

Discovering wearable sensor-derived temporal digital phenotypes and their association with hypertensive disorders in pregnancy

Nidhi Soley

nsoley1@jhu.edu

Johns Hopkins University

Angie Jelin

Johns Hopkins Medicine

Vishal M. Patel

Johns Hopkins University

Casey O. Taylor

Johns Hopkins University School of Medicine

Article

Keywords: wearable sensors, machine learning, multimodal sensor time series, continuous health monitoring, reproductive health monitoring, hypertensive disorders of pregnancy, digital phenotyping

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9452274/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Discovering wearable sensor-derived temporal digital phenotypes and their association with hypertensive disorders in pregnancy

Nidhi Soley^{1,2*}, Angie Jelin³, Vishal M. Patel⁴, Casey O. Taylor^{1,2,5}

1. Institute for Computational Medicine, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD 21218, United States

2. Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, MD 21218, United States

3. Department of Gynecology and Obstetrics, Johns Hopkins University School of Medicine, Baltimore, MD 21218, United States

4. Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218, United States

5. Department of Medicine, Division of General Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21218, United States

*Corresponding author:

Nidhi Soley, MS

Hackerman Hall, Rm 318, 3101 Wyman Park Dr, Baltimore, MD 21218

Email: nsoley1@jhu.edu

Telephone: [+18016717776]

Abstract

Hypertensive disorders of pregnancy (HDP) remain a leading cause of maternal morbidity, and episodic clinic measurements often miss early physiologic changes. We investigated whether longitudinal wearable data could identify interpretable physiological clusters (digital phenotypes), and trimester-to-trimester changes associated with subsequent HDP. We conducted a retrospective cohort study using pregnant participants in the NIH All of Us Research Program with wearable (Fitbit) and electronic health record data. Wearable measurements were obtained from 14-day windows preceding the end of each trimester and postpartum. We trained an attention-based recurrent autoencoder separately for physical activity, sleep, and pulse rate variability, and then clustered the latent embeddings using Gaussian mixture models to derive modality-specific digital phenotypes. We examined whether transitions in cluster membership between adjacent periods were associated with HDP by the next boundary. Activity clusters were externally validated in the Washington University Prematurity Research Cohort. In the discovery cohort (272 participants; 1,088 windows), we identified three activity, two sleep, and two PRV clusters with moderate-to-strong latent separation (silhouette 0.63–0.65; area under the curve 0.69–0.72). The clusters reflected gradients in activity and sedentary behavior, sleep quantity and architecture, and autonomic tone. Transitions toward lower activity, poorer sleep, and reduced autonomic variability were associated with increased odds of subsequent HDP. In external validation (131 pregnancies; 393 windows), activity clusters retained similar structure and discrimination (silhouette 0.54; area under the curve 0.68), and assignments were stable under resampling. These wearable sensor-derived digital phenotypes support the potential of machine learning-enabled continuous monitoring to complement routine prenatal care.

Keywords: wearable sensors; machine learning; multimodal sensor time series; continuous health monitoring; reproductive health monitoring; hypertensive disorders of pregnancy; digital phenotyping

Introduction

HDP comprises blood-pressure disorders that arise in or affect pregnancy, including gestational hypertension, chronic hypertension, and the preeclampsia spectrum (including HELLP and eclampsia)¹. HDP is a leading cause of maternal and perinatal morbidity and mortality¹. In the U.S., HDP accounts for approximately 7% of maternal deaths from 2017–2019^{1,2}, with many HDP-related fatalities occurring in the days and weeks around delivery. Because four out of five maternal deaths are considered preventable³, there is an urgent need for earlier detection and prevention strategies.

Wearable devices, including fitness trackers and smartwatches, provide high-frequency measures of physical activity (PA), sleep, and pulse rate, offering a view of maternal physiology between prenatal visits^{4,5}. Prior work suggests that reduced activity⁶, sleep disturbance⁷, and altered autonomic function^{8,9} may precede hypertensive complications. Epidemiological and mechanistic studies further support these relationships. PA during pregnancy has been associated with a reduced risk of preeclampsia in meta-analytic evidence⁶ and examined in relation to preeclampsia and gestational hypertension in case-control studies¹⁰. Exercise has also been studied as a modifiable factor affecting blood pressure during pregnancy¹¹. Systematic reviews and meta-analyses have linked sleep disturbance to preeclampsia^{7,12} and other adverse maternal and fetal outcomes¹³, while autonomic dysregulation has been reported in preeclampsia¹⁴. However, existing studies generally do not examine HDP-specific wearable phenotypes or changes in these phenotypes across trimester and postpartum boundaries.

Many HDP prediction approaches rely on limited sets of static clinical features rather than continuous time-series data. Recent machine learning-enabled maternal risk assessment models^{15,16}, and broader reviews of artificial intelligence in HDP risk stratification^{17,18} highlight the promise of advanced analytics and the continuing need for clinically meaningful representations of longitudinal physiology. Pregnancy wearable data nevertheless present a challenging machine-learning problem: multimodal streams are dense but noisy, adherence varies across gestation, derived features differ across sensing platforms, and clinically useful outputs must remain interpretable^{4,5,12,19,20}. HDP-associated phenotyping therefore requires robust temporal representation learning, physiologically meaningful subtype discovery, and evaluation across external sensing cohorts.

Deep representation learning is increasingly used in biomedical informatics to compress high-dimensional time series into latent spaces amenable to downstream tasks such as clustering, subtyping, and risk stratification²⁰. Attention mechanisms provide a way to emphasize informative time points²¹, although attention weights should not be interpreted as causal explanations²². Outcome-guided deep temporal clustering has also been proposed to identify disease-progression subtypes²³, supporting the use of outcome-relevant representations for temporal phenotyping. Translational use of clustering additionally requires reproducibility: resampling-based approaches can quantify clustering stability^{23,24}, and agreement metrics such as the adjusted Rand index can compare cluster assignments²⁵.

In this study, we analyzed longitudinal Fitbit data linked with electronic health records (EHR) for pregnant participants in the NIH All of Us Research Program²⁶. We applied an attention-based gated recurrent unit autoencoder (AttentiveGRUAE)²⁷ to learn latent representations of PA, sleep, and pulse rate variability (PRV) from 14-day windows across pregnancy and postpartum, then clustered these representations to identify modality-specific digital phenotypes. Relative to prior pregnancy wearable studies^{4,10,12,13,20}, this work focuses on HDP and emphasizes transitions between learned phenotypes

across trimester and postpartum boundaries rather than only cross-sectional associations. It also complements HDP risk-prediction approaches anchored to sparse clinical measurements^{16,17} by characterizing longitudinal wearable patterns. Specifically, we aimed to derive interpretable physiological clusters, quantify associations between cluster transitions and subsequent HDP at clinically meaningful boundaries, and evaluate the external reproducibility and stability of activity phenotypes in the independent Washington University Prematurity Research Cohort¹⁹.

Results

We applied an attention-based recurrent autoencoder separately to physical activity, sleep, and pulse rate variability, combining reconstruction with an auxiliary HDP classification objective. We then clustered the learned representations using Gaussian mixture models to derive modality-specific digital phenotypes (**Figure 1**).

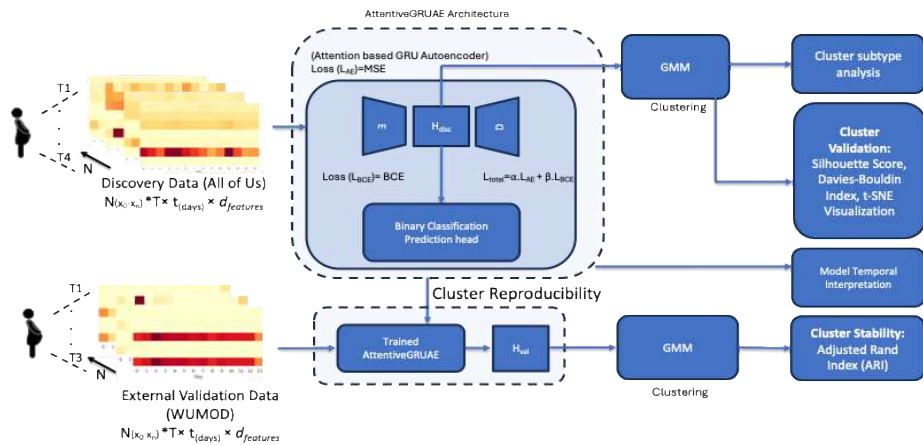


Figure 1. The AttentiveGRUAE pipeline learns wearable-derived digital phenotypes and evaluates their characteristics and reproducibility.

Study timeline and cohort

Of 42,839 EHR-identified pregnancy episodes, 541 had linked Fitbit data within the study period and metadata for outcome labeling (**Figure 2 a,b**). Among these, 272 patients had complete longitudinal coverage for all analysis windows and wearable metrics (14-day windows before the end of T1, T2, T3, and postpartum), forming 1,088 total sequences (272 x 4 windows). By term status, completeness comprised 140 full-term episodes and 132 preterm episodes; no participant contributed both. HDP prevalence increased across gestation and postpartum, from 16.9% by the end of trimester 1 to 33.1% by postpartum. HDP status was carried forward across periods; these percentages therefore represent cumulative HDP status.

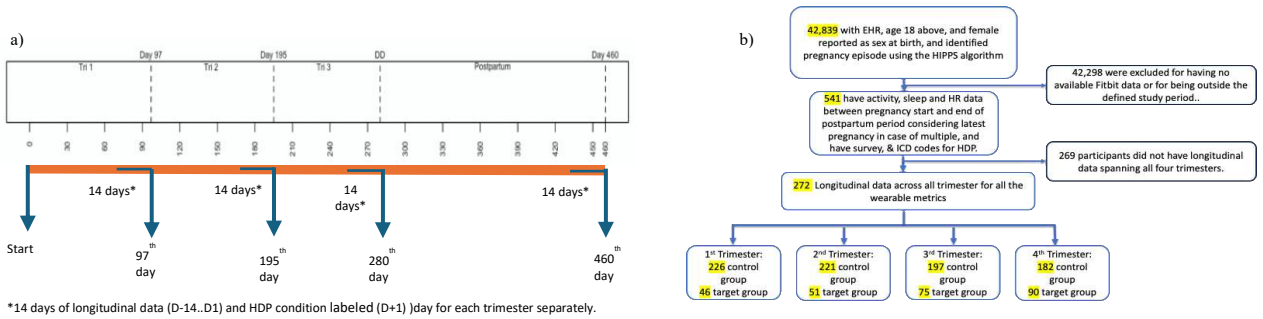


Figure 2. Study design, cohort selection, and data windowing.

Latent structure and internal validation

The learned representations yielded distinct clusters for each modality. After projecting 14-day windows (Supplementary Table 5) into AttentiveGRUAE latent space and fitting a GMM, internal validity indices indicated meaningful structure, corroborated but not inferred from t-SNE visualizations (Figure 3 a,b,c). Silhouette scores and Davies–Bouldin indices (DBI) describe separation in the learned latent space, whereas AUC values describe HDP discrimination by the auxiliary classification head. The t-SNE projections in Figure 3 are provided only as two-dimensional visualizations and were not used to select the number of clusters or assess cluster validity. BIC supported 3 activity clusters and 2 clusters each for sleep and PRV (Supplementary Figure 1a-c). AUC from the auxiliary prediction head on a held-out test set was strongest for sleep and PRV (Table 1). MSE was lowest for sleep. Participant-level 5-fold cross-validation produced similar results, with mean AUC values of 0.694, 0.716, and 0.711, respectively. Cross-validation summary results are reported in Supplementary Table 6.

Table 1: Model reconstruction and clustering performance across modalities.

Data	MSE	AUC	Silhouette	DBI
Discovery				
Physical Activity	0.066	0.692	0.650	0.590
Sleep	0.006	0.720	0.638	1.217
PRV (SDANN)	0.060	0.720	0.650	1.901
External validation				
Physical Activity	0.106	0.681	0.543	0.574

a) Physical Activity

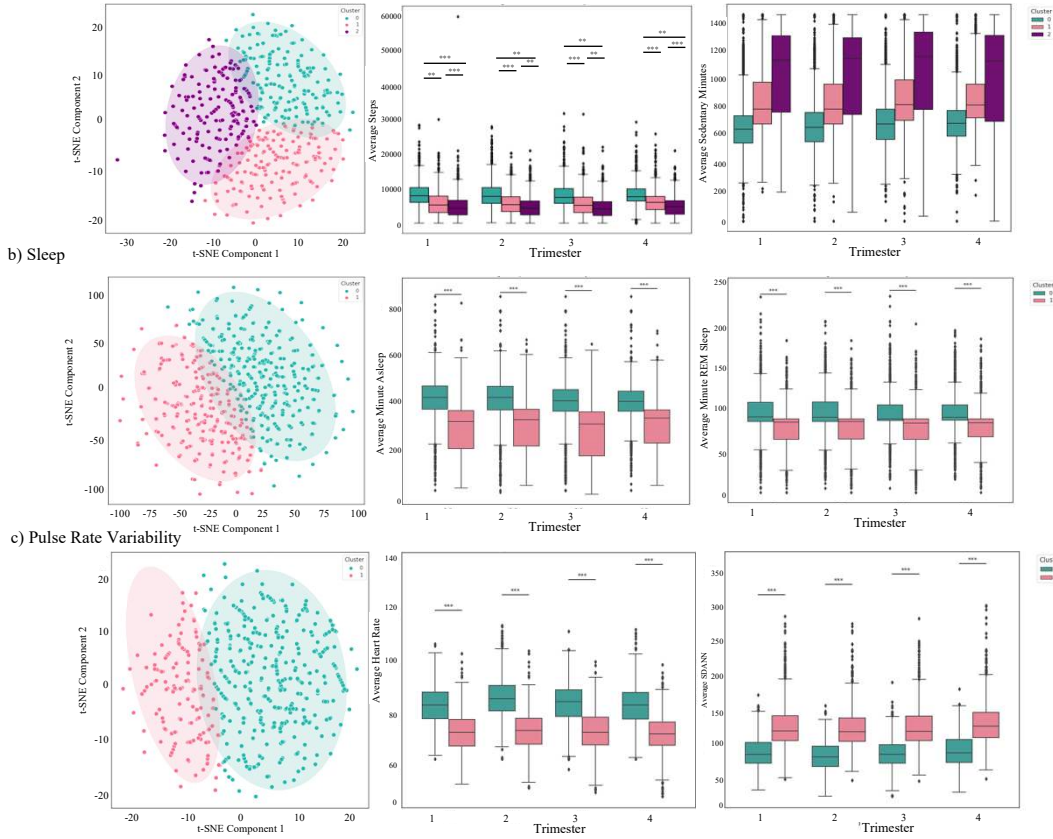


Figure 3. Learned wearable representations distinguish clusters with different physiological profiles across pregnancy and postpartum.

Cluster characteristics

PA ($K=3$; mode-cluster assignment: $C0=103$, $C1=85$, $C2=84$): Clusters differed strongly in activity and sedentary behavior (**Supplementary Table 1**). $C0$ had the most favorable profile (e.g., higher daily steps and more light activity, lower sedentary time), whereas $C2$ showed the inverse. After FDR adjustment, multiple PA features remained highly significant (e.g., sedentary minutes and lightly active minutes with large η^2). PRV (SDANN) was modestly higher in the high-activity group and lower in the low-activity group. Pre-pregnancy chronic conditions (hypertension/diabetes/obesity) were over-represented in the low-activity cluster, while age group, race/ethnicity, preterm status, and gestational-age distribution showed no statistically significant differences after FDR control.

Sleep ($K=2$; mode-cluster assignment: $C0=167$, $C1=105$): Sleep clusters were distinguished by quantity and architecture (**Supplementary Table 2**). $C0$ (“longer and more sleep”) had higher minutes asleep and REM sleep, higher deep sleep, and lower awake/restless time compared with $C1$ (“shorter sleep”), with large effect sizes for key sleep features. After FDR correction, cross-domain PA and PRV differences

were small and mostly non-significant; demographic distributions, preterm status, and comorbidity rates did not differ significantly after correction.

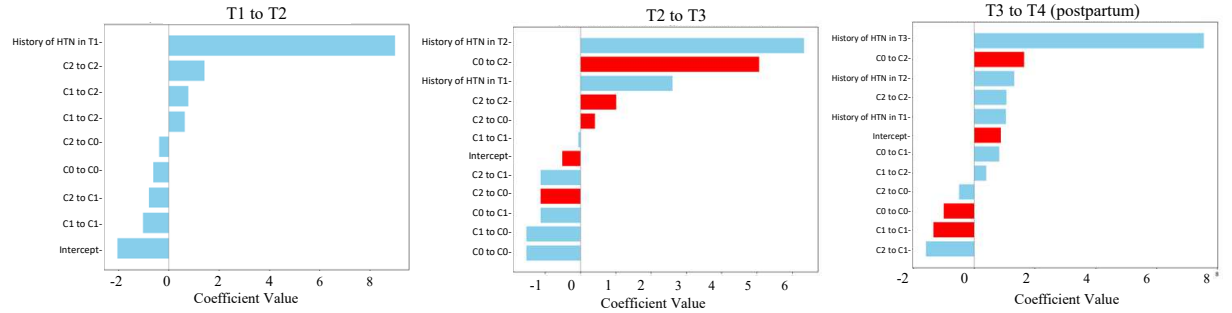
PRV (K=2; mode-cluster assignment): PRV clusters captured a strong autonomic gradient (**Supplementary Table 3**). The higher-PRV cluster (C1) had markedly higher SDANN and substantially lower mean heart rate (and lower minimum/maximum heart rates). Activity and sleep quantities were similar across PRV clusters after FDR correction, and demographic characteristics were broadly comparable after adjustment. A small association with non-hypertensive gestational comorbidities reached the FDR threshold.

Cluster transitions and HDP risk

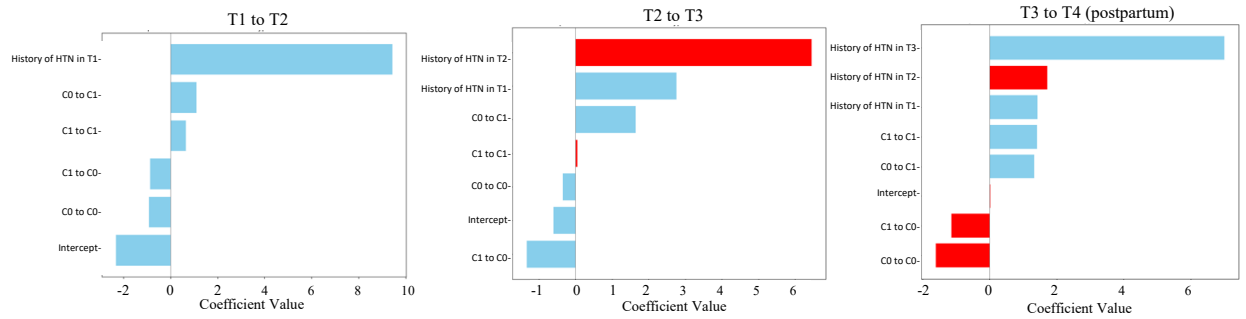
We examined whether changes in cluster membership between adjacent periods were associated with HDP status at the subsequent boundary, adjusting for prior hypertension. In boundary-specific models, prior chronic hypertension was strongly associated with HDP at each interval. Beyond this baseline risk, the clearest transition-related effects were observed for worsening activity phenotypes (C0 to C2), particularly between trimester 2 and trimester 3 and from trimester 3 to postpartum, where transitions from more active to less active states showed positive coefficients, while transitions toward more favorable activity states tended to be negative. For sleep, worsening transitions toward fragmented sleep (C0 to C1) were also associated with increased odds of HDP, with the strongest effects again emerging in later intervals rather than between trimester 1 and trimester 2. PRV transition effects were comparatively smaller and less consistent, although lower-PRV trajectories in later pregnancy showed positive associations (C1 to C0) with HDP risk (**Figure 4 a,b,c**).

At the posterior-probability threshold of 0.70, the activity and sleep transition associations remained directionally consistent and statistically significant when the analysis was restricted to high-confidence cluster assignments (**Supplementary Table 7**). PRV transition estimates remained smaller and less consistent.

a) Physical Activity (C0= Very Active; C1= Moderately Active; C2= Less Active). Blue bars indicate statistically significant coefficients ($p<0.05$)



b) Sleep (C0= Restorative Sleep; C1= Fragmented Sleep). Blue bars indicate statistically significant coefficients ($p<0.05$)



c) PRV (C0= Lower HRV; C1= Higher HRV). Blue bars indicate statistically significant coefficients ($p<0.05$)

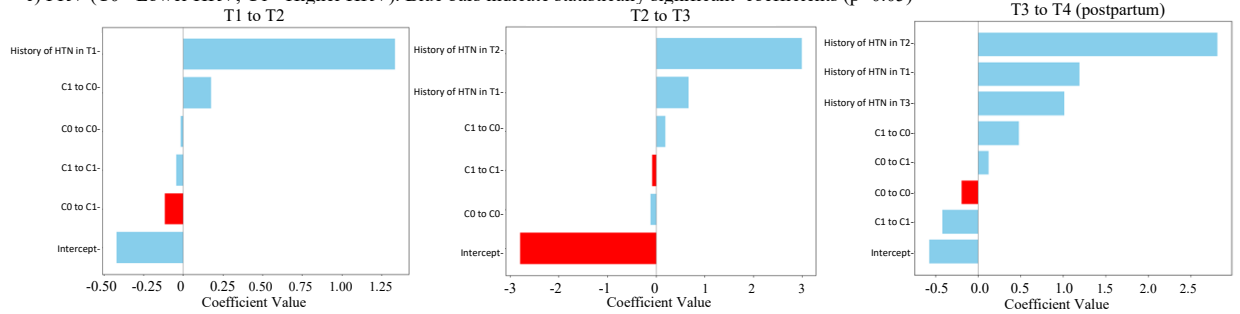


Figure 4. Cluster transitions are associated with HDP across successive pregnancy and postpartum boundaries.

Cross-physiology concordance

Concordance categories distinguished participants with (a) *Concordant Healthy*: all three modalities indicated favorable physiology (e.g., higher PA, less sedentary time, higher PRV, less fragmented sleep, greater deep-sleep duration); (b) *Concordant Unhealthy*: the opposite pattern (e.g., low activity, more sedentary time, reduced PRV, shorter, fragmented sleep); (c) *Discordant*: any other combination. Most pregnancies were discordant across modalities. Across pairings and the three-modality integration, concordance analyses showed a graded pattern of risk: HDP prevalence was lowest in concordant healthy profiles, highest in concordant unhealthy profiles, and intermediate in discordant profiles (**Table 2**). This pattern was statistically significant for activity plus PRV concordance and for the 3-modality integration.

Table 2. HDP Prevalence Across Cross-Physiology Phenotype Concordance Groups				
Cluster Source	Concordant Healthy (HDP Positive/Total)	Concordant Unhealthy (HDP Positive/Total)	Discordant (HDP Positive/Total)	p-value
PRV + PA	<20/51 = 21%	25/57 = 44%	57/164 = 35%	<0.05
Sleep + PA	23/75 = 31%	<20/36 = 37%	52/161 = 32%	0.183
PRV + Sleep	<20/68 = 25%	25/63 = 40%	49/141 = 35%	0.264
PRV+Sleep+PA	<20/38=14%	<20/21=52%	73/213 =34%	<0.05

Temporal interpretability

Attention profiles differed by modality. When attention weights were averaged over sequences, PA attention generally emphasized later days within the observation window. Sleep attention varied across periods and outcome groups, with later peaks in some comparisons (**Figure 5 a,b**). PRV attention was comparatively flat, indicating the model relies on distributed autonomic information accumulated across many days rather than any single day (**Figure 5 c**). These patterns were stable across trimesters for all three modality.

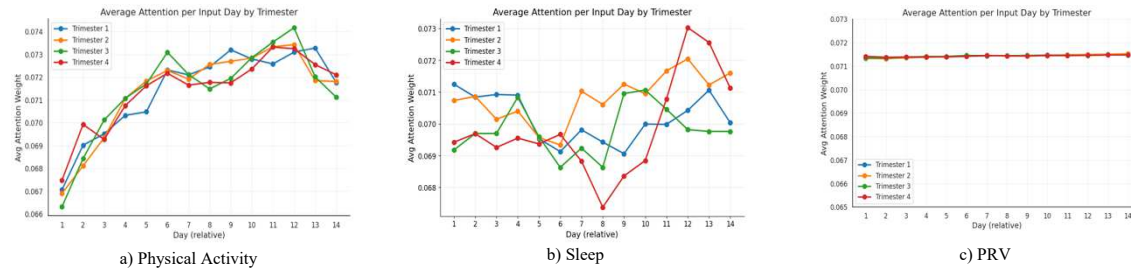


Figure 5. Average temporal attention profiles differ across wearable modalities and pregnancy periods.

When stratified by outcome, early pregnancy (T1/T2) HDP-positive windows showed earlier attention peaks (around Days 6-7), whereas HDP-negative windows peaked later (often Days 9-12) (**Supplementary Figures 2,3,4**). In late pregnancy and postpartum (T3/T4), attention in HDP-positive episodes shifted toward the last window days (e.g., Day 12 in T4), while non-HDP showed flatter late-window profiles. PRV peak-day differences by outcome were small and variable across trimesters.

External validation and stability

In the external cohort, 131 pregnancies contributed 393 activity windows across trimesters 1-3. Activity distributions were qualitatively similar across cohorts (**Supplementary Table 4**). In external validation, cluster separation remained moderate (silhouette 0.543 vs 0.650 in discovery; DBI 0.574 vs 0.590) and discriminative performance generalized with minimal loss (AUC 0.681 vs 0.692), while reconstruction error increased (MSE 0.106 vs 0.066) (**Table 1**). Cluster assignments were highly reproducible under participant resampling (mean ARI 0.948 across leave-out experiments; **Supplementary Figure 5**).

Discussion

Using longitudinal wearable data collected across pregnancy and postpartum, we identified digital phenotypes spanning PA, sleep, and PRV (**Figure 3 a-c**). We evaluated their association with HDP. Our findings support the clinical relevance of these digital phenotypes. First, each physiologic domain yielded distinct latent subgroups with coherent behavioral and autonomic profiles. Second, concordance across domains revealed a graded pattern of risk, with the highest HDP burden observed among individuals with consistently unfavorable physiology across activity, sleep, and PRV. Third, transitions in cluster membership across gestation were associated with near-term HDP risk. Finally, external validation demonstrated preservation of latent structure and stable cluster assignments despite device and protocol differences, supporting the transportability of the overall clustering framework.

Longitudinally, transitions between clusters were informative for near-term HDP risk (**Figure 4 a-c**). These findings support a view of HDP risk as a dynamic process, in which loss of physiological stability over time may be more informative than absolute values at discrete time points. Worsening trajectories across sleep, activity, and autonomic regulation reflect diminishing adaptive reserve during pregnancy, consistent with the increasing cardiovascular and metabolic demands of gestation²⁸⁻³¹. These findings also extend prior work linking reduced activity, sleep disturbance, and autonomic dysfunction to hypertensive complications in pregnancy^{4,7-14,20,32}. This perspective reinforces the limitation of snapshot-based screening approaches and motivates longitudinal monitoring to identify early deviations from expected physiological adaptation.

The concordance results further support the concept that HDP reflects multi-system dysregulation rather than impairment within a single behavioral or autonomic domain (**Table 2**). Alignment of unfavorable profiles across autonomic, behavioral, and sleep-related processes is consistent with shared upstream mechanisms, including heightened sympathetic activity, vascular dysfunction, and inflammatory stress^{4,10,11,13,14}. Together, these observations support a systems-level model of HDP risk and highlight the potential of continuous, wearable-based phenotyping to capture evolving multi-domain vulnerability before clinical manifestation.

Model interpretability using attention weight showed that aggregated attention concentrated on the last few days of each window for PA and sleep, consistent with the clinical impression that short-horizon behavioral decompensation can precede HDP documentation³² (**Figure 5 a,b**). In contrast, PRV attention was comparatively flat across the 14 days, suggesting reliance on more chronic autonomic patterns rather than brief excursions (**Figure 5 c**). This profile is coherent with PRV's multiday structure and role as an integrative marker of autonomic balance³³, and with observations of reduced PRV in hypertensive disorders and postpartum pre-eclampsia^{9,14}.

External validation in an independent actigraphy cohort supported generalizability. Despite domain shift (different device, protocol), cluster separation was preserved, discrimination declined only modestly (AUC 0.68 vs 0.69 in discovery), and reconstruction error increased as expected under distribution shift (**Table 1**). The external assignments remained highly stable under resampling (**Supplementary Figure 5**). These findings suggest that the learned representation captures reproducible structure beyond cohort-specific noise and may be reusable across data sources with limited recalibration.

This study has several limitations. The sample size and participants with sufficient wearable adherence may differ systematically from the broader pregnant population, introducing selection bias. The discovery cohort was modest for training a recurrent neural network, and the reported results should therefore be interpreted as evidence of feasibility rather than definitive estimates of performance. HDP labels derived from EHR codes and carry-forward rules are susceptible to misclassification and temporal misalignment with physiology. PRV derived from wrist PPG is less precise than ECG. Activity and sleep features can also vary by firmware, algorithms, motion artifacts, nocturnal signal loss, or non-wear time, which could bias features and cluster assignments. The encoder was trained on 14-day windows and evaluated under device and protocol differences, causing increased reconstruction error under domain shift. External validation was limited to activity because the external cohort did not contain the sleep-stage and heart-rate data required to evaluate the sleep and PRV models. External validation of sleep and PRV phenotypes will require a cohort with wearable-derived sleep staging and heart-rate data, which we identify as a priority for future work. In addition, our model treats each 14-day window independently and clusters and compares phenotypes only at predefined trimester and postpartum boundaries. This design was chosen because these boundaries align with points where clinical decisions are made, and because our sample size limited the complexity of models we could fit reliably. A continuous-time latent-state model (e.g., a state-space formulation) could capture this evolution directly and is a natural direction for future work. Combining all the modalities and training multimodal models are also important directions for future work. Finally, attention supports interpretability but is not a causal attribution.

Overall, our study demonstrates that continuous wearable data can be transformed into digital phenotypes and short-horizon trajectory changes that are associated with subsequent HDP documentation. The most actionable signal in our analyses was near-term deterioration in activity and sleep phenotypes, consistent with the possibility that behavioral and recovery patterns shift proximally to clinical recognition. Since activity and sleep are potentially modifiable, these phenotypes may also provide targets for monitoring and supportive behavioral counseling in higher-risk pregnancies. This framework can potentially be applicable as an informatics layer on top of routine prenatal care, using passively collected wearable streams to summarize current physiologic state and flag worsening trajectories between scheduled visits. Future work should prospectively validate these phenotypes, evaluate transportability across devices and populations, and test whether adding longitudinal phenotype and transition features to established clinical predictors improves discrimination, calibration, and clinically meaningful lead time for detection.

Methods

Study design and timeline

We conducted a retrospective study using longitudinal wearable data linked to EHRs from the NIH *All of Us* Research Program²⁶ (discovery dataset). Pregnancy episodes were identified using the Hierarchy and rule-based pregnancy episode Inference features integrated with Pregnancy Progression Signatures (HIPPS) algorithm³⁴, which infers gestational timing from EHR data³⁵.

Episodes <259 days were flagged as preterm; otherwise, full-term. If gestational age was unavailable, the delivery date was set as inferred episode start + 280 days. These definitions follow international guidance defining preterm birth as <37 completed weeks (<259 days)²⁸. Trimesters were defined by obstetric convention, with explicit handling of preterm delivery: T1, day 0–97; T2, day 98–195 or delivery (if preterm); T3, only if delivery occurred after day 195; otherwise, not applicable. Postpartum (T4) began on

delivery date and extended 180 days²⁹. For each period boundary day D (end of T1, T2, T3, and T4), we constructed a 14-day look-back window (days D-14 through D-1), skipped day D to prevent label leakage, and adjudicated HDP status on day D+1. The 14-day window was selected a priori based on internal pilots (7/14/28 days) and alignment with prior actigraphy studies in pregnancy³⁰.

Cohorts and outcome definitions

In the discovery data, we included participants (female sex at birth, age ≥ 18 years) with: (i) linked EHRs and HIPPS-identified pregnancy episodes, (ii) Fitbit activity, sleep, and heart-rate data covering pregnancy onset through the postpartum period, and (iii) survey data or ICD codes sufficient to label HDP within selected windows. We excluded participants without four-period coverage from longitudinal clustering. For individuals with multiple pregnancies, we retained the most recent episode.

The primary binary outcome was the presence of any hypertensive disorder by the end of each period. We anchored ICD-10 CM families O10–O16³¹ and included ICD-9-CM equivalents 642 (0x–7x,9x)³⁶. We did not analyze superimposed hypertension as a separate subtype (O11*). Self-report of chronic hypertension, pre-eclampsia/eclampsia, or in-pregnancy diagnoses (including postpartum) also set HDP as positive for the relevant period. We applied a carry-forward rule; once HDP was documented as positive, it remained positive for all subsequent periods.

Wearable features

In the discovery data, the predictors were daily summary features of PA, sleep, and heart rate-derived PRV from the wearable data from NIH *All of Us*. For each participant and trimester in the discovery data, we formed a 14 x features daily matrix for each trimester. Missing daily aggregates within retained windows were imputed by time-aware linear interpolation per feature. For PA, we selected *sedentary_minutes*, *lightly_active_minutes*, *fairly_active_minutes*, *very_active_minutes*, *steps*, and *calories_out*. Variables highly collinear with these (e.g., *activity_calories*, *marginal_calories*) were removed. For sleep, from Fitbit's sleep/stage summaries, we retained *minute_asleep*, *minute_awake*, *sleep_fragmented*, *minute_deep*, *minute_light*, *minute_rem*, and *minute_wake*, removing *minute_in_bed* and *minute_restless* as redundant. We calculated PRV from Fitbit-derived pulse-rate measurements following prior *All of Us* aligned processing³⁷ and standard guidance³³. We computed consecutive 5-min mean heart rates and converted each to a mean inter-beat interval (IBI; RR interval). The RR interval was derived from the relationship that 1 millisecond per minute equals 60,000. Thus, to find the duration of one heartbeat (RR interval) in milliseconds, we used:

$$\text{AverageRR} = \frac{60,000}{\text{AverageHR}} \text{ms} \quad (1)$$

where the average is the mean pulse rate (beats per minute) over that 5-min window. Using this, we then calculated the standard deviation of the 5-min mean RR values across 24 hours (SDANN). We also retained the minimum, maximum, and mean pulse rate (24 h). Beat-to-beat IBI metrics (e.g., RMSSD) were not used because raw IBIs were unavailable on *All of Us*. SDANN is a recognized long-term time-domain index and correlates with other PRV measures^{33,37}. Because this measure was derived from photoplethysmography-based pulse-rate measurements rather than electrocardiographic interbeat intervals, it is described as PRV³⁷.

For experiments, the per-period tensor was [N*trimesters (T), 14 days, No. of feature].

Representation learning and clustering

We used the AttentiveGRUAE architecture²⁷ (**Figure 1**), an attention-based GRU autoencoder trained jointly for sequence reconstruction and end-of-window outcome prediction²⁷. The encoder produces a sequence of hidden states over which a multi-head attention block is applied. The attention-refined states are summarized into a fixed-length latent vector (H_{disc}). A GRU decoder reconstructs the input sequence from H_{disc} (reconstruction head), and a feed-forward layer maps it to a binary outcome at the window boundary (prediction head). H_{disc} is an outcome-guided representation²⁷. The reconstruction objective encourages H_{disc} to retain the broader temporal and physiological structure of the sequence, whereas the prediction objective makes the representation sensitive to variation associated with HDP at the window boundary. Joint training was used to balance these goals. The GMM does not receive HDP labels directly; it is fitted to a latent space that has been shaped by outcome supervision. Accordingly, the resulting clusters are interpreted as outcome-guided physiological phenotypes.

For training, the discovery cohort was partitioned at the episode level into a training set (70%) used to fit AttentiveGRUAE, validation set for validation (15%) and a hold-out test set (15%) for reporting. All preprocessing (scalers, imputation rules) was derived only from the training split and applied unchanged to the internal validation set. We used the Adam optimizer with gradient clipping, dropout/weight regularization, early stopping on validation AUC, and ReduceLROnPlateau (max 50 epochs, batch size 64) to avoid overfitting. Training minimized a weighted sum of reconstruction (MSE) and prediction (BCE) losses (Equations 2,3).

$$\text{Loss} = \alpha \text{MSE}_{\text{recon}} + \beta \text{BCE}_{\text{pred}} \quad (2)$$

$$\text{Loss} = \alpha \cdot \frac{1}{NTF} \sum_{i=1}^N \sum_{t=1}^T \sum_{f=1}^F (x_{itf} - \hat{x}_{itf})^2 + \beta \cdot \frac{1}{N} \sum_{i=1}^N [-y_i \log(\hat{y}_i) - (1 - y_i) \log(1 - \hat{y}_i)] \quad (3)$$

where x_{itf} and $\hat{x}_{itf}$ were the true and reconstructed values for feature f at day t for episode i (one subject-trimester 14-day window), y_i was the HDP label, $\hat{y}_i$ was the predicted probability, N was the number of episodes (number of subjects $\times$ number of trimesters), T was the number of days in the window (14), and F was the number of features. To balance the two objectives, we tuned the loss weights (α, β) by cross-validation grid search from values ($\alpha \in \{0.3, 0.5, 0.7, 1.0\}$, $\beta \in \{0.5, 0.7, 1.0\}$) on the discovery training set, selecting the pair that maximized validation AUC while maintaining low reconstruction MSE. This preserved the AttentiveGRUAE objective while adapting the task-specific trade-off for our cohort.

To assess robustness to the choice of data partition, we additionally performed participant-level 5-fold cross-validation. All windows from a participant were retained in the same fold, and the model and preprocessing steps were fitted independently within each training fold.

After training, latent embeddings were clustered with Gaussian mixture models (GMM), and the number of clusters was chosen by minimizing the Bayesian information criterion (BIC). Cluster quality was summarized with the silhouette score and Davies-Bouldin index (DBI); t-stochastic neighbor embedding (t-SNE) was used for visualization only.

Cluster transitions, prior hypertension history, and trimester-end HDP risk

We examined whether changes in cluster membership between adjacent trimesters were associated with the risk of HDP at each trimester boundary. For every pregnancy episode n , we obtained (1) trimester-specific cluster labels at T1, T2, T3, T4, and (2) transition indicators for the intervals T1 to T2, T2 to T3, and T3 to T4. With K clusters, each transition is a categorical variable with levels (i to j), and a “no-change” level (i to i). For episode n and boundary b , let $C_{n,b} \in \{1, \dots, K\}$ denote the cluster at the start of the interval and $C_{n,b+1}$ the cluster at the end. For each ordered pair (i to j), we defined a binary transition indicator:

$$T_{n,b}^{(i \rightarrow j)} = \begin{cases} 1, & \text{if } C_{n,b} = i \text{ and } C_{n,b+1} = j \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

We also included a baseline covariate indicating prior hypertensive history, H_n . For each boundary $b \in \{T1, T2, T3, T4\}$, we then fitted a multivariable logistic regression model with binary outcome $Y_{n,b}$ (HDP documented at the end of boundary b) and predictors given by the transition indicators and H_n :

$$\text{logit}(P(Y_{n,b} = 1)) = \beta_{0,b} + \sum_{(i,j)} \beta_b^{(i \rightarrow j)} T_{n,b}^{(i \rightarrow j)} + \gamma_b H_n \quad (5)$$

Here, $\beta_{0,b}$ was the intercept for boundary b , $\beta_b^{(i \rightarrow j)}$ was the log-odds change in HDP risk associated with transition (i to j) relative to the reference transition, and γ_b was the coefficient for prior chronic hypertension. As a model-robust cross-check of statistical significance for the transition coefficients, we performed a generalized least-squares (GLS) analysis^{38,39}.

As a sensitivity analysis, we repeated the transition models after restricting the analysis to transitions for which the assigned cluster had a GMM posterior probability of at least 0.70 at both the beginning and end of the interval.

Cluster subtype

To summarize cluster characteristics, we assigned each subject to its mode cluster membership at each trimester (ties broken by the earliest occurrence; results were unchanged with maximum posterior membership). We then compared demographics (age, gestational age, race, ethnicity, preterm status, chronic conditions) and physiology (PA, sleep, PRV) across subtypes. For categorical variables, group differences were tested with Pearson’s chi-squared test. For continuous variables, two-group comparisons used two-sample t-tests and Cohen’s d with pooled SD as effect size; for clusters > 2 , we used one-way ANOVA and reported eta (η^2) as effect size. All test families within a variable domain were adjusted for false discovery rate using the Benjamini-Hochberg method ($q=0.05$).

Cross-physiology concordance

For inter-cluster agreement analysis, we defined three cross-physiology phenotypes as (a) *Concordant Healthy*: all three modalities indicated favorable physiology (e.g., higher PA, less sedentary time, higher PRV, less fragmented sleep, greater deep-sleep duration); (b) *Concordant Unhealthy*: the opposite pattern (e.g., low activity, more sedentary time, reduced PRV, shorter, fragmented sleep); (c) *Discordant*: any other combination. We evaluated concordance using pairwise contingency tables and three-way agreement, and computed HDP proportions within phenotype groups at any trimester. Group differences

in HDP prevalence were tested using Pearson’s chi-square and Fisher’s exact test when expected counts <5 , with two-sided p-values.

Temporal interpretability

For each 14-day input window, we extracted attention weights from the model’s multi-head attention block and aggregated them to the day level by averaging across heads, then renormalized the resulting vector to a probability. This yields, per sequence, a 14-dimensional attention profile $\alpha=[\alpha_1,\dots,\alpha_{14}]$ over input days (attention weights are produced by a softmax, hence non-negative and summing to 1)²¹. Using the aggregated α , we computed per-day importance by averaging α over sequences within each subgroup (by trimester and HDP status). “Peak attention days” were defined as day(s) with highest averaged weight. For visualization, we overlaid group-mean trajectories with vertical markers at the peak attention day(s) identified from α .

External validation and stability analysis

For external validation, we used the Washington University Prematurity Research Cohort (WUMOD). We processed WUMOD activity sequences to match the discovery input shape and annotated T1–T3 periods from delivery metadata as described for that cohort¹⁹. WUMOD does not include postpartum actigraphy. Trimester-specific HDP labels were constructed from ICD-10-CM O10–O16 codes^{30,31} and, where available, blood-pressure evidence meeting guideline thresholds (SBP ≥ 140 mmHg and/or DBP ≥ 90 mmHg)¹, applying the same carry-forward rule and 14-day labeling used in discovery.

The raw actigraphy data comes from MotionWatch-8 (MW8). We developed a preprocessing pipeline for this data and extracted features to align it with the discovery data. We identified non-wear time using prior works^{40,41}. If light was missing, we marked minute-level non-wear when both activity ≈ 0 and light ≈ 0 for ≥ 90 minutes. If light was available, we used activity ≈ 0 and low rolling variability (15-min SD below a small threshold) for ≥ 90 minutes to mark non-wear time. Tri-axial MW8 counts were aggregated into 1-minute epochs and categorized into activity intensities using device-specific cut-points calibrated for MW8⁴⁰. Specifically, sedentary behavior was defined as counts per minute (CPM) ≤ 178.5 , light physical activity (LIPA) as $178.5 < \text{CPM} < 562.5$, and moderate-to-vigorous physical activity (MVPA) as $\text{CPM} \geq 562.5$. For each day, we derived accelerometry-based summaries consistent with NHANES-derived metrics⁴¹: total activity as the sum of $\log(1 + \text{CPM})$ ($TLAC$), minutes above MVPA threshold ($MVPA_{minutes}$), minutes in the light physical activity range ($LIPA_{minutes}$), and sedentary minutes ($ST_{minutes}$). We also computed transition-based indicators of behavioral fragmentation, including the sedentary-to-active transition probability, measured as the number of sedentary-to-active transitions per sedentary minute ($SATP$), and the active-to-sedentary transition probability, measured as transitions per active minute ($ASTP$). Valid wear-time was calculated as the sum of all epochs not flagged as non-wear ($WT_{minutes}$). To guard against partition drift and ensure internal consistency, we enforced the constraint:

$$ST_{minutes} + LIPA_{minutes} + MVPA_{minutes} = WT_{minutes} \quad (6)$$

Since the external validation cohort, WUMOD, provided actigraphy data only, downstream analyses used activity-derived features exclusively.

We froze the discovery encoder and embedded WUMOD windows into the same latent space, then assigned clusters via GMM ($K=3$). Let H_{val} denote the set of latent embeddings for the external-validation episodes, and let $z \in H_{\text{val}}$ be the latent vector produced by the encoder for a given episode. For each z , we used the discovery-fit GMM to compute posterior cluster-membership probabilities $\gamma(z) = [\gamma_1(z), \dots, \gamma_K(z)]$, where $\gamma_K(z)$ is the probability of belonging to cluster K . We then assigned a hard cluster label via:

$$c(z) = \operatorname{argmax}_K (\gamma_K(z)) \quad (7)$$

while retaining $\gamma(z)$ for soft cluster-probability analyses. For the external cohort, we quantified cluster separation in the latent space using the silhouette score and DBI, two established internal validity metrics. For visualization only, we projected the latent on (t-SNE) space.

We performed a separate analysis to evaluate the stability of cluster structure within the external latent embeddings. The discovery-trained encoder remained frozen and was used to generate external embeddings once. We then fitted a three-component GMM to the complete external embedding set. For each value $n \in \{1, \dots, 50\}$, we randomly removed n participants, refitted GMM using the reduced external embedding set, and compared the labels of the retained participants with their labels from the complete-external-cohort GMM using the ARI. ARI is invariant to permutations of cluster labels, no manual label matching was required. The procedure was repeated 200 times for each value of n .

The *All of Us* Institutional Review Board Operations Office staff have determined that activities utilizing the “Controlled Tier” (as we would in this work) do not constitute research involving human subjects⁴². Therefore, additional IRB approval and participant consent were not required for this study. For the external validation dataset, the original Washington University Prematurity Research Cohort study received approval from the Washington University in St. Louis Institutional Review Board, and participants provided written informed consent¹⁹. Data for the present external validation analysis were obtained under a data-use agreement with Washington University in St. Louis.

Data availability

This study used training/discovery data from the *All of Us* Research Program’s Controlled Tier Dataset [Version 7], available to authorized users on the Researcher Workbench. For external validation, we used data from the March of Dimes Prematurity Research Center at Washington University, available after signing the Data Use Agreement.

Code availability

The code for the model is shared via GitHub: <https://github.com/tirilab/AttentiveGRUAE--An-Attention-Based-GRU-Autoencoder-for-Temporal-Clustering>⁴³.

Acknowledgments

We also acknowledge editing support by Anne N. Connor, a Senior Science Editor on staff at Johns Hopkins University, funded by the Office of the Vice Provost for Research at Johns Hopkins University.

Funding

This research was funded by a Johns Hopkins University Pivot Grant. No specific grant number is available.

Competing interests

All authors declare no financial or non-financial competing interests.

Author contributions

N.S. conceptualized the study, curated the data, performed formal analysis, conducted the investigation, developed the methodology, validated the study, created the visualizations, and drafted the manuscript. A.J. contributed to study validation, manuscript review and editing. V.M.P. contributed to manuscript review and editing. C.O.T. conceptualized the study, curated the data, administered the project, provided resources, supervised the work, and reviewed and edited the manuscript. All authors read and approved the final manuscript.

References

1. Mammaro, A. et al. Hypertensive disorders of pregnancy. *J. Prenat. Med.* **3**, 1–5 (2009).
2. Radparvar, A. A. et al. Hypertensive disorders of pregnancy. *JACC Adv.* **3**, 100864 (2024). [doi:10.1016/j.jacadv.2024.100864](https://doi.org/10.1016/j.jacadv.2024.100864).
3. Centers for Disease Control and Prevention. Preventing pregnancy-related deaths. [CDC webpage](https://www.cdc.gov/pregnancy/preventing-pregnancy-related-deaths/) (2024; accessed 16 April 2026).
4. Maugeri, A., Barchitta, M. & Agodi, A. How wearable sensors can support the research on foetal and pregnancy outcomes: a scoping review. *J. Pers. Med.* **13**, 218 (2023). [doi:10.3390/jpm13020218](https://doi.org/10.3390/jpm13020218).
5. Alim, A. & Imtiaz, M. H. Wearable sensors for the monitoring of maternal health—a systematic review. *Sensors* **23**, 2411 (2023). [doi:10.3390/s23052411](https://doi.org/10.3390/s23052411).
6. Aune, D., Saugstad, O. D., Henriksen, T. & Tonstad, S. Physical activity and the risk of preeclampsia: a systematic review and meta-analysis. *Epidemiology* **25**, 331–343 (2014). [doi:10.1097/EDE.0000000000000036](https://doi.org/10.1097/EDE.0000000000000036).
7. Abbasi, M. et al. Association between sleep disorders and preeclampsia: a systematic review and meta-analysis. *J. Matern. Fetal Neonatal Med.* **37**, 2419383 (2024). [doi:10.1080/14767058.2024.2419383](https://doi.org/10.1080/14767058.2024.2419383).
8. Rowan, S. P., Lilly, C. L., Claydon, E. A., Wallace, J. & Merryman, K. Monitoring one heart to help two: heart rate variability and resting heart rate using wearable technology in active women across the perinatal period. *BMC Pregnancy Childbirth* **22**, 887 (2022). [doi:10.1186/s12884-022-05183-z](https://doi.org/10.1186/s12884-022-05183-z).
9. Moors, S. et al. Heart rate variability in hypertensive pregnancy disorders: a systematic review. *Pregnancy Hypertens.* **20**, 56–68 (2020). [doi:10.1016/j.preghy.2020.03.003](https://doi.org/10.1016/j.preghy.2020.03.003).

10. Spracklen, C. N., Ryckman, K. K., Triche, E. W. & Saftlas, A. F. Physical activity during pregnancy and subsequent risk of preeclampsia and gestational hypertension: a case control study. *Matern. Child Health J.* **20**, 1193–1202 (2016). doi:10.1007/s10995-016-1919-y.
11. Zhu, Z. et al. Effects of physical exercise on blood pressure during pregnancy. *BMC Public Health* **22**, 1733 (2022). doi:10.1186/s12889-022-14074-z.
12. Guo, Y., Xu, Q., Dutt, N., Kehoe, P. & Qu, A. Longitudinal changes in objective sleep parameters during pregnancy. *Womens Health (Lond.)* **19**, 17455057231190952 (2023). doi:10.1177/17455057231190952.
13. Lu, Q. et al. Sleep disturbances during pregnancy and adverse maternal and fetal outcomes: a systematic review and meta-analysis. *Sleep Med. Rev.* **58**, 101436 (2021). doi:10.1016/j.smrv.2021.101436.
14. Musa, S. M., Adam, I. & Lutfi, M. F. Heart rate variability and autonomic modulations in preeclampsia. *PLoS ONE* **11**, e0152704 (2016). doi:10.1371/journal.pone.0152704.
15. Gangakhedkar, G. R. & Kulkarni, A. P. Physiological changes in pregnancy. *Indian J. Crit. Care Med.* **25**, S189–S192 (2021). doi:10.5005/jp-journals-10071-24039.
16. Montgomery-Csobán, T. et al. Machine learning-enabled maternal risk assessment for women with pre-eclampsia (the PIERS-ML model): a modelling study. *Lancet Digit. Health* **6**, e238–e250 (2024). doi:10.1016/S2589-7500(23)00267-4.
17. Alkhodari, M. et al. The role of artificial intelligence in hypertensive disorders of pregnancy: towards personalized healthcare. *Expert Rev. Cardiovasc. Ther.* **21**, 531–543 (2023). doi:10.1080/14779072.2023.2223978.
18. Boyer, T. M. et al. Hypertensive disorders of pregnancy and premature cardiovascular disease in a diverse cohort of young US women. *Circulation* **153**, 480–492 (2026). doi:10.1161/CIRCULATIONAHA.125.078057.
19. Stout, M. J. et al. A multidisciplinary Prematurity Research Cohort Study. *PLoS ONE* **17**, e0272155 (2022). doi:10.1371/journal.pone.0272155.
20. Ravindra, N. G. et al. Deep representation learning identifies associations between physical activity and sleep patterns during pregnancy and prematurity. *npj Digit. Med.* **6**, 171 (2023). doi:10.1038/s41746-023-00911-x.
21. Vaswani, A. et al. Attention is all you need. In *Advances in Neural Information Processing Systems* **30** (Curran Associates, 2017). Conference proceedings.
22. Jain, S. & Wallace, B. C. Attention is not explanation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* 3543–3556 (Association for Computational Linguistics, 2019). doi:10.18653/v1/N19-1357.
23. Wang, D., Ma, X., Schulz, P. E., Jiang, X. & Kim, Y. Clinical outcome-guided deep temporal clustering for disease progression subtyping. *J. Biomed. Inform.* **158**, 104732 (2024). doi:10.1016/j.jbi.2024.104732.
24. Ben-Hur, A., Elisseeff, A. & Guyon, I. A stability based method for discovering structure in clustered data. In *Biocomputing 2002* 6–17 (World Scientific, 2001). doi:10.1142/9789812799623_0002.
25. Steinley, D. Properties of the Hubert-Arabie adjusted Rand index. *Psychol. Methods* **9**, 386–396 (2004). doi:10.1037/1082-989X.9.3.386.
26. The All of Us Research Program Investigators. The “All of Us” Research Program. *N. Engl. J. Med.* **381**, 668–676 (2019). doi:10.1056/NEJMSr1809937.

27. Soley, N., Patel, V. M. & Taylor, C. O. AttentiveGRUAE: an attention-based GRU autoencoder for temporal clustering and behavioral characterization of depression from wearable data. Preprint at doi:10.48550/arXiv.2510.02558 (2025).
28. Quinn, J. A. et al. Preterm birth: case definition & guidelines for data collection, analysis, and presentation of immunisation safety data. *Vaccine* **34**, 6047–6056 (2016). doi:10.1016/j.vaccine.2016.03.045.
29. Liang, O. S., Sheffield, J. S. & Taylor, C. O. Detecting patterns of prescription drug use during pregnancy and lactation with visualization techniques. *AMIA Jt. Summits Transl. Sci. Proc.* **2019**, 478–487 (2019).
30. Okun, M. L., Kohl, V. & Feliciano, L. Comparison of longitudinal diary and actigraphy-assessed sleep in pregnant women. *Sleep Med.* **88**, 149–156 (2021). doi:10.1016/j.sleep.2021.09.015.
31. ICD10Data.com. 2026 ICD-10-CM codes O10–O16: edema, proteinuria and hypertensive disorders in pregnancy, childbirth and the puerperium. *ICD-10-CM code listing* (accessed 30 October 2025).
32. Hauspurg, A. & Jeyabalan, A. Postpartum preeclampsia or eclampsia: defining its place and management among the hypertensive disorders of pregnancy. *Am. J. Obstet. Gynecol.* **226**, S1211–S1221 (2022). doi:10.1016/j.ajog.2020.10.027.
33. Mejía-Mejía, E., May, J. M., Torres, R. & Kyriacou, P. A. Pulse rate variability in cardiovascular health: a review on its applications and relationship with heart rate variability. *Physiol. Meas.* **41**, 07TR01 (2020). doi:10.1088/1361-6579/ab998c.
34. Jones, S. E. et al. Who is pregnant? Defining real-world data-based pregnancy episodes in the National COVID Cohort Collaborative (N3C). *JAMIA Open* **6**, ooad067 (2023). doi:10.1093/jamiaopen/ooad067.
35. Smith, L. H., Wang, W. & Keefe-Oates, B. Pregnancy episodes in All of Us: harnessing multi-source data for pregnancy-related research. *J. Am. Med. Inform. Assoc.* **31**, 2789–2799 (2024). doi:10.1093/jamia/ocae195.
36. ICD9Data.com. 2012 ICD-9-CM Volume 1 diagnosis codes. *ICD-9-CM code listing* (accessed 30 October 2025).
37. Friligkou, E. et al. Integrating genome-wide information and wearable device data to explore the link of anxiety and antidepressants with pulse rate variability. *Mol. Psychiatry* **30**, 2309–2315 (2025). doi:10.1038/s41380-024-02836-7.
38. Taboga, M. Generalized least squares. In *Lectures on Probability Theory and Mathematical Statistics*, online appendix (Kindle Direct Publishing, 2021). *StatLect* (accessed 30 October 2025).
39. Susko, E. Confidence regions and hypothesis tests for topologies using generalized least squares. *Mol. Biol. Evol.* **20**, 862–868 (2003). doi:10.1093/molbev/msg093.
40. Landry, G. J., Falck, R. S., Beets, M. W. & Liu-Ambrose, T. Measuring physical activity in older adults: calibrating cut-points for the MotionWatch 8©. *Front. Aging Neurosci.* **7**, 165 (2015). doi:10.3389/fnagi.2015.00165.
41. Tabacu, L., Ledbetter, M., Leroux, A., Crainiceanu, C. & Smirnova, E. Quantifying the varying predictive value of physical activity measures obtained from wearable accelerometers on all-cause mortality over short to medium time horizons in NHANES 2003–2006. *Sensors* **21**, 4 (2021). doi:10.3390/s21010004.

42. All of Us Institutional Review Board Operations Office. Not Human Subjects Research Determination: Research Hub: Controlled Tier Data Access Process for the All of Us Research Program. Determination letter (2021; revised 31 August 2021; accessed 10 December 2025).
43. tirilab. AttentiveGRUAE--An-Attention-Based-GRU-Autoencoder-for-Temporal-Clustering. GitHub repository (accessed 10 December 2025).

Figure Legends

Figure 1. The AttentiveGRUAE pipeline learns wearable-derived digital phenotypes and evaluates their characteristics and reproducibility. Physical activity, sleep, and pulse rate variability measurements from 14-day windows are processed separately using an attention-based gated recurrent unit autoencoder (AttentiveGRUAE). Training combines mean squared error (MSE) reconstruction loss and binary cross-entropy (BCE) classification loss, weighted by alpha and beta, respectively. Gaussian mixture models (GMMs) cluster the learned representations. Downstream analyses characterize cluster subtypes, evaluate cluster structure using the silhouette score and Davies–Bouldin index, visualize embeddings using t-distributed stochastic neighbor embedding (t-SNE). The lower pathway applies the trained model to external wearable data from the Washington University Prematurity Research Cohort (WUMOD) and evaluates cluster stability using the adjusted Rand index (ARI). Yellow-to-red heatmaps display wearable input sequences derived from study data; blue boxes identify processing and analysis stages, and connecting arrows indicate the flow of data and representations. Dashed outlines group the model architecture and external representation-extraction stages. E and D denote the encoder and decoder; H denotes latent representations, with subscripts indicating classification and external-validation representations. N denotes the number of participants, w the number of observation windows, t the number of days per window, and d the number of features. T1–T3 denote the pregnancy trimesters; T4 denotes postpartum.

Figure assembled by the authors using Microsoft Word and PowerPoint; pregnant-person icons sourced from the built-in Microsoft Office icon library. Heatmaps generated by the authors from study data.

Figure 2. Study design, cohort selection, and data windowing. a) Timeline illustrating the first, second, and third trimesters and the postpartum period, together with the 14-day wearable observation windows preceding the specified assessment boundaries. The orange horizontal line represents the study timeline, and blue arrows mark the illustrated boundaries. Dashed vertical lines show trimester ends across pregnancy and postpartum periods. b) Cohort-selection flowchart showing the inclusion and exclusion steps used to identify pregnancies with linked wearable measurements and electronic health records. HDP, hypertensive disorders of pregnancy; T1–T3, first through third trimesters; D is day; DD, delivery date.

Figure 3. Learned wearable representations distinguish clusters with different physiological profiles across pregnancy and postpartum. a) Physical activity embeddings and cluster-specific distributions of daily step counts and sedentary minutes. b) Sleep embeddings and distributions of minutes asleep and rapid eye movement sleep. c) Pulse rate variability embeddings and distributions of heart rate and SDANN. In each panel, the left plot displays a two-dimensional t-distributed stochastic neighbor embedding (t-SNE) projection. These projections are shown only for visualization and were not used to select the number of clusters or assess cluster validity. Teal, pink, and purple identify physical activity clusters C0, C1, and C2, respectively. Teal and pink identify clusters C0 and C1 for sleep and pulse rate variability. Cluster identifiers are specific to each modality. The middle and right plots show feature distributions across the first, second, and third trimesters and postpartum. One, two, and three asterisks (*) indicate p values below 0.05, 0.01, and 0.001, respectively; “ns” indicates a nonsignificant comparison. PRV, pulse rate variability; REM, rapid eye movement; SDANN, standard deviation of average normal-to-normal intervals.

Figure 4. Cluster transitions are associated with HDP across successive pregnancy and postpartum boundaries. a) Physical activity cluster transitions, with C0 representing higher activity, C1 moderate activity, and C2 lower activity. b) Sleep cluster transitions, with C0 representing restorative sleep and C1 fragmented sleep. c) Pulse rate variability cluster transitions, with C0 representing lower variability and C1 higher variability. Within each panel, plots show the T1-to-T2, T2-to-T3, and T3-to-postpartum comparisons from left to right. Horizontal bars represent fitted regression coefficients for cluster transitions, hypertension-history variables, and the intercept. Labels such as "C0 to C1" indicate a transition from the first cluster to the second; repeated cluster identifiers ("C0 to C0") indicate stable membership. Blue bars indicate $p < 0.05$; red bars indicate non-significant transitions. HDP, hypertensive disorders of pregnancy; HTN, hypertension; PRV, pulse rate variability; T1–T3, first through third trimesters.

Figure 5. Average temporal attention profiles differ across wearable modalities and pregnancy periods. a) Physical activity. b) Sleep. c) Pulse rate variability. Each plot displays average model attention weights across the 14 days of the wearable input window. Blue, orange, green, and red lines represent the first trimester, second trimester, third trimester, and postpartum period, respectively. Circular markers indicate individual input days, and connecting lines show the temporal profiles. Physical activity attention increases toward later input days, sleep attention varies across days and periods, and pulse rate variability attention is approximately uniform. Attention weights describe the model's allocation of attention and should not be interpreted as causal effects or evidence of clinical deterioration. PRV, pulse rate variability.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [reviseddiscoveringwerablesphenotypesHDPsupplementarynpjai.pdf](#)