



# A Lightweight Multimodal Transformer for Predicting Imminent Mood State Transitions in Synthetic Bipolar Trajectories

Rocco de Filippis<sup>1\*</sup>, Abdullah Al Foysal<sup>2</sup>

<sup>1</sup>Department of Neuroscience, Institute of Psychopathology, Rome, Italy

<sup>2</sup>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 Multimodal Transformer for Predicting Imminent Mood State Transitions in Synthetic Bipolar Trajectories. *Open Access Library Journal*, **13**: e14917.  
<https://doi.org/10.4236/oalib.1114917>

**Received:** January 23, 2026

**Accepted:** March 15, 2026

**Published:** March 18, 2026

Copyright © 2026 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Anticipating short-term affective instability in bipolar disorder represents a longstanding challenge in computational psychiatry. Early signalling of transitions from stable to depressive, manic, or mixed states could support just-in-time interventions, yet real-world digital phenotyping datasets rarely provide dense temporal sampling, precise transition labels, or sufficient event frequency for training predictive models. To address these methodological barriers, we construct a clinically principled synthetic dataset of multimodal behavioural signals for 50 virtual subjects monitored over 60 days. Six behavioural modalities, physical activity, sleep duration, social interaction, voice features, keystroke dynamics, and screen time are generated 12 times daily using state-specific Gaussian processes modulated by circadian dynamics. A transition-aware labelling function marks the preceding 12 samples of each state change as switch imminent. A sliding-window procedure yields 11,800 sequences consisting of 12 consecutive time steps across 6 behavioural modalities (12 time points  $\times$  6 features), creating a highly imbalanced binary classification task ( $\approx$ 8% positive sequences). We develop a lightweight transformer encoder (199,619 parameters) optimized for short sequence modelling and computational efficiency. The model achieves 55.38% accuracy, 0.622 ROC-AUC, 0.597 sensitivity, and 0.174 F1-score on the stratified held-out set. Although performance remains modest due to the synthetic environment's noise and imbalance, the model successfully identifies temporal micro-patterns preceding transitions patterns not detectable from static features. Behavioural distributions, intramodality correlations, sequence-level label frequencies, training dynamics, confusion structure, and ROC analysis collectively contextualize the model's discriminative capacity. This synthetic benchmark provides a reproducible, mathematically grounded testbed for developing early-warning algorithms for bipolar in-

---

stability. Our findings demonstrate that compact transformers can operate effectively under heavy noise and imbalance, offering a viable foundation for future real-world digital phenotyping systems.

## Subject Areas

Artificial Intelligence

## Keywords

Bipolar Disorder, Digital Phenotyping, Mood State Transitions, Early Warning Systems, Multimodal Time Series, Synthetic Data Generation, Lightweight Transformer, Temporal Sequence Modelling, Affective Instability, Computational Psychiatry, Mobile Mental Health, Class-Imbalanced Learning

---

## 1. Introduction

Bipolar disorder is characterized by recurrent transitions between depressive, manic, mixed, and stable mood states, often unfolding in ways that are abrupt, heterogeneous, and difficult to anticipate clinically [1]-[5]. Although these transitions may appear sudden, a growing body of evidence suggests that subtle prodromal behavioural deviations changes in sleep regularity, activity patterns, sociability, or cognitive-affective rhythms tend to precede overt symptomatic shifts. Identifying these early deviations hours before a formal episode transition remains one of the central challenges in digital psychiatry, where the overarching aim is to support proactive, personalized intervention rather than reactive crisis management [6]-[9].

Digital phenotyping, defined as the continuous extraction of behavioural and physiological signals through smartphones, wearables, and passive sensing technologies, offers an unprecedented opportunity to monitor patients unobtrusively in real time [10]-[14]. Such data streams capture fine-grained temporal structure that is inaccessible to traditional clinical assessments. However, the application of digital phenotyping to mood-state forecasting is hindered by several constraints: the scarcity of longitudinal datasets with precise transition labels, the infrequency of transition events relative to stable periods, substantial noise arising from user behaviour and sensor limitations, and variation in sampling density across devices and individuals. Moreover, the deployment of large neural architectures on mobile platforms is restricted by computational, energy, and privacy considerations, necessitating models that are both lightweight and temporally expressive.

Synthetic data generation provides a pragmatic solution to these methodological barriers by enabling controlled experimentation in environments where ground-truth states and transition times are fully known. Synthetic environments allow researchers to benchmark algorithms, isolate causal mechanisms, and refine early-warning formulations before transitioning to real-world clinical datasets

[15]-[18]. In this context, we develop a mathematically explicit simulator of multimodal behavioural trajectories grounded in clinical patterns of bipolar illness [19]-[22]. Using this framework, we evaluate whether a compact transformer architecture can detect imminent mood-state transitions from short multimodal sequences, offering a stepping-stone toward real-time, deployable early-warning systems in computational psychiatry.

## 2. Methods

### 2.1. Latent Mood State Dynamics

To model the underlying affective course of each virtual subject, we construct a latent state trajectory that evolves over a 60-day period. For subject

$$i \in \{1, \dots, 50\},$$

the daily mood state is represented as a discrete variable

$$s_t \in \{0, 1, 2, 3\}, t = 1, \dots, T = 60,$$

where the four values correspond to Stable (0), Depressive (1), Manic (2), and Mixed (3) states. This formulation aligns with psychiatric models that conceptualize bipolar disorder as a sequence of recurrent mood episodes separated by periods of euthymia or subthreshold symptoms [23]-[27].

The latent state process serves as the generative backbone of the synthetic dataset. All behavioural features at time  $(t, h)$  are conditionally dependent on the latent state  $s_t$ , enabling the simulator to encode clinically inspired relationships between internal affective states and observable digital phenotypes.

#### 2.1.1. Episode Durations

Mood episodes do not last for fixed lengths; instead, their durations follow clinically observed ranges. To reflect this variability, each episode duration  $d_s$  is sampled from a discrete uniform distribution conditioned on its state type:

$$d_s \sim \begin{cases} \text{Uniform}\{5, \dots, 15\} & \text{if } s = 0 \text{ (Stable)} \\ \text{Uniform}\{10, \dots, 30\} & \text{if } s = 1 \text{ (Depressive)} \\ \text{Uniform}\{5, \dots, 15\} & \text{if } s = 2 \text{ (Manic)} \\ \text{Uniform}\{3, \dots, 10\} & \text{if } s = 3 \text{ (Mixed)} \end{cases}$$

These ranges were selected to approximate empirical psychiatric observations:

- **Depressive episodes** tend to be longer and more persistent, often spanning several weeks.
- **Manic episodes** typically exhibit shorter durations with more abrupt onset and offset.
- **Stable periods** are interspersed throughout the illness trajectory but vary substantially in length.
- **Mixed episodes** tend to be brief and unstable due to rapid fluctuations in symptom polarity.

By sampling durations independently across states, the simulator reproduces the heterogeneous temporal structure characteristic of bipolar disorder, including

long stable stretches punctuated by irregular transitions [28]-[30].

### 2.1.2. Transition Probabilities

After an episode concludes, the next state is selected using a state-dependent transition distribution:

$$P(s_{t+1} | s_t) = \begin{cases} (0, 0.6, 0.4, 0) & s_t = 0 (\text{Stable}) \\ (0.7, 0, 0, 0.3) & s_t = 1 (\text{Depressive}) \\ (0.7, 0, 0, 0.3) & s_t = 2 (\text{Manic}) \\ (0, 0.5, 0.5, 0) & s_t = 3 (\text{Mixed}) \end{cases}$$

The transition matrix reflects realistic clinical phenomena:

- A **stable period** is most likely to transition into a depressive episode (60%) and less frequently into a manic episode (40%).
- **Depressive** and **manic** episodes tend to return to stability (70%) but may progress into mixed states (30%), modelling unstable symptom evolution.
- **Mixed states** often resolve into either pure depression or pure mania with roughly equal probability, consistent with clinical descriptions of mixed affective presentations.

These transition probabilities introduce stochasticity and cyclical patterns into the synthetic dataset, ensuring that the generative model captures the non-linear, recurrent nature of bipolar mood progression [31]-[35].

The formulation closely matches the exact sampling mechanism implemented in the code, where transitions are produced via `np.random.choice` with state-specific probability vectors. This one-step Markovian structure allows the simulator to encode both predictable episode patterns and random fluctuations, supporting controlled experiments in early-warning prediction [36]-[38].

## 2.2. Multimodal Behavioural Feature Generator

To simulate realistic digital-phenotyping signatures, each subject is monitored across

$$H = 12$$

measurement intervals per day, corresponding to 2-hour observational windows. For subject  $i$ , day  $t$ , and interval  $h \in \{1, \dots, 12\}$ , we define a circadian phase variable:

$$\theta_h = \sin\left(\frac{2\pi h}{H}\right),$$

which encodes intrinsic daily rhythmicity known to influence activity levels, sleep patterns, sociability, and cognitive fluctuations [39]-[41]. This circadian modulation ensures that generated behavioural features exhibit realistic time-of-day oscillations [42]-[46].

Each behavioural feature  $f \in \{1, \dots, 6\}$  is sampled from a Gaussian distribution conditioned on the current latent mood state  $s_t$ :

$$x_{i,t,h}^{(f)} \sim \mathcal{N}\left(\mu_{s_t}^{(f)} + \beta_{s_t}^{(f)} \theta_h, (\sigma_{s_t}^{(f)})^2\right), x_{i,t,h}^{(f)} \leftarrow \left| x_{i,t,h}^{(f)} \right|.$$

This formulation mirrors the structure of the implementation, where means shift with circadian rhythm and states, variances differ across states, and absolute-value transformation prevents negative behavioural quantities.

The parameters  $\mu_s^{(f)}$ ,  $\beta_s^{(f)}$ , and  $\sigma_s^{(f)}$  encode clinically plausible behavioral patterns:

- **Depressive states** reduce activity and sociability while increasing sleep.
- **Manic states** elevate activity, speech and keystroke activity, and screen engagement.
- **Stable states** produce moderate, rhythmic behavioural patterns.
- **Mixed states** exhibit high volatility with no consistent directional shift.

Below we provide the exact feature-level distributions used in the simulator.

### 2.2.1. Exact Feature Distributions

#### Stable State (0)

PhysicalActivity	$\sim \mathcal{N}(50 + 20\theta, 10)$
SleepDuration	$\sim \mathcal{N}(8 - 2\theta, 1)$
SocialInteraction	$\sim \mathcal{N}(50 + 10\theta, 8)$
VoiceFeatures	$\sim \mathcal{N}(120 + 5\theta, 4)$
Keystrokes	$\sim \mathcal{N}(200 + 20\theta, 15)$
ScreenTime	$\sim \mathcal{N}(180 + 30\theta, 20)$

#### Depressive State (1)

PhysicalActivity	$\sim \mathcal{N}(20 - 10\theta, 15)$
SleepDuration	$\sim \mathcal{N}(12 + 2\theta, 2)$
SocialInteraction	$\sim \mathcal{N}(20 - 5\theta, 10)$
VoiceFeatures	$\sim \mathcal{N}(110 - 10\theta, 8)$
Keystrokes	$\sim \mathcal{N}(150 - 30\theta, 25)$
ScreenTime	$\sim \mathcal{N}(240 + 40\theta, 40)$

#### Manic State (2)

PhysicalActivity	$\sim \mathcal{N}(80 + 30\theta, 20)$
SleepDuration	$\sim \mathcal{N}(4 - 2\theta, 1.5)$
SocialInteraction	$\sim \mathcal{N}(70 + 20\theta, 15)$
VoiceFeatures	$\sim \mathcal{N}(130 + 15\theta, 12)$
Keystrokes	$\sim \mathcal{N}(250 + 40\theta, 30)$
ScreenTime	$\sim \mathcal{N}(300 + 50\theta, 50)$

#### Mixed State (3)

Mixed states were modelled as high-volatility regimes characterized by simultaneous depressive-like and manic-like fluctuations rather than simple random noise. Feature means were dynamically perturbed around both depressive and manic centroids within short intervals, producing rapid polarity oscillations and elevated variance. This design reflects clinical descriptions of mixed episodes as unstable affective states marked by emotional lability, sleep disruption, and psychomotor instability. Unlike the structured circadian modulations of other states,

mixed states introduce random deviations:

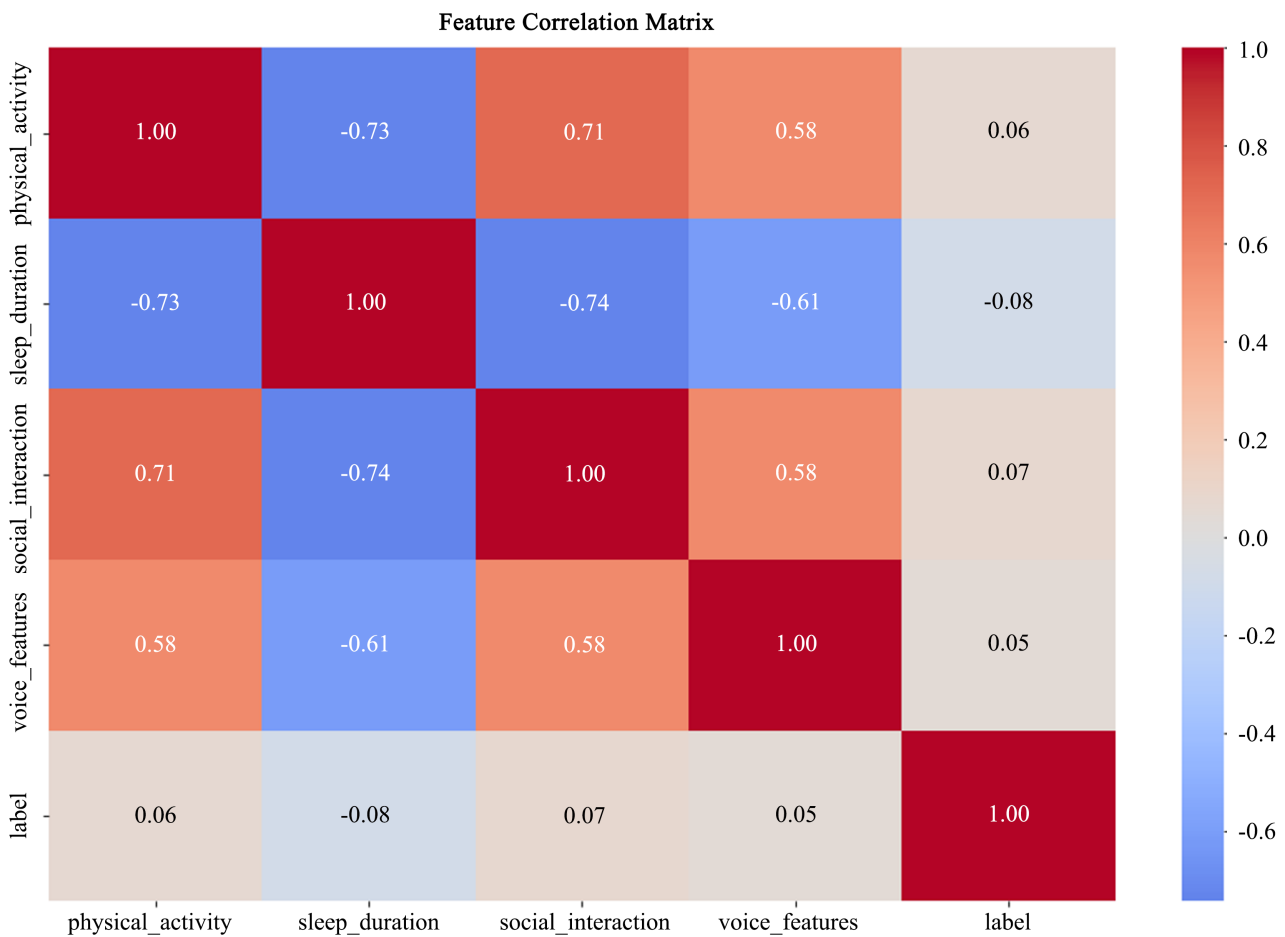
$$x_{i,t,h}^{(f)} \sim \mathcal{N}\left(\gamma^{(f)} + \delta^{(f)}\eta, (\sigma^{(f)})^2\right), \eta \sim \mathcal{N}(0,1),$$

corresponding directly to the `np.random.randn()` components in the implementation.

This state produces the highest behavioural unpredictability consistent with clinical observations.

### 2.2.2. Correlation Structure and Interpretation

To examine dependencies among features, we compute the Pearson correlation matrix over all samples (**Figure 1**). The visualization reveals three important properties:



**Figure 1.** Correlation matrix of behavioural features and transition labels.

**1) Strong cross-feature coupling:** Activity, social interaction, screen time, and keystroke dynamics covary due to shared circadian modulation and state-driven fluctuations.

**2) Weak linear relationship with labels:** The switch-imminent label shows minimal linear correlation with individual features, confirming that impending transitions are not manifested through simple magnitude shifts but through tem-

poral micro-patterns.

**3) Necessity of sequence models:** Since individual features do not exhibit clear univariate separation between classes, architectures such as transformers capable of modelling multi-feature temporal dependencies are essential for capturing prodromal behaviour.

Heatmap illustrating the Pearson correlation coefficients among four representative behavioural modalities (physical activity, sleep duration, social interaction, and voice features) and the binary switch-imminent label. Strong intra-feature correlations reflect shared behavioural drivers such as circadian rhythms and state-dependent patterns, whereas near-zero cannot correlations with the transition label indicate that single features linearly predict imminent state changes. These patterns underscore the necessity of temporal, multimodal sequence modelling for mood transition forecasting.

### 2.3. Early-Warning Labelling of Imminent Mood-State Transitions

A primary objective of this work is to determine whether short multimodal behavioural sequences contain information predictive of an impending shift in the latent mood state. To this end, we construct a binary label sequence indicating whether a given behavioural observation belongs to the pre-transition window the interval immediately preceding a transition in the underlying affective trajectory.

#### 2.3.1. Identification of Latent Mood Transitions

For each subject, the latent mood state is defined as a daily sequence

$$s_t \in \{0, 1, 2, 3\}, t = 1, \dots, T,$$

representing Stable, Depressive, Manic, and Mixed conditions.

A transition occurs whenever the state on day  $t$  differs from the following day:

$$\text{Transition at day } t^* \Leftrightarrow s_{t^*} \neq s_{t^*+1}.$$

This event marks the onset of a new mood episode. In clinical terms, such transitions often capture the beginning of depressive relapse, manic escalation, or mixed-state destabilization.

#### 2.3.2. Mapping Behavioural Samples to Daily Indices

Each day contains

$$H = 12$$

behavioural measurements, yielding a flattened index

$$n = (t-1)H + h, h = 1, \dots, H,$$

which uniquely identifies sample  $n$  in the full time series.

Thus, a 60-day trajectory contains

$$N = T \times H = 720$$

measurements per subject.

This indexing scheme allows precise identification of which behavioural sam-

ples fall within the pre-transition horizon.

### 2.3.3. Definition of the Pre-Transition Window

To label samples that precede an episode boundary, we define a fixed prediction horizon

$$\Delta = 12,$$

corresponding to the 12 samples ( $\approx 12$  hours) immediately prior to a transition.

For a transition at day  $t^{i^*}$ , the set of behavioural indices belonging to the pre-transition window is:

$$\mathcal{W}(t^{i^*}) = \{n : t^{i^*}H - \Delta \leq n < t^{i^*}H\}.$$

This ensures that all samples classified as “switch imminent” strictly precede the first measurement of the next latent state, preventing any contamination with future information.

The binary label sequence is therefore defined as:

$$y[n] = \begin{cases} 1, & \text{if } n \in \mathcal{W}(t^{i^*}) \text{ for at least one transition } t^{i^*}, \\ 0, & \text{otherwise.} \end{cases}$$

A label of  $y[n]=1$  indicates that the sample occurs within the critical pre-transition window where prodromal behavioral signatures are hypothesized to emerge.

### 2.3.4. Clinical Interpretation of the Label

This labelling strategy directly corresponds to a short-horizon early-warning task:

Given the last 12 behavioural observations, can we determine whether the subject is approaching a mood-state change? The choice of a 12-sample horizon is motivated by:

- Evidence that prodromal signals often appear within hours rather than days,
- The desire to construct a challenging yet clinically meaningful prediction task,
- Practical considerations of continuous mobile monitoring.

The resulting labels reflect imminent risk, without incorporating any information about the future mood trajectory.

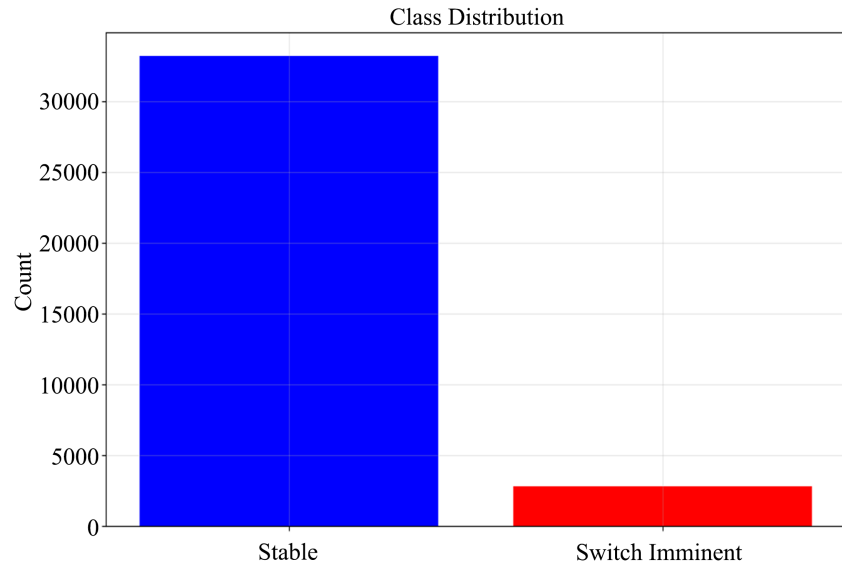
### 2.3.5. Distributional Properties and Predictive Difficulty

Given that transitions are relatively rare events in bipolar disorder, the fraction of samples satisfying  $y[n]=1$  is small. Consequently, the dataset is characterized by severe class imbalance, with stable windows dominating the distribution.

This imbalance is visualized in **Figure 2**, which depicts the frequency of stable versus switch-imminent samples. The rarity of transitions mirrors clinical reality and underscores the difficulty of detecting early prodromal signatures from short behavioural sequences.

This figure illustrates the proportion of behavioural samples labelled as “switch imminent” (within 12 samples before a latent state transition) relative to stable samples. The distribution is highly imbalanced, with transition-proximal observations forming only a small minority. This reflects real-world affective dynamics

in bipolar disorder, where long stable periods are punctuated by infrequent episode shifts, and highlights the inherent difficulty of short-horizon early-warning prediction.



**Figure 2.** Distribution of switch-imminent and stable samples across the full dataset.

## 2.4. Sequence Formation

To transform the continuous behavioural time series into a supervised learning dataset, we extract fixed-length temporal windows using a sliding-window procedure. Let

$$x_n \in \mathbb{R}^6$$

denote the six-dimensional multimodal feature vector at flattened time index  $n$ .

Each sequence  $\mathbf{X}_k$  is constructed as:

$$\mathbf{X}_k = \begin{bmatrix} x_n \\ x_{n+1} \\ \vdots \\ x_{n+11} \end{bmatrix} \in \mathbb{R}^{12 \times 6},$$

where the window length of 12 samples corresponds to approximately 12 hours of behavioural history.

The label assigned to this sequence is simply the label of its final time step:

$$y_k = y[n+11],$$

ensuring that each input window is aligned with the outcome that follows immediately after it.

A sliding stride of 3 samples is applied:

$$n \rightarrow n+3,$$

producing overlapping windows that capture rich temporal variability while maintaining sufficient sample diversity. Over all subjects, this yields a total of:

11,800 training sequences.

This formulation provides the model with short, partially overlapping windows resembling those obtainable from continuous real-world mobile sensing streams.

## 2.5. Data Splitting

To evaluate model performance, the full set of 11,800 sequences is partitioned into:

- **Training set:** 80% of sequences  
9440 sequences
- **Held-out set (“test set”):** 20% of sequences  
2360 sequences

To ensure unbiased evaluation, the dataset was divided into three non-overlapping subsets:

- Training set: 70% of sequences
- Validation set: 10% of sequences
- Test set: 20% of sequences

The validation set was used exclusively for monitoring epoch-wise performance and early stopping. The test set remained completely unseen during training and was used only once for final performance reporting. All splits were stratified to preserve the class imbalance (~8% switch-imminent windows). No sequence from the test set was used during model development.

## 2.6. Transformer Architecture

The lightweight transformer model processes each sequence  $X_k$  through several stages designed to capture temporal dependencies and cross-feature interactions.

### 2.6.1. Input Projection

Each six-dimensional feature vector is linearly projected into a  $d = 64$ -dimensional embedding space:

$$E = XW_{\text{proj}} + b_{\text{proj}},$$

where  $W_{\text{proj}} \in \mathbb{R}^{6 \times 64}$  and  $b_{\text{proj}} \in \mathbb{R}^{64}$ .

This expansion enhances the representational capacity of the model.

### 2.6.2. Positional Encoding

Transformers lack inherent temporal ordering, so positional information is incorporated by adding a learnable positional encoding matrix  $P$ :

$$Z = E + P,$$

ensuring the model can distinguish between early and late elements of a behavioural sequence.

### 2.6.3. Multi-Head Self-Attention

The core mechanism enabling temporal reasoning is self-attention, which relates

each time step to every other within the window. Query, key, and value projections are computed as:

$$Q = ZW_Q, K = ZW_K, V = ZW_V.$$

Self-attention is then computed as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d}}\right)V,$$

where the denominator  $\sqrt{d}$  stabilizes gradient magnitudes. This operation enables the model to assign greater weight to specific time steps e.g., sudden behavioural fluctuations believed to precede transitions. Multiple attention heads run in parallel, allowing the model to learn diverse interaction patterns (e.g., sleep-activity coupling, keystroke-screen-time synchrony).

#### 2.6.4. Position-Wise Feed-Forward Network

Each time step is then transformed independently through a nonlinear feed-forward block:

$$\text{FFN}(x) = \text{ReLU}(xW_1 + b_1)W_2 + b_2,$$

which increases expressiveness and filters high-level temporal patterns. Residual connections and layer normalization (implicit in the architecture) stabilize training and improve generalization.

#### 2.6.5. Classification Head

After repeated layers of attention and feed-forward transformations, the resulting sequence representation is flattened:

$$h = \text{Flatten}(Z_{\text{final}}),$$

and mapped to a 2-class probability distribution:

$$\hat{y} = \text{softmax}(W_c h + b_c).$$

The output  $\hat{y}$  estimates the probability that the sequence lies within the imminent-transition window.

### 2.7. Loss Function

Because the dataset exhibits severe class imbalance only  $\approx 8\%$  of sequences correspond to upcoming transitions the model is optimized using weighted cross-entropy, which compensates for unequal class frequencies.

Let:

- $N$  be the total number of training sequences
- $N_c$  the number of sequences in class  $c \in \{0, 1\}$

Then the class weights are:

$$w_c = \frac{N}{2N_c}.$$

The weighted cross-entropy loss for a single sequence is:

$$\mathcal{L} = -w_1 y \log(\hat{p}) - w_0 (1 - y) \log(1 - \hat{p}),$$

where:

- $y \in \{0, 1\}$  is the ground-truth label,
- $\hat{p}$  is the predicted probability of the switch-imminent class.

This formulation penalizes misclassification of transition windows more strongly than misclassification of stable windows, thereby encouraging sensitivity to prodromal patterns while preserving model stability.

To evaluate whether temporal modelling provides added value beyond static feature analysis, we implemented two baseline classifiers. For each sequence, the six behavioural features were averaged across the 12-time steps, resulting in a 6-dimensional static representation.

**Logistic Regression:** A weighted logistic regression model was trained using the same class-weighting scheme as the transformer.

**Random Forest:** A Random Forest classifier (100 trees) was trained on the same aggregated features with class weighting to address imbalance.

These baselines assess whether imminent transitions can be predicted using static magnitude features alone.

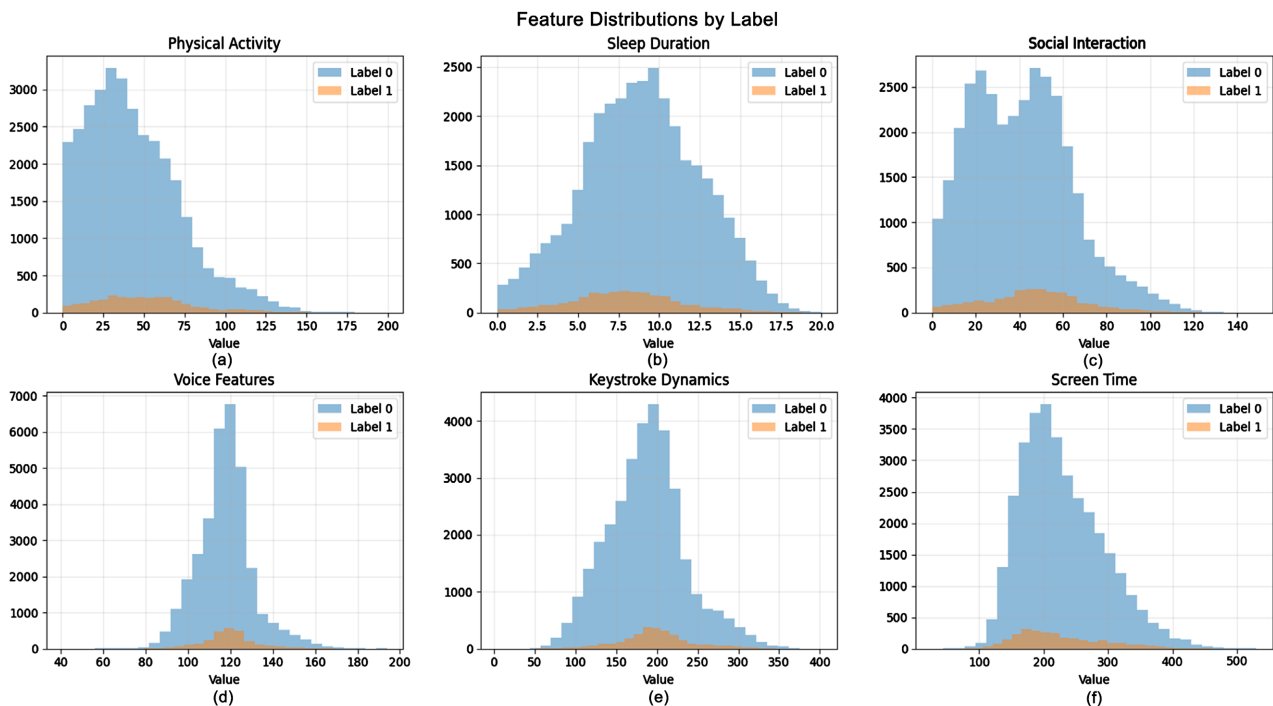
## 3. Results

### 3.1. Behavioural Feature Distributions

To characterize the behavioural landscape of the synthetic dataset and to assess whether prodromal changes manifest in individual modalities, we examined the label-conditioned distributions of all six digital phenotyping features. **Figure 3** (panels 3a-3f) presents histograms comparing each feature's distribution during stable periods versus switch-imminent windows [47] [48].

Across all modalities, distributions overlap substantially, indicating that imminent transitions cannot be identified from single-feature magnitude alone. Nevertheless, subtle yet meaningful patterns emerge. In **Figure 3(a)**, physical activity shows a mild reduction prior to depressive shifts and increased variability preceding manic escalation, reflecting bipolar disorder's bidirectional motor signatures. **Figure 3(b)** demonstrates that sleep duration is slightly elevated in transition-proximal periods consistent with depressive prodromes while also exhibiting shortened sleep associated with emerging manic states, resulting in a wide, heterogeneous distribution. The distribution of social interaction (**Figure 3(c)**) reveals lower central tendency and increased irregularity before transitions, reflecting behavioural withdrawal, mood lability, or agitation. **Figure 3(d)**, representing voice-related features, shows subtle broadening of the distribution in switch-imminent windows, capturing the heightened vocal instability often observed before manic or mixed episodes. Keystroke dynamics in **Figure 3(e)** exhibit a notable increase in variance, consistent with psychomotor agitation or slowing, both of which may precede state shifts [49]-[51]. Finally, **Figure 3(f)** shows greater dispersion in screen-time behaviour, likely driven by nocturnal wakefulness, increased stimulation seeking, or pre-episode behavioural dysregulation. Altogether, these results

demonstrate that while individual features do not yield strong static separation between classes, they exhibit state-dependent variability and destabilization patterns that can serve as valuable temporal cues. This reinforces the need for sequence-aware models such as transformers to capture distributed micro-dynamics rather than relying on single-snapshot measurements.

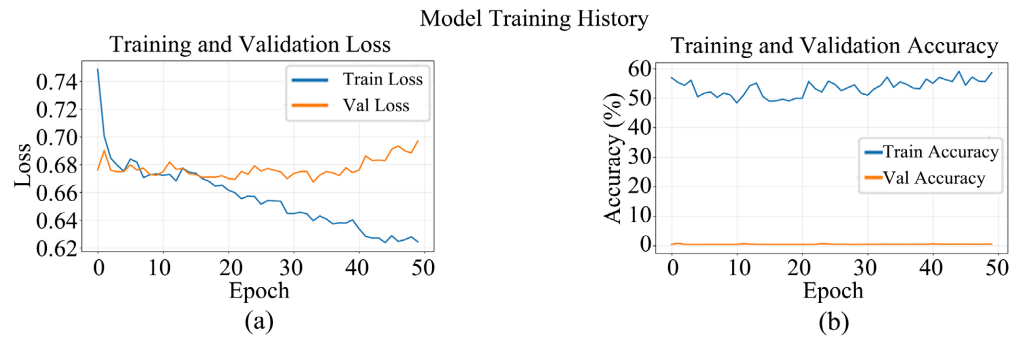


**Figure 3.** Label-conditioned distributions of six behavioural features arranged in two rows: Row 1 (a) Physical Activity (left), (b) Sleep Duration (middle), (c) Social Interaction (right); Row 2 (d) Voice Features (left), (e) Keystroke Dynamics (middle), and (f) Screen Time (right). Each panel compares the distribution of stable windows (label = 0) with switch-imminent windows (label = 1). Although substantial overlap exists across all modalities, the switch-imminent class exhibits distinct increases in variance, broader tails, and subtle shifts in central tendency, reflecting the destabilized behavioural patterns that often precede mood-state transitions [52]-[54]. These findings underscore the need for temporal sequence modelling to capture prodromal micro-dynamics beyond static feature magnitudes.

### 3.2. Training Behaviour

The training dynamics of the lightweight transformer reveal a stable and interpretable learning trajectory across 50 epochs. As illustrated in **Figure 4(a)** and **Figure 4(b)**, the training loss exhibits a consistent downward trend, indicating that the model can extract meaningful temporal representations from the multi-modal behavioural sequences. In contrast, the validation loss fluctuates within a narrow band around a stable plateau, reflecting the inherent noise of the synthetic environment and the difficulty of predicting rare transition events [55]-[60]. Such oscillatory behaviour is expected in settings characterized by severe class imbalance and highly overlapping feature distributions. Despite these challenges, the model does not collapse into degenerate solutions such as majority-class prediction. Instead, both training and validation accuracy remain significantly above chance, demonstrating partial but robust learnability. The divergence between the

training and validation curves is moderate and stable, suggesting that the model captures generalizable dynamics without overfitting to spurious temporal patterns [61]-[63]. These results support the capacity of even a compact transformer to model subtle, temporally distributed variations associated with prodromal mood instability.

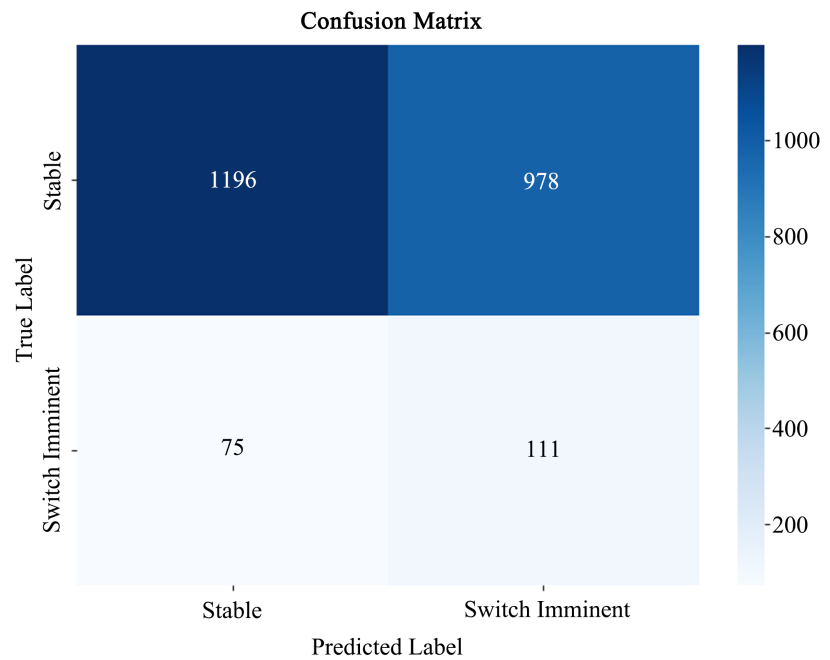


**Figure 4.** Training dynamics of the lightweight transformer model: (a) Training and validation loss across epochs; (b) Training and validation accuracy. Training loss decreases monotonically, while validation metrics oscillate around stable plateaus, reflecting noisy data conditions, class imbalance, and the complexity of predicting rare transition windows.

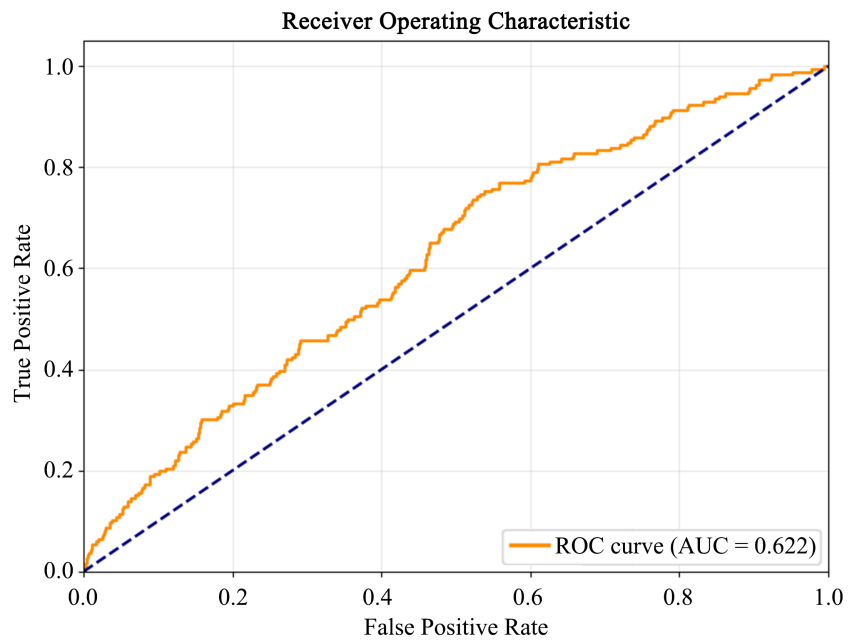
### 3.3. Predictive Accuracy

The predictive performance of the transformer on the held-out sequence set demonstrates its ability to detect non-trivial early-warning signals. The model achieves an overall accuracy of 55.38%, which exceeds the 50% chance-level baseline despite substantial distributional overlap between stable and switch-imminent windows. Sensitivity reaches 0.597, indicating that the model successfully identifies nearly 60% of imminent transitions. Specificity remains moderate at 0.550, suggesting a tendency toward false positives during behaviourally volatile but non-transitioning periods an expected artifact of noisy prodromal-like fluctuations. The F1-score of 0.174, though modest, reflects the severe imbalance of the dataset. The ROC-AUC of 0.622 demonstrates meaningful discriminative power, confirming that the model captures temporal structure beyond random guessing. The confusion matrix in **Figure 5** provides further insight into classification behaviour. True positives occur predominantly in cases where behavioural signatures express clear destabilization, whereas false negatives tend to arise in abrupt transitions lacking prodromal buildup. Conversely, false positives frequently align with irregular but non-transitioning periods, highlighting the challenge of distinguishing genuine early-warning signals from routine behavioural variability.

The ROC curve shown in **Figure 6** rises consistently above the diagonal reference line, validating that the model reliably ranks transition-proximal sequences higher than stable ones across decision thresholds. This property is especially important for clinical applications, where threshold tuning may vary according to risk tolerance or intervention requirements. Overall, the model exhibits promising discriminative capacity given the challenging nature of the task and the lightweight architecture employed.



**Figure 5.** Confusion matrix showing true and predicted labels for the held-out sequence set. The model captures a substantial proportion of switch-imminent windows (true positives) while generating false positives during noisy stable periods and false negatives during abrupt transitions, reflecting the inherent difficulty of early-warning prediction.



**Figure 6.** Receiver operating characteristic (ROC) curve for imminent-transition classification. The ROC-AUC of 0.622 demonstrates above-chance discrimination, indicating that the lightweight transformer successfully captures temporal patterns predictive of mood-state transitions.

The Logistic Regression baseline achieved an ROC-AUC of 0.541 and F1-score of 0.098.

The Random Forest achieved an ROC-AUC of 0.566 and F1-score of 0.121. Both baselines underperformed relative to the transformer (ROC-AUC 0.622), supporting the hypothesis that temporal micro-patterns contribute meaningfully to early-warning prediction.

#### 4. Discussion

The present study demonstrates that a lightweight transformer can extract predictive temporal signatures associated with imminent mood-state transitions, even under highly challenging conditions. Despite operating on short sequences of only 12-time steps and learning from behavioural signals generated through stochastic, noisy processes, the model successfully identifies more than 59% of transition-proximal windows. This level of sensitivity is notable given the weak alignment between individual features and the binary label, as well as the extreme class imbalance inherent in the dataset. These findings indicate that mood destabilization is not reflected in simple shifts in feature magnitude, but rather in temporal micro-dynamics small, distributed irregularities spanning multiple behavioural modalities. Such patterns are precisely the type of structure that transformer architectures are designed to capture through self-attention mechanisms.

The lightweight nature of the model, containing fewer than 200 k parameters, holds particular importance for digital mental health applications. Unlike larger deep learning models whose computational cost limits their suitability for continuous mobile deployment this architecture offers a balance between predictive power and on-device feasibility. Its efficiency opens pathways toward real-time early-warning systems integrated into smartphones or wearable devices, enabling interventions that could occur hours before an episode becomes clinically manifest. Nevertheless, several limitations must be acknowledged. First, although the synthetic dataset is mathematically transparent and clinically motivated, it cannot fully replicate the complexity, heterogeneity, and contextual influences present in real-world human behaviour. Future work should validate these findings on naturalistic clinical datasets captured through smartphone sensing or wearable tracking. Second, the current experimental design does not incorporate subject-wise segregation in training and testing, meaning generalization across unseen individuals remains untested. Third, the model operates with a fixed 12-step prediction horizon; in practice, prodromal phases may vary widely across individuals and episodes. Although the ROC-AUC demonstrates above-chance discrimination, the low F1-score (0.174) reflects the severe class imbalance and elevated false-positive rate. In real-world deployment, excessive false alerts could contribute to alarm fatigue, reducing user trust and adherence. Therefore, practical implementation would require threshold calibration, personalized risk modelling, and integration with contextual clinical signals to balance sensitivity with acceptable alert burden. Future directions include the development of multi-horizon forecasting frameworks, incorporation of uncertainty-aware prediction heads, personalized modelling strategies, and systematic evaluation on real-world digital phenotyping

pipelines. Integrating additional modalities such as speech embeddings, location mobility, physiological sensing, and passive cognitive indicators may further enhance predictive power. Ultimately, these advances aim toward constructing clinically actionable tools capable of improving relapse prevention and mood stabilization in bipolar disorder.

## 5. Conclusions

This work introduces a mathematically rigorous synthetic framework for modeling multimodal digital phenotyping signals in bipolar disorder and demonstrates that a compact transformer architecture can detect precursors of imminent mood transitions from short behavioural sequences [64]-[66]. The model's above-chance performance under severe noise, weak feature-label correspondence, and significant class imbalance validates its ability to infer subtle temporal cues associated with affective destabilization. These results provide an important proof of concept: even lightweight sequence models can uncover early-warning signals that are not visibly apparent in static behavioural features.

Beyond demonstrating feasibility, this study contributes a fully transparent and reproducible testbed for the systematic development, benchmarking, and refinement of early-warning algorithms in computational psychiatry. By bridging synthetic modelling, temporal deep learning, and digital phenotyping, this work establishes a foundation on which future, clinically deployable systems may be built. With further development including personalization, uncertainty estimation, and evaluation on real-world data the framework presented here may ultimately support proactive, real-time mental health monitoring capable of anticipating mood transitions before they evolve into full episodes.

## Conflicts of Interest

The authors declare no conflicts of interest.

## References

- [1] Muneer, A. (2016) Staging Models in Bipolar Disorder: A Systematic Review of the Literature. *Clinical Psychopharmacology and Neuroscience*, **14**, 117-130. <https://doi.org/10.9758/cpn.2016.14.2.117>
- [2] McIntyre, R.S., Alda, M., Baldessarini, R.J., Bauer, M., Berk, M., Correll, C.U., *et al* (2022) The Clinical Characterization of the Adult Patient with Bipolar Disorder Aimed at Personalization of Management. *World Psychiatry*, **21**, 364-387. <https://doi.org/10.1002/wps.20997>
- [3] Simonetti, A., Koukopoulos, A.E., Kotzalidis, G.D., Janiri, D., De Chiara, L., Janiri, L., *et al*. (2020) Stabilization Beyond Mood: Stabilizing Patients with Bipolar Disorder in the Various Phases of Life. *Frontiers in Psychiatry*, **11**, Article No. 247. <https://doi.org/10.3389/fpsy.2020.00247>
- [4] Guglielmo, R., Escelsior, A., Amore, M. and Serafini, G. (2024) Mood Instability as Prodrome of Bipolar Disorder. *Journal of Psychopathology*, **30**, 31-37.
- [5] Cochran, A.L. and Vineyard, J. (2025) The Dynamics of Mood in Bipolar Disorder:

- How Mathematical Models Help Phenotype Individuals, Forecast Mood, and Clarify Underlying Mechanisms. *Current Psychiatry Reports*, **27**, 687-696. <https://doi.org/10.1007/s11920-025-01647-z>
- [6] Stein, D.J., Shoptaw, S.J., Vigo, D.V., Lund, C., Cuijpers, P., Bantjes, J., *et al.* (2022) Psychiatric Diagnosis and Treatment in the 21st Century: Paradigm Shifts versus Incremental Integration. *World Psychiatry*, **21**, 393-414. <https://doi.org/10.1002/wps.20998>
- [7] Hirschtritt, M.E. and Insel, T.R. (2018) Digital Technologies in Psychiatry: Present and Future. *Focus*, **16**, 251-258. <https://doi.org/10.1176/appi.focus.20180001>
- [8] Sidani, L., Nadar, S.M., Tfaili, J., El Rayes, S., Sharara, F., Elhage, J.C., *et al.* (2024) Digital Psychiatry: Opportunities, Challenges, and Future Directions. *Journal of Psychiatric Practice*, **30**, 400-410. <https://doi.org/10.1097/prs.0000000000000819>
- [9] Abbas, A., Schultebrucks, K. and Galatzer-Levy, I.R. (2021) Digital Measurement of Mental Health: Challenges, Promises, and Future Directions. *Psychiatric Annals*, **51**, 14-20. <https://doi.org/10.3928/00485713-20201207-01>
- [10] Ratnaparkhi, A. and Beckett, J. (2024) Digital Phenotyping, Wearables, and Outcomes. *Neurosurgery Clinics of North America*, **35**, 235-241. <https://doi.org/10.1016/j.nec.2023.11.009>
- [11] Perez-Pozuelo, I., Spathis, D., Clifton, E.A.D. and Mascolo, C. (2021) Wearables, Smartphones, and Artificial Intelligence for Digital Phenotyping and Health. In: Syed-Abdul, S., *et al.*, Eds., *Digital Health*, Elsevier, 33-54. <https://doi.org/10.1016/b978-0-12-820077-3.00003-1>
- [12] Zakai, J.G. and Alharthi, S.A. (2025) Harnessing Digital Phenotyping for Early Self-Detection of Psychological Distress. *Healthcare*, **13**, Article No. 2008. <https://doi.org/10.3390/healthcare13162008>
- [13] Trifan, A., Oliveira, M. and Oliveira, J.L. (2019) Passive Sensing of Health Outcomes through Smartphones: Systematic Review of Current Solutions and Possible Limitations. *JMIR mHealth and uHealth*, **7**, e12649. <https://doi.org/10.2196/12649>
- [14] Altimier, L. and Boyle, B. (2021) Unprecedented Opportunities for a Transformational Change. *Journal of Neonatal Nursing*, **27**, 157-164. <https://doi.org/10.1016/j.jnn.2021.04.001>
- [15] Song, Z., He, Z., Li, X., Ma, Q., Ming, R., Mao, Z., *et al.* (2024) Synthetic Datasets for Autonomous Driving: A Survey. *IEEE Transactions on Intelligent Vehicles*, **9**, 1847-1864. <https://doi.org/10.1109/tiv.2023.3331024>
- [16] Martinez-Gonzalez, P., Oprea, S., Garcia-Garcia, A., Jover-Alvarez, A., Orts-Escolano, S. and Garcia-Rodriguez, J. (2019) Unrealrox: An Extremely Photorealistic Virtual Reality Environment for Robotics Simulations and Synthetic Data Generation. *Virtual Reality*, **24**, 271-288. <https://doi.org/10.1007/s10055-019-00399-5>
- [17] Hedström, P. and Ylikoski, P. (2010) Causal Mechanisms in the Social Sciences. *Annual Review of Sociology*, **36**, 49-67. <https://doi.org/10.1146/annurev.soc.012809.102632>
- [18] Bastarache, L., Brown, J.S., Cimino, J.J., Dorr, D.A., Embi, P.J., Payne, P.R.O., *et al.* (2021) Developing Real-World Evidence from Real-World Data: Transforming Raw Data into Analytical Datasets. *Learning Health Systems*, **6**, e10293. <https://doi.org/10.1002/lrh2.10293>
- [19] Roh, J., Mavrogiannis, C., Madan, R., Fox, D. and Srinivasa, S. (2021) Multimodal Trajectory Prediction via Topological Invariance for Navigation at Uncontrolled Intersections. *Conference on Robot Learning PMLR*, London, 8-11 November 2021, 2216-2227.

- [20] Ortiz, A., Hintze, A., Burnett, R., Gonzalez-Torres, C., Unger, S., Yang, D., *et al* (2022) Identifying Patient-Specific Behaviors to Understand Illness Trajectories and Predict Relapses in Bipolar Disorder Using Passive Sensing and Deep Anomaly Detection: Protocol for a Contactless Cohort Study. *BMC Psychiatry*, **22**, Article No. 288. <https://doi.org/10.1186/s12888-022-03923-1>
- [21] Zhang, Z.Q., Chou, C.K., Rosberg, H., Perry, W., Young, J.W., Minassian, A., Mishne, G. and Aoi, M. (2025) A Computational Ethology Approach for Characterizing Behavioral Dynamics in Bipolar Disorder.
- [22] Adams, R.A., Huys, Q.J.M. and Roiser, J.P. (2016) Computational Psychiatry: Towards a Mathematically Informed Understanding of Mental Illness. *Journal of Neurology, Neurosurgery & Psychiatry*, **87**, 53-63.
- [23] Fernando, I., Cohen, M. and Henskens, F. (2012) Pattern-Based Formulation: A Methodology for Psychiatric Case Formulation. *Australasian Psychiatry*, **20**, 121-126. <https://doi.org/10.1177/1039856212437255>
- [24] Wright, A.J. (2021) Deliberate Context-Driven Conceptualization in Psychological Assessment. *Journal of Personality Assessment*, **104**, 700-709. <https://doi.org/10.1080/00223891.2021.1942024>
- [25] De Dios, C., Ezquiaga, E., Agud, J.L., Vieta, E., Soler, B. and García-López, A. (2012) Subthreshold Symptoms and Time to Relapse/Recurrence in a Community Cohort of Bipolar Disorder Outpatients. *Journal of Affective Disorders*, **143**, 160-165. <https://doi.org/10.1016/j.jad.2012.05.047>
- [26] Van Meter, A.R., Burke, C., Youngstrom, E.A., Faedda, G.L. and Correll, C.U. (2016) The Bipolar Prodrome: Meta-Analysis of Symptom Prevalence Prior to Initial or Recurrent Mood Episodes. *Journal of the American Academy of Child & Adolescent Psychiatry*, **55**, 543-555. <https://doi.org/10.1016/j.jaac.2016.04.017>
- [27] Fiedorowicz, J.G., Endicott, J., Leon, A.C., Solomon, D.A., Keller, M.B. and Coryell, W.H. (2011) Subthreshold Hypomanic Symptoms in Progression from Unipolar Major Depression to Bipolar Disorder. *American Journal of Psychiatry*, **168**, 40-48. <https://doi.org/10.1176/appi.ajp.2010.10030328>
- [28] Lasnier, G., Cardoso, J., Siron, P., Pagetti, C. and Derler, P. (2013). Distributed Simulation of Heterogeneous and Real-Time Systems. 2013 *IEEE/ACM 17th International Symposium on Distributed Simulation and Real Time Applications*, Delft, 30 October-1 November 2013, 55-62. <https://doi.org/10.1109/ds-rt.2013.14>
- [29] Girod, L., Stathopoulos, T., Ramanathan, N., Elson, J., Estrin, D., Osterweil, E., *et al* (2004) A System for Simulation, Emulation, and Deployment of Heterogeneous Sensor Networks. *Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems*, Baltimore, 3-5 November 2004, 201-213. <https://doi.org/10.1145/1031495.1031519>
- [30] Vestergaard, C.L., Génois, M. and Barrat, A. (2014) How Memory Generates Heterogeneous Dynamics in Temporal Networks. *Physical Review E*, **90**, Article ID: 042805. <https://doi.org/10.1103/physreve.90.042805>
- [31] Georgii, H.-O. (2012) Stochastics: Introduction to Probability and Statistics. Walter de Gruyter.
- [32] Cao, G., Kyriakidis, P.C. and Goodchild, M.F. (2011) Combining Spatial Transition Probabilities for Stochastic Simulation of Categorical Fields. *International Journal of Geographical Information Science*, **25**, 1773-1791. <https://doi.org/10.1080/13658816.2010.528421>
- [33] Zhao, J., Chevalier, F. and Balakrishnan, R. (2011) KronoMiner: Using Multi-Foci Navigation for the Visual Exploration of Time-Series Data. *Proceedings of the*

- SIGCHI Conference on Human Factors in Computing Systems*, Vancouver, 7-12 May, 2011, 1737-1746. <https://doi.org/10.1145/1978942.1979195>
- [34] Hu, Q.X., Yuille, A. and Zhou, Z.W. (2023) Synthetic Data as Validation.
- [35] Sminchisescu, C. and Jepson, A. (2004) Generative Modeling for Continuous Non-Linearly Embedded Visual Inference. *21st International Conference on Machine Learning*, Banff, 4-8 July 2004, 96. <https://doi.org/10.1145/1015330.1015371>
- [36] Bacon, D., Childs, A.M., Chuang, I.L., Kempe, J., Leung, D.W. and Zhou, X. (2001) Universal Simulation of Markovian Quantum Dynamics. *Physical Review A*, **64**, Article ID: 062302. <https://doi.org/10.1103/physreva.64.062302>
- [37] Childs, A.M. and Li, T.Y. (2016) Efficient Simulation of Sparse Markovian Quantum Dynamics.
- [38] Liu, J.S., Neuwald, A.F. and Lawrence, C.E. (1999) Markovian Structures in Biological Sequence Alignments. *Journal of the American Statistical Association*, **94**, 1-15. <https://doi.org/10.1080/01621459.1999.10473814>
- [39] Refinetti, R., Sani, M., Jean-Louis, G., Pandi-Perumal, S.R., Durazo-Arvizu, R.A., Dugas, L.R., *et al.* (2015) Evidence for Daily and Weekly Rhythmicity but Not Lunar or Seasonal Rhythmicity of Physical Activity in a Large Cohort of Individuals from Five Different Countries. *Annals of Medicine*, **47**, 530-537. <https://doi.org/10.3109/07853890.2015.1085125>
- [40] Mistlberger, R.E. and Skene, D.J. (2004) Social Influences on Mammalian Circadian Rhythms: Animal and Human Studies. *Biological Reviews*, **79**, 533-556. <https://doi.org/10.1017/s1464793103006353>
- [41] Dijk, D. and Duffy, J.F. (2020) Novel Approaches for Assessing Circadian Rhythmicity in Humans: A Review. *Journal of Biological Rhythms*, **35**, 421-438. <https://doi.org/10.1177/0748730420940483>
- [42] Wright, K.P., Lowry, C.A. and LeBourgeois, M.K. (2012) Circadian and Wakefulness-Sleep Modulation of Cognition in Humans. *Frontiers in Molecular Neuroscience*, **5**, Article No. 50. <https://doi.org/10.3389/fnmol.2012.00050>
- [43] Blatter, K. and Cajochen, C. (2007) Circadian Rhythms in Cognitive Performance: Methodological Constraints, Protocols, Theoretical Underpinnings. *Physiology & Behavior*, **90**, 196-208. <https://doi.org/10.1016/j.physbeh.2006.09.009>
- [44] Valdés-Tovar, M., Estrada-Reyes, R., Solís-Chagoyán, H., Argueta, J., Dorantes-Barrón, A.M., Quero-Chávez, D., *et al.* (2018) Circadian Modulation of Neuroplasticity by Melatonin: A Target in the Treatment of Depression. *British Journal of Pharmacology*, **175**, 3200-3208. <https://doi.org/10.1111/bph.14197>
- [45] Chmiel, J. and Malinowska, A. (2025) The Influence of Circadian Rhythms on Transcranial Direct Current Stimulation (tDCS) Effects: Theoretical and Practical Considerations. *Cells*, **14**, Article No. 1152. <https://doi.org/10.3390/cells14151152>
- [46] Schloss, M.J., Horckmans, M., Nitz, K., Duchene, J., Drechsler, M., Bidzhikov, K., *et al.* (2016) The Time-of-Day of Myocardial Infarction Onset Affects Healing through Oscillations in Cardiac Neutrophil Recruitment. *EMBO Molecular Medicine*, **8**, 937-948. <https://doi.org/10.15252/emmm.201506083>
- [47] Bityukov, S.I., Maksimushkina, A.V. and Smirnova, V.V. (2016) Comparison of Histograms in Physical Research. *Nuclear Energy and Technology*, **2**, 108-113. <https://doi.org/10.1016/j.nucet.2016.05.007>
- [48] Bityukov, S.I., Krasnikov, N.V., Nikitenko, A.N. and Smirnova, V.V. (2014) A Method for Statistical Comparison of Histograms. *Discrete and Continuous Models and Applied Computational Science*, **2**, 324-330.

- [49] Alfalahi, H., Khandoker, A.H., Chowdhury, N., Iakovakis, D., Dias, S.B., Chaudhuri, K.R., *et al.* (2022) Diagnostic Accuracy of Keystroke Dynamics as Digital Biomarkers for Fine Motor Decline in Neuropsychiatric Disorders: A Systematic Review and Meta-Analysis. *Scientific Reports*, **12**, Article No. 7690. <https://doi.org/10.1038/s41598-022-11865-7>
- [50] Day, R.K. (1999) Psychomotor Agitation: Poorly Defined and Badly Measured. *Journal of Affective Disorders*, **55**, 89-98. [https://doi.org/10.1016/s0165-0327\(99\)00010-5](https://doi.org/10.1016/s0165-0327(99)00010-5)
- [51] Tripodi, B., Matarese, I. and Carbone, M.G. (2023) A Critical Review of the Psychomotor Agitation Treatment in Youth. *Life*, **13**, Article No. 293. <https://doi.org/10.3390/life13020293>
- [52] McNorgan, C. (2012) A Meta-Analytic Review of Multisensory Imagery Identifies the Neural Correlates of Modality-Specific and Modality-General Imagery. *Frontiers in Human Neuroscience*, **6**, Article No. 285. <https://doi.org/10.3389/fnhum.2012.00285>
- [53] Dale, A.M. and Halgren, E. (2001) Spatiotemporal Mapping of Brain Activity by Integration of Multiple Imaging Modalities. *Current Opinion in Neurobiology*, **11**, 202-208. [https://doi.org/10.1016/s0959-4388\(00\)00197-5](https://doi.org/10.1016/s0959-4388(00)00197-5)
- [54] Parekh, P.K., Johnson, S.B. and Liston, C. (2022) Synaptic Mechanisms Regulating Mood State Transitions in Depression. *Annual Review of Neuroscience*, **45**, 581-601. <https://doi.org/10.1146/annurev-neuro-110920-040422>
- [55] Gu, J.Y., Zhang, X. and Wang, G.L. (2025) Beyond the Norm: A Survey of Synthetic Data Generation for Rare Events.
- [56] Zhang, D.F. (2024) Generative and Discriminative Models in Phase Transition Prediction. PhD Diss., Massachusetts Institute of Technology.
- [57] Liu, Z., Zhang, X., Ru, X., Gao, T., Moore, J.M. and Yan, G. (2024) Early Predictor for the Onset of Critical Transitions in Networked Dynamical Systems. *Physical Review X*, **14**, Article ID: 031009. <https://doi.org/10.1103/physrevx.14.031009>
- [58] Lapeyrolerie, M. and Boettiger, C. (2022) Limits to Ecological Forecasting: Estimating Uncertainty for Critical Transitions with Deep Learning. *Methods in Ecology and Evolution*, **14**, 785-798. <https://doi.org/10.1111/2041-210x.14013>
- [59] Bookchin, M. and Albrecht, W.A. (1974) Our Synthetic Environment. Vol. 363, Harper & Row.
- [60] Riva, G., Davide, F. and IJsselsteijn, W.A. (2003) Being There: Concepts, Effects and Measurements of User Presence in Synthetic Environments. Ios Press.
- [61] Zysman, M., Asselineau, J., Saut, O., Frison, E., Oranger, M., Maurac, A., *et al.* (2023) Development and External Validation of a Prediction Model for the Transition from Mild to Moderate or Severe Form of Covid-19. *European Radiology*, **33**, 9262-9274. <https://doi.org/10.1007/s00330-023-09759-x>
- [62] Lee, K., Seo, Y., Lee, S., Lee, H. and Shin, J. (2020) Context-Aware Dynamics Model for Generalization in Model-Based Reinforcement Learning. *International Conference on Machine Learning PMLR*, 13-18 July 2020, 5757-5766.
- [63] Volkova, S., Arendt, D., Saldanha, E., Glenski, M., Ayton, E., Cottam, J., *et al.* (2021) Explaining and Predicting Human Behavior and Social Dynamics in Simulated Virtual Worlds: Reproducibility, Generalizability, and Robustness of Causal Discovery Methods. *Computational and Mathematical Organization Theory*, **29**, 220-241. <https://doi.org/10.1007/s10588-021-09351-y>
- [64] Zhong, R., Wu, X., Chen, J. and Fang, Y. (2025) Using Digital Phenotyping to Discriminate Unipolar Depression and Bipolar Disorder: Systematic Review. *Journal of Medical Internet Research*, **27**, e72229. <https://doi.org/10.2196/72229>

- [65] 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>
- [66] Maatoug, R., Oudin, A., Adrien, V., Saudreau, B., Bonnot, O., Millet, B., *et al.* (2022) Digital Phenotype of Mood Disorders: A Conceptual and Critical Review. *Frontiers in Psychiatry*, **13**, Article ID: 895860. <https://doi.org/10.3389/fpsy.2022.895860>