.X. (2025) AI for Medical Imaging: Foundation Models, 2D-to-3D Reconstruction, and Clinical Applications. PhD Dissertation, Duke University.

[31] Ayuso-Mateos, J.L., Nuevo, R., Verdes, E., Naidoo, N. and Chatterji, S. (2010) From Depressive Symptoms to Depressive Disorders: The Relevance of Thresholds. *British Journal of Psychiatry*, **196**, 365-371. https://doi.org/10.1192/bjp.bp.109.071191

[32] Tusa, N., Koponen, H., Kautiainen, H., Korniloff, K., Raatikainen, I., Elfving, P., *et al.* (2019) The Profiles of Health Care Utilization among a Non-Depressed Population and Patients with Depressive Symptoms with and without Clinical Depression. *Scandinavian Journal of Primary Health Care*, **37**, 312-318. https://doi.org/10.1080/02813432.2019.1639904

[33] Kline, E.R., Seidman, L.J., Cornblatt, B.A., Woodberry, K.A., Bryant, C., Bearden, C.E., *et al.* (2018) Depression and Clinical High-Risk States: Baseline Presentation of Depressed Vs. Non-Depressed Participants in the NAPLS-2 Cohort. *Schizophrenia Research*, **192**, 357-363. https://doi.org/10.1016/j.schres.2017.05.032

[34] Rubin, E.H., Veiel, L.L., Kinscherf, D.A., Morris, J.C. and Storandt, M. (2001) Clinically Significant Depressive Symptoms and Very Mild to Mild Dementia of the Alzheimer Type. *International Journal of Geriatric Psychiatry*, **16**, 694-701. https://doi.org/10.1002/gps.408

[35] Daghistani, T. and Alshammari, R. (2020) Comparison of Statistical Logistic Regression and Randomforest Machine Learning Techniques in Predicting Diabetes. *Journal of Advances in Information Technology*, **11**, 78-83. https://doi.org/10.12720/jait.11.2.78-83

[36] Huang, J., Tsai, Y., Wu, P., Lien, Y., Chien, C., Kuo, C., *et al.* (2020) Predictive Modeling of Blood Pressure during Hemodialysis: A Comparison of Linear Model, Random Forest, Support Vector Regression, XGBoost, LASSO Regression and Ensemble Method. *Computer Methods and Programs in Biomedicine*, **195**, Article 105536. https://doi.org/10.1016/j.cmpb.2020.105536

[37] Maulina, M., Hiola, Y.P. and Alamudi, A. (2025) Performance Evaluation of Multinomial Logistic Regression, Random Forest, and XGBoost Methods in Data Classification. *Journal of Mathematics, Computations and Statistics*, **8**, 355-373. https://doi.org/10.35580/jmathcos.v8i2.8459

[38] Sitompul, L.R., Nababan, A.A., Manihuruk, M.L., Ponsen, W.A. and Supriyandi, S. (2025) Comparison of XGBoost, Random Forest and Logistic Regression Algorithms in Stroke Disease Classification. *Sinkron*, **9**, 957-968. https://doi.org/10.33395/sinkron.v9i2.14794

[39] Boer, Y., Valencia, L., Setiadi, M.R., Eka Setiawan, K. and Hasani, M.F. (2023) Classification of Heart Disease: Comparative Analysis Using KNN, Random Forest, Gaussian Naive Bayes, XGBoost, SVM, Decision Tree, and Logistic Regression. 2023 5th *International Conference on Cybernetics and Intelligent System* (*ICORIS*), Pangkalpinang, 6-7 October 2023, 1-5. https://doi.org/10.1109/icoris60118.2023.10352195

[40] Snyder, C.K. and Chang, A. (2019) Mobile Technology, Sleep, and Circadian Disruption. In: *Sleep and Health*, Elsevier, 159-170. https://doi.org/10.1016/b978-0-12-815373-4.00013-7

[41] Kim, Y., Kim, E., Lee, Y. and Park, J. (2025) Role of Late-Night Eating in Circadian Disruption and Depression: A Review of Emotional Health Impacts. *Physical Activity and Nutrition*, **29**, 18-24. https://doi.org/10.20463/pan.2025.0003

[42] Reichenberger, D.A., Snyder, C.K. and Chang, A. (2026) Screen Use, Sleep, and Circadian Disruption. In: *Sleep and Health*, Elsevier, 207-215. https://doi.org/10.1016/b978-0-443-13954-3.00018-8

[43] Carrington, A.M., Manuel, D.G., Fieguth, P.W., Ramsay, T., Osmani, V., Wernly, B.,

et al. (2022) Deep ROC Analysis and AUC as Balanced Average Accuracy, for Improved Classifier Selection, Audit and Explanation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **45**, 329-341. https://doi.org/10.1109/tpami.2022.3145392

[44] Chicco, D. and Jurman, G. (2023) The Matthews Correlation Coefficient (MCC) Should Replace the ROC AUC as the Standard Metric for Assessing Binary Classification. *BioData Mining*, **16**, Article No. 4. https://doi.org/10.1186/s13040-023-00322-4

[45] Eve, R., Trevizani, R., Greenbaum, J.A., Carter, H., *et al*. (2023) The ROC-AUC Accurately Assesses Imbalanced Datasets.

[46] Baxter, J.L. (1993) Aggregate Behaviour. In: *Behavioural Foundations of Economics*, Palgrave Macmillan, 213-236. https://doi.org/10.1007/978-1-349-22627-6_12

[47] Bader, M., Abdelwanis, M., Maalouf, M. and Jelinek, H.F. (2024) Detecting Depression Severity Using Weighted Random Forest and Oxidative Stress Biomarkers. *Scientific Reports*, **14**, Article No. 16328. https://doi.org/10.1038/s41598-024-67251-y

[48] Jie, N., Zhu, M., Ma, X., Osuch, E.A., Wammes, M., Theberge, J., *et al*. (2015) Discriminating Bipolar Disorder from Major Depression Based on SVM-Foba: Efficient Feature Selection with Multimodal Brain Imaging Data. *IEEE Transactions on Autonomous Mental Development*, **7**, 320-331. https://doi.org/10.1109/tamd.2015.2440298

[49] Kaspar, M. (2019) Enabling Feature-Level Interpretability in Non-Linear Latent Variable Models: A Synthesis of Statistical and Machine Learning Techniques. PhD Dissertation, University of Oxford.

[50] Oldfield, J., Tzelepis, C., Panagakis, Y., Nicolaou, M.A. and Patras, I. (2024) Bilinear Models of Parts and Appearances in Generative Adversarial Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **46**, 8568-8579. https://doi.org/10.1109/TPAMI.2024.3415506

[51] Arafat, J., Tasmin, F. and Poudel, S. (2025) Feature Selection and Regularization in Multi-Class Classification: An Empirical Study of One-vs-Rest Logistic Regression with Gradient Descent Optimization and L1 Sparsity Constraints.

[52] De Bernardinis, M., Violi, V., Roncoroni, L., Boselli, A.S., Giunta, A. and Peracchia, A. (1999) Discriminant Power and Information Content of Ranson's Prognostic Signs in Acute Pancreatitis: A Meta-Analytic Study. *Critical Care Medicine*, **27**, 2272-2283. https://doi.org/10.1097/00003246-199910000-00035

[53] Solberg, H.E. (1978) Discriminant Analysis. *CRC Critical Reviews in Clinical Laboratory Sciences*, **9**, 209-242. https://doi.org/10.3109/10408367809150920

[54] Tasche, D. (2009) Estimating Discriminatory Power and PD Curves When the Number of Defaults Is Small.

[55] Davy, R., Esau, I., Chernokulsky, A., Outten, S. and Zilitinkevich, S. (2017) Diurnal Asymmetry to the Observed Global Warming. *International Journal of Climatology*, **37**, 79-93. https://doi.org/10.1002/joc.4688

[56] Chulliat, A., Blanter, E., Le Mouël, J.L. and Shnirman, M. (2005) On the Seasonal Asymmetry of the Diurnal and Semidiurnal Geomagnetic Variations. *Journal of Geophysical Research*: *Space Physics*, **110**, A5.

[57] Song, J. (1998) Diurnal Asymmetry in Surface Albedo. *Agricultural and Forest Meteorology*, **92**, 181-189. https://doi.org/10.1016/s0168-1923(98)00095-1

[58] Long, K.M. and Meadows, G.N. (2018) Simulation Modelling in Mental Health: A Systematic Review. *Journal of Simulation*, **12**, 76-85.

https://doi.org/10.1057/s41273-017-0062-0

[59] Almeda, N., García-Alonso, C.R., Salinas-Pérez, J.A., Gutiérrez-Colosía, M.R. and Salvador-Carulla, L. (2019) Causal Modelling for Supporting Planning and Management of Mental Health Services and Systems: A Systematic Review. *International Journal of Environmental Research and Public Health*, **16**, Article 332. https://doi.org/10.3390/ijerph16030332

[60] Onnela, J. (2021) Opportunities and Challenges in the Collection and Analysis of Digital Phenotyping Data. *Neuropsychopharmacology*, **46**, 45-54. https://doi.org/10.1038/s41386-020-0771-3

[61] Cohen, A.S., Cox, C.R., Masucci, M.D., Le, T.P., Cowan, T., Coghill, L.M., *et al.* (2020) Digital Phenotyping Using Multimodal Data. *Current Behavioral Neuroscience Reports*, **7**, 212-220. https://doi.org/10.1007/s40473-020-00215-4

[62] Martinez-Martin, N., Insel, T.R., Dagum, P., Greely, H.T. and Cho, M.K. (2018) Data Mining for Health: Staking Out the Ethical Territory of Digital Phenotyping. *npj Digital Medicine*, **1**, Article No. 68. https://doi.org/10.1038/s41746-018-0075-8

[63] Boche, H., Fono, A. and Kutyniok, G. (2024) A Mathematical Framework for Computability Aspects of Algorithmic Transparency. 2024 *IEEE International Symposium on Information Theory* (*ISIT*), Athens, 7-12 July 2024, 3089-3094. https://doi.org/10.1109/isit57864.2024.10619190

[64] Vershynin, R. (2020) Collaborative Research: A Mathematical Framework for Generating Synthetic Data. NSF Award Number 2027299. *Directorate for Mathematical and Physical Sciences*, **20**, Article 27299.

[65] Strohmer, T. (2020) ATD: A Mathematical Framework for Generating Synthetic Data. NSF Award Number 2027248. *Directorate for Mathematical and Physical Sciences*, **20**, Article 27248.

[66] Calhoun, Z.D., Lahrichi, S., Ren, S., Malof, J.M. and Bradbury, K. (2022) Self-Supervised Encoders Are Better Transfer Learners in Remote Sensing Applications. *Remote Sensing*, **14**, Article 5500. https://doi.org/10.3390/rs14215500

[67] Li, Z.Z., Zhao, K., Chen, P.D., Wang, D.W., *et al.* (2025) Disentangled Representation Learning for Capturing Individualized Brain Atrophy via Pseudo-Healthy Synthesis. *IEEE Journal of Biomedical and Health Informatics*, **29**, 5056-5068. https://doi.org/10.1109/jbhi.2025.3543218