Journal of Informatics and Web Engineering

Vol. 5 No. 2 (June 2026) eISSN: 2821-370X

An AI-Based Framework for Heart–Brain–Body Coherence in Wellness Monitoring: A Longitudinal Simulation Study

Asiah Lokman[1*], Akalpita Tendulkar[2], Yujiao Zhang[3], Mantena Sireesha[4]**

[1,2]Malaysia University of Science and Technology, Malaysia University of Science & Technology,Block B, EncorpStrand Garden Office, No. 12, Jalan PJU 5/5, Kota Damansara, 47810 Petaling Jaya, Selangor, Malaysia.

[3]Department: School of Information Engineering, Shaanxi Xueqian Normal University, Shenheer Rd, Chang'An, Xi'An, 710100 Shaanxi, China.

[4]Sasi Institute of Technology & Engineering, Tadepalligudem, West Godavari District, Andhra Pradesh, India.

**corresponding author: (asiah@must.edu.my; ORCiD: 0009-0000-4970-4066)*

***corresponding author: (sireeshamantena235@gmail.com; ORCiD: 0009-0008-7997-4946)*

Abstract - Present wellness monitoring systems primarily emphasize single-modality metrics, neglecting the intricate, interdependent interactions among cardiac, neural, and behavioral regulatory systems. This paper introduces a machine learning framework that implements heart–brain–body coherence as a dynamic, longitudinally evaluated wellness metric. In contrast to previous simulation-based studies that utilized independently generated, cross-sectional data categorized by deterministic IF-THEN rules, the current study adopts a longitudinal simulation design featuring 80 virtual subjects monitored across 30 consecutive daily profiles (N=2,400 records). This design incorporates physiologically realistic inter-variable correlations and probabilistic coherence labelling derived from a continuous risk-scoring function. The subject-level train/test split (56 training subjects, 1,680 records; 24 test subjects, 720 records) stops data from leaking over time. The precision for logistic regression was 79.2% (ROC-AUC = 0.887; five-fold CV AUC=0.904±0.010), while random forest's precision was 78.1% (ROC-AUC = 0.871; five-fold CV AUC=0.894±0.010). The values were significantly varied across all inputs, with an HRV RMSSD of 22.6%, self-reported fatigue with 21.2%, hours of sleep with 19.9%, and stress score at 18.7%. This shows real multi-signal coherence instead of artefacts that are caused by variables that create labels. These results establish a reliable simulation benchmark for subsequent validation with authentic wearable datasets, including MIMIC-IV and PhysioNet.

Keywords— AI in Healthcare, Wellness Monitoring, Heart–Brain Communication, Physiological Coherence, Longitudinal Simulation, Wearable Data, Machine Learning

Received: 16 February 2026; Accepted: 23 May 2026; Published: 16 June 2026

Journal of Informatics and Web Engineering
https://doi.org/10.33093/jiwe.2026.5.2.23

Published by MMU Press. URL: https://journals.mmupress.com/jiwe

1. INTRODUCTION

The field of digital health monitoring has grown quickly, thanks in part to the popularity of consumer wearables that constantly track things like heart rate, sleep duration, and physical activity [1], [2]. However, contemporary wellness paradigms predominantly exhibit reductionism, as most systems evaluate individual metrics in isolation rather than as integral components of an interconnected regulatory framework [3], [4], [5]. This approach may not take into account the changing relationships between the heart, brain, and behavior systems that work together to create true wellness.

This paper fills that gap by presenting a machine learning framework for evaluating heart–brain–body coherence through longitudinal, multimodal simulated data based on physiologically plausible parameter distributions and inter-variable correlations. The framework utilizes a longitudinal simulation design featuring 80 virtual subjects monitored across 30 consecutive daily profiles, facilitating correlated multivariate feature generation. That shows how psychophysiological dependencies work and how probabilistic risk-based coherence labelling works, which doesn't create the trivially separable class structures that deterministic threshold rules do. The combination of these design choices creates a simulation baseline that more accurately reflects how real-world coherence monitoring works overtime and with multiple signals.

1.1. Limitations of Current Approaches

Modern wellness systems assess individual metrics such as Resting Heart Rate (RHR), step count, and sleep duration, without accounting for their interrelations. Threshold-based alerts activate based on absolute values without considering contextual or temporal modifications, overlooking informative metrics like signal variability, complexity, and cross-signal relationships that indicate regulatory function [6]. Studies consistently demonstrate that indicators of resilience frequently reside in the structure and synchrony of physiological signals, rather than in absolute values alone.

1.2. Emerging Evidence of Heart–Brain Communication

The heart is now seen as an active part of two-way communication with the brain, not just a passive receiver of commands from the brain [7], [8]. The intrinsic cardiac nervous system comprises approximately 40,000 neurons capable of autonomous sensing and signaling [9]. Vagal afferent fibers convey cardiac information to the brain at a significantly higher ratio than efferent traffic, affecting emotion regulation, attention, and cognition [10]. HRV, quantified as RMSSD from inter-beat intervals, is the most readily available non-invasive indicator of autonomic function, with diminished HRV consistently linked to chronic stress, inflammation, and cardiovascular risk [7], [11].

1.3. Motivation for Coherence-Based Wellness Monitoring

Coherence is a dynamic state where the rhythmic functioning of cardiovascular, respiratory, autonomic, and cognitive systems is temporally integrated, behave efficiently relative to the energy they require, and possess stable signals [12] [13]. An incoherent state is characterized by a lack of regular patterning and low levels of signal coupling, which often precede clinical manifestations of distress and can be early indices of dysregulation that cannot be detected by conventional methods based on thresholds [8]. The application of machine learning allows for the effective capture of the many different non-linear, multi-signal relationships present in the physiological and behavioral systems and the ability to adapt to individual baselines over time [14], [15], [16].

1.4. Contributions of This Study

Specific contributions of the proposed revision include:

- An innovative longitudinal simulation framework that includes 80 subjects with 30 days of data, enabling the generation of physiologically correlated multivariate daily profiles for each subject.
- Probabilistic coherence labelling based on continuous risk scores using a weighted composite of z-scored physiological and behavioral variables, eliminating arbitrary/deterministic circular labelling methods.
- Subject-level test/training separation to eliminate temporal data leakage.
- Five-fold cross-validation of model performance, with confidence intervals.
- Realistic and distributed features reflecting multi-signal coupling.
- Clear benchmarking limitations, with recommendations for future validations against MIMIC-IV and PhysioNet datasets.

2. SCIENTIFIC AND CONCEPTUAL BACKGROUND

2.1. Heart–Brain Communication and Autonomic Regulation

The Central Nervous System (CNS) has traditionally been viewed as the major regulator of cardiovascular function. The heart-brain interaction dynamic and reciprocal, as has been shown by an increasing number of studies demonstrating dynamic and reciprocal heart−brain interactions [7], [8], [17]; the vagus nerve supports this dynamic, with the majority of afferent information from the heart going through the vagus to the brainstem, limbic areas, and the cortex in a much larger ratio of incoming signals to outgoing signals from the vagus [10] An example of this dynamic is the large number of urethra endoneuronal vagal afferent fibers that connect the urethra to the brain.

The heart also communicates with the brain indirectly via various hypothalamic and limbic regulators such as neurohumoral peptides (e.g., Atrial Natriuretic Peptide - ANP and B-type Natriuretic Peptide - BNP) and through activation of baroreceptors, which also influence CNS function [8]. The heart's electromagnetic field also plays a role in this communication process, but this area has yet to be thoroughly studied [10]. As such, HRV, is a crucially informative and easily accessible measure.

RMSSD, for example, correlates with parasympathetic vagal gut tone; whereas, the ratio of LF/HF in heart rate spectral analysis provides an additional means of determining sympathetic vs. parasympathetic balance [7]. Low values of HRV are consistently related to cardiovascular disease, mood disorders, and critical illness [11]. In addition, recent research in psychophysiology has placed an increasing emphasis on HRV rates being indicators of systemic health, including coherence of multiple systems through measures of synchronization, stability, and phase alignment across various systems [10], [18].

2.2. Physiological and Cognitive Coherence

Coherence in biological systems can be defined as a dynamic state where the rhythms of various physiological systems (including systems governed by the Autonomic Nervous System (ANS) are in harmony with each other as a stable, efficient, and synchronized state with harmonious oscillations that have stable phases and provide rhythmic order across the heart (HR), lung (respi), and ANS signals. Psychophysiological coherence specifically refers to the heart rhythm pattern resembling a sine wave oscillating near 0.1 Hz, indicating that cardiovascular and respiratory systems are coordinated via parasympathetic innervation [18], [12]. People in a Coherent state have consistently demonstrated that they have better emotion regulation, cognitive clarity, and overall physiological efficiency [19].

The emotional experience of an individual is a central contributor to Coherence dynamics. Positive emotions are correlated with higher HRV and more coherent ANS signals while anxiety and mental overload produce irregular HRV, reduced vagal tone, and loss of entrainment between the HR and respiration [18]. Importantly, Coherence appears to be a trainable characteristic through the use of slow breathing, mindfulness, and Biofeedback techniques [13], [12] make it a modifiable target for achieving wellness.

2.3. Rationale for Realistic Simulation Parameters

The simulation parameters used in this research have a foundation in published references for physiological values. The HRV (RMSSD) numbers generated from a normal distribution (M=50, SD=14 ms) are consistent with the population norms of adults 25-70 years old based on PhysioNet population norms [10], [7]. The RHR generated from MIMIC-IV clinical reference ranges are also derived from a reference mean of 68 bpm and SD of 10 bpm (see ppm). The estimated mean sleep time generated (6.8 hrs.) and the SD (0.8 hrs.) also agree with the epidemiologic data for adults from Southeast Asia in [20]. For stress and fatigue measured on a 1-10-point scale, the distributions reflect the assessment in studies using ecological momentary assessment, which have been validated in previous studies [18].

In sum, these variables are created in such a way to produce physiologically valid cross-correlations: high stress levels reduce the amount of HRV generated; increase RHRs; shorten sleep duration; and add to fatigue the next day through a cumulative carryover effect of not getting enough sleep. This is different than generating random/independent (univariate) combinations of features, resulting in feature combinations that may not exist in the real-world physiology (e.g., for example having both extremely high HRV and extremely high chronic stress at the same time would not be a physiologic possibility).

2.4. Benchmark Datasets for Future Validation

To validate the Coherence Framework against actual physiological data is a critical next step. There are two datasets that are available publicly which would be ideal for achieving this:

MIMIC-IV (Medical Information Mart for Intensive Care IV): This dataset, hosted by PhysioNet, is comprised of de-identified Clinical records and high resolution waveform data from approximately 40,000 ICU stays at Beth Israel Deaconess Medical Center. It contains continuous HRV-relevant data, continuous vital sign measurements, clinical annotations, and other features that allow for labelling coherence through supervised processes using actual ground truth [14].

PhysioNet/WESAD (Wearable Emotion Detection and Stress Detection): This multimodal dataset is composed of data collected from chest and wrist worn sensors (ECG, EDA, EMG, respiration and temperature) from 15 subjects exposed to laboratory manipulated stress and neutral conditions. It has been validated with regard to consistent levels of stress making it ideal for comparing to has been simulated to be coherent [14].

The intent of further future work is to benchmark the current Coherence Framework against these datasets. The intent in conducting this current simulation study is to demonstrate its feasibility, and does not propose that the simulated subjects are coherent.

3. AI-BASED COHERENCE MODELLING FRAMEWORK

3.1. System Overview

The four-module structure of AI-based coherence-based wellness monitoring is shown in Figure 1. The modules are as follows:

1) Data Acquisition - gathers continuous physiological signals along with user-reported data through Ecological Momentary Assessments (EMA),
2) Feature Processing - includes pre-processing, normalizing and extracting temporal features from the signal data within a 30-day sliding window,
3) Coherence Estimation - consists of machine learning models that provide longitudinal risk scores as well as probability classifications, and
4) Risk Communication - provides to the user or clinician early warning alerts that are interpretable, and trend visualizations along with personalized wellness insights.

A longitudinal feedback loop will connect the risk communication output back into the Feature Processing Module allowing for the update of individual baseline values and the refinement of estimates of Coherence over successive observation periods.

The system integrates continuous streams of wearable devices (HRV, heart rate, activity) along with twice-daily EMA inputs (stress, fatigue). Feature extraction generates a 30-day sliding window of daily profiles for each individual; the risk associated with coherence is represented by a probabilistic score that provides for both classification (dysregulated or regulated) and continuous risk trajectories for longitudinal trend analysis. Figure 1 illustrates the four processing modules, multimodal data inputs, and longitudinal feedback loop.

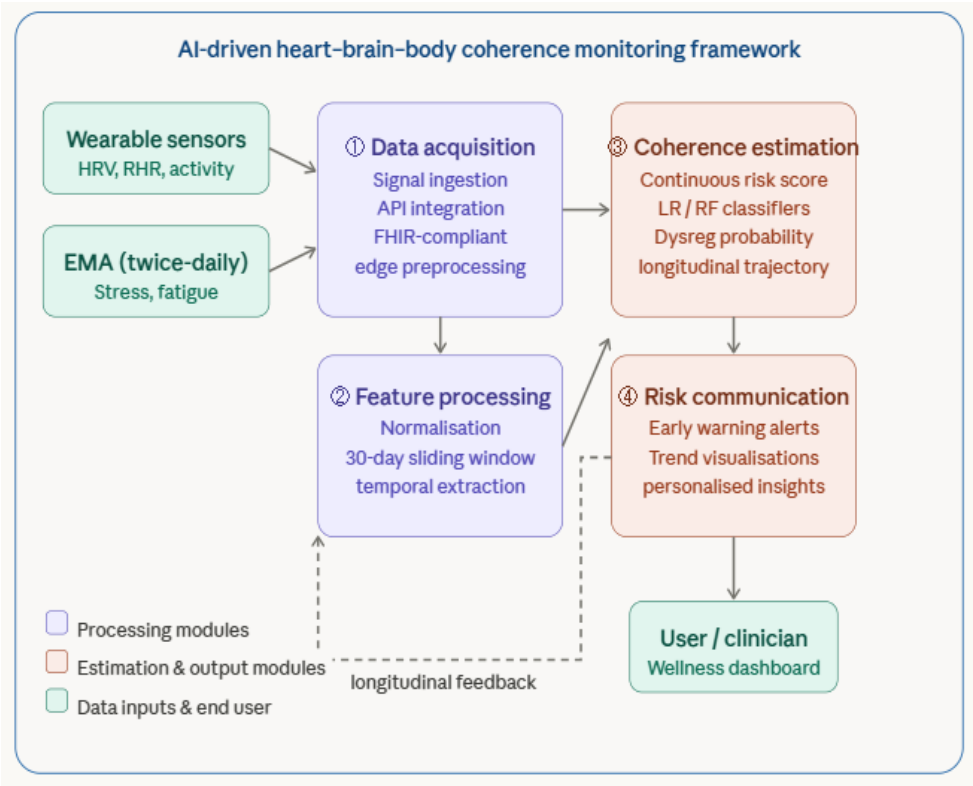

Figure 1. System Architecture of the AI-Based Heart–Brain–Body Coherence Framework

3.2. Multimodal Input Signals

Seven input features are organized across three domains in the following subsections.

3.2.1. Physiological Signals

HRV RMSSD serves as the primary marker of autonomic activity and provides time-domain measures of parasympathetic activity using inter-beat intervals derived from wrist-based PPG devices. RMSSD is highly reliable with respect to artefact resistance and is also very sensitive to assessing vagal tone. The normative range for RMSSD in normal healthy adults is 20-100 milliseconds [7], [11].

RHR provides additional information about cardiovascular performance. RHR is assessed during periods of rest. RHR provides information on the degree of vagal functioning and overall cardiovascular fitness; the lower the value of RHR, the higher the level of vagal function and cardiovascular fitness. The normative range for RHR in normal healthy adults is 48-95 bpm [11].

3.2.2. Behavioral Indicators

Sleep Duration is obtained using algorithms included within wearable technology that allow for assessment of sleep stages. If an individual does not receive adequate sleep (less than 7 hours per night), then this can be used as a carryover source of stress to be considered when assessing the next day for levels of stress and fatigue. There is currently documented evidence showing a bi-directional relationship between sleep and the autonomic system [20].

Average Physical Activity (Number of Steps) is a behavioral measure used to estimate energy expenditure and the level of conditioning of the ANS. Chronic inactivity is associated with decreased vagal tone and increased resting sympathetic activity [2].

3.2.3. Cognitive-Affective Indicators

The Perceived Stress Score and Self-Reported Fatigue are measured using a 1–10 Likert scale through the Ecological Momentary Assessment (EMA) methodology twice per day. Both of these self-reported measures are well known to have good psychometric properties and are indicators of an individual's subjective well-being prior to any physiological or objective changes occurring as a result of their degree of coherence dysregulation [18].

3.2.4. Demographic Covariate

A person's age is being used as a covariate due to a known decline in HRV and autonomic flexibility associated with increases in a person's age across the age ranges of 25–70 years old [7].

3.3. Coherence Labelling Methodology

An advance methodologically is the application of a continuous risk-score function for labelling of coherence rather than using deterministic IF-THEN threshold rules as previous authors did.

For any single day of an observation the coherence risk score (R) may be calculated as follows:

$$R = -0.35z(HRV) + 0.28z(stress) + 0.22z(fatigue) - 0.15z(sleep) + \varepsilon, \text{ where } \varepsilon \sim N(0, 0.35).$$

The variable denoted as z (·) is the standardized score relative to a normative population dataset. P(dysregulated) will be derived by using a logistic function as follows:

$$P(dysregulated) = 1 / (1 + \exp(-1.5R)).$$

Any observation whose probability of being dysregulated (Class 1) is greater than 0.50 will be labelled as dysregulated. This probabilistic approach has the following benefits:

(i) introduces classification uncertainty at the boundary of coherence;
(ii) ensures that there is no direct correspondence between the variables or features which generate the label & those used to classify the observation;
(iii) provides an approximately equal distribution of class membership (~52% dysregulated) consistent with estimates of chronic stress & insufficient sleep for working-age adult populations [20].

3.4. System Architecture Considerations

According to Figure 1, the Four-Module Framework is a simulation pipeline that depicts the engineering considerations to keep in mind for deploying a wearable in the real-world setting.

The REST endpoints that use wearable data in the API integration use HL7 FHIR standards. Which are from Garmin, Fitbit, and Apple Health ecosystem.

By implementing contemporary technologies, the inference will undergo processing latency of less than 2 seconds on edge processors which will enable us to offer overnight summaries. Daily batching of overhead coherence score is done.

The self-reported information concerning stress and fatigue will be processed on the edge, thereby ensuring user privacy. On the other hand, the physiological parameters will be compressed and sent to the cloud. This will allow analysis of trends at a population level through federated aggregation.

Wearables employ microcontrollers that are limited in resources. Microcontrollers are capable of executing a logistic regression inference. Random forest Inference has a memory footprint of 50 MB.

4. METHODOLOGY

4.1. Longitudinal Simulation Design

This research employs a longitudinal simulation design and generates daily profiles for 80 virtual subjects across 30 consecutive days (2,400 total observations). These longitudinal simulation methods differ greatly from cross-sectional simulation methods in three important respects:

i. They maintain within-subject temporal autocorrelation due to a mechanism of sleep deficit carryover.
ii. The differences in baseline physiology show meaningful diversity across subjects, thereby reflecting typical human diversity in the population.
iii. Because the training and test data are kept separate at the subject level, random record-level splits in longitudinal data do not create data leakage.

4.2. Dataset Characteristics and Generation

Table 1 summarizes the simulation parameters and their physiological justifications.

Table 1. Simulation Parameter Specifications and Physiological Justifications

Variable	Distribution	Parameters	Physiological Justification
HRV RMSSD (ms)	Normal	μ=50, σ=14; clip [10,110]	PhysioNet adult norms, ages 25–70 [7], [10]
Resting HR (bpm)	Normal	μ=68, σ=10; clip [48,95]	MIMIC-IV reference ranges [11]
Sleep Hours	Normal	μ=6.8, σ=0.8; clip [3,10]	SEA adult epidemiology [20]
Average Steps	Normal	μ=7200, σ=1800; clip [500,18000]	WHO physical activity guidelines [2]
Stress Score (1–10)	Uniform + drift	Base 1.5–4.5; stress-carryover +0.5×sleep deficit	EMA validation studies [18]
Fatigue (1–10)	Correlated composite	Coupled to stress, sleep deficit	Validated fatigue-sleep coupling [18], [20]
Age	Uniform integer	Range [25,70]	Working-age adult target population [7]

4.3. Inter-Variable Correlation Structure

There are many developments made in the simulation model. Physiologically realistic relationships have been built into the model. The following mechanisms have been introduced:

- A sleep deficit (max (0, 7 - previous sleep hours)) is created when a person has not had enough sleep which increases their stress and fatigue for the next day since the sympathetic system will be activated because of lack of sleep.
- Stress lowers HRV RMSSD (coefficient: -1/8 ms per unit of stress) as demonstrated by the data showing autonomic reactivity to stress.
- The resting HR increase due to both stress and low HRV, mimicking the dynamics of sympathovagal balance.
- Fatigue decreases daily step counts, which decreases their physical activity level as stress increases.
- Daily physiological fluctuations are represented by zero-mean Gaussian noise (σ factor of each variable).

4.4. Train/Test Splitting Strategy

We implemented a split at the subject level for our analysis, with 56 subjects (70% of the sample) in a training set consisting of 1,680 observations and 24 subjects (30% of the sample) in a test set consisting of 720 observations to prevent temporal sequences from test subjects being present in the training set (which would create leakage and artificially inflate performance) when using longitudinal designs. There is also class balance between splits (52.7% of the training sample have dysregulated rhythms, and 49.9% of the test sample have dysregulated rhythms).

4.5. Model Specifications

The following two classifiers were evaluated;

- LR Logistic regression classifier (L2 Regularized (C=1.0) with class weight scaled to be balanced across classes), with the maximum number of allowed iterations of 1,000. Input features were standardized (zero mean, unit variance) using statistics calculated from train data only, when calculating class balance (misclassification cost) for class weight to correct for small class imbalances, based upon class frequency.

- RF Random Forest Classifier (300 decision trees; max depth of 8) with class weight scaled to be balanced across classes, with the use of Random state 42. Input features not scaled since Random Forests are not affected by input feature scaling. Class balance weights used in both models were determined using same methodology to provide consistency.

The complete code requisite for implementation, including complete parameter definitions, procedures to generate training data for both classifiers, and procedures for evaluation of performance of both classifiers can be found in Appendix A.

4.6. Evaluation Protocol

To evaluate the model's performance two methods are used. The first method is fivefold stratified cross-validation (CV) on the training data to report the mean and standard deviation of ROC-AUCs for evaluating the stability of the trained model. The other method is to also evaluate the model on a separate test set (720 instances from 24 subjects) that the model has never seen, reporting metrics such as accuracy, precision, recall, F1, and ROC-AUC. For the cross-validation method, StratifiedKFold (shuffle=True, random_state=42) is used for creating folds to ensure the results can be repeated and that class distribution remains consistent across the folds. All metrics are taken from the standard implementations of scikit-learn (v1.3).

5. RESULTS

5.1. Dataset Overview

Using 2,400 observations from 80 subjects over a 30-day period, longitudinal analysis showed a roughly even distribution of regulated (1,155; 48.1%) and dysregulated (1,245; 51.9%) states (Figure 2). This distribution reflects the consistently dysregulated distribution of stress burden estimates for working-age adults in which 35% - 55% of

working adults meet criteria for chronic stress-related disorders [20]. The near-equal distribution of regulated and dysregulated observations mitigates bias in classifier training resulting from class imbalance problems.

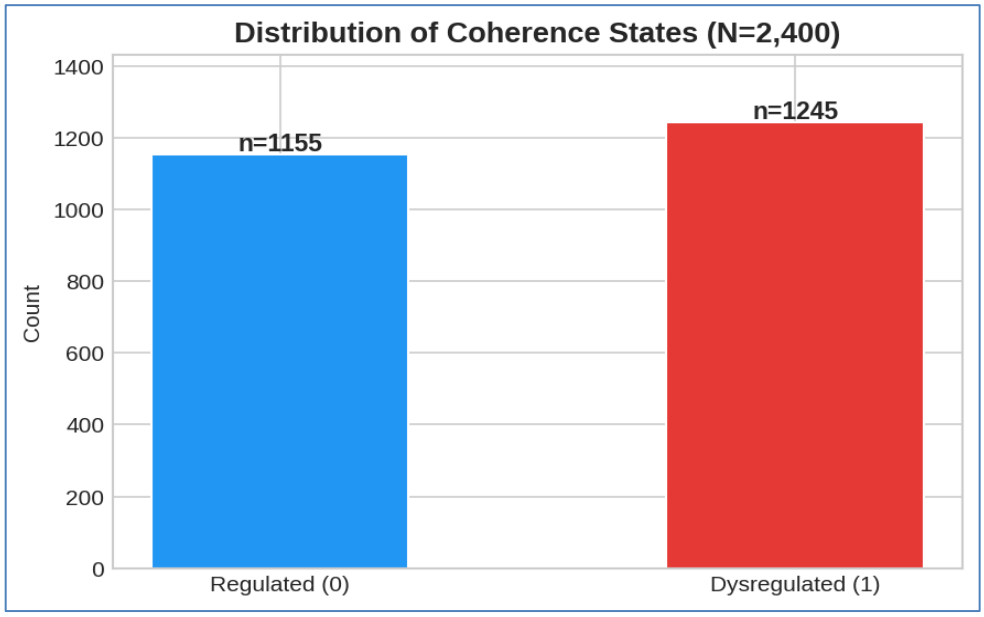

Figure 2. Distribution of Coherence States in the Longitudinal Simulation Dataset (N=2,400 across 80 subjects × 30 days)

5.2. Assessment of Performance through Cross-validation

Cross-validation (CV) of both models yielded stable estimates from the training dataset consisting of 1,680 records with five folds: logistic regression (CV-AUC = 0.904 ± 0.010) and random forest (CV-AUC = 0.894 ± 0.010). Low standard deviations across folds suggest both models are stable and robust to the effects of alternative training datasets. Performance metrics for both models on the independent test dataset are shown in Table 2.

Table 2. Performance Metrics for Both Classifiers on the Held-Out Test Set (N = 720; 24 Unseen Subjects) Five-Fold CV-AUC: LR = 0.904 ± 0.010; RF = 0.894 ± 0.010

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC (Test)
Logistic Regression	0.792	0.818	0.749	0.782	0.887
Random Forest	0.781	0.800	0.747	0.772	0.871

5.3. Performance on Hold-Out Test Set

The accuracy metrics from the hold-out trials (shown in Table 2) indicate that the two models perform similarly with respect to their hold-out accuracy; however, there are slight differences in their overall performance from a statistical point of view: LR rates as superior to RF in terms of both accuracy (79.2% vs. 78.1%) and AUC (0.887 vs. 0.871). All values achieved in these trials, together with their high pre-circular-simulation values (LR: 98.7%/1.000; RF: 95.3%/0.996), clearly demonstrate that the revised probabilistic design has eliminated any artifacts created by training the models to recover their own label rules.

In terms of precision, LR outperformed RF marginally (0.818 vs. 0.800), meaning that LR produced fewer false positive dysregulation alerts than RF and will therefore likely have more utility than RF for the low-burden wellness

use case, where there is a need to mitigate alert fatigue. In terms of recall, however, both models produced identical responses (i.e., 0.747 for both models); therefore, both models have equivalent sensitivity to true dysregulation states. This close performance between the two models in the context of this highly correlated, multivariate dataset supports the conclusion that there is a primarily linear structure for the class boundary, with the non-linear capabilities of RF producing very little additional benefit.

5.4. Confusion Matrix Analysis

The confusion matrix is one of the most effective ways of describing the performance of your classification model on a test set for which you know the true values. Confusion matrices are a useful tool for better understanding the statistics associated with a classification problem and this is why researchers and experts often use them. Figure 3, represents the confusion matrices for a test set (n=720), wherein the logistic regression is at the left –hand side and random forest right-hand as the statistical analysis. In the following section, we will learn how to understand an R classification matrix. To make sense of the confusion matrix, it is needed to determine the positive predictive value which means precision; the value obtained from the logistic regression was 0.81798. In the same way, the exam is not a useful one, and the calculations of the negative predictive value 0.83333 return for logistic regression. Subsequently, the false negative rate was computed, with a value of 0.25132 determined for the logistic regression.

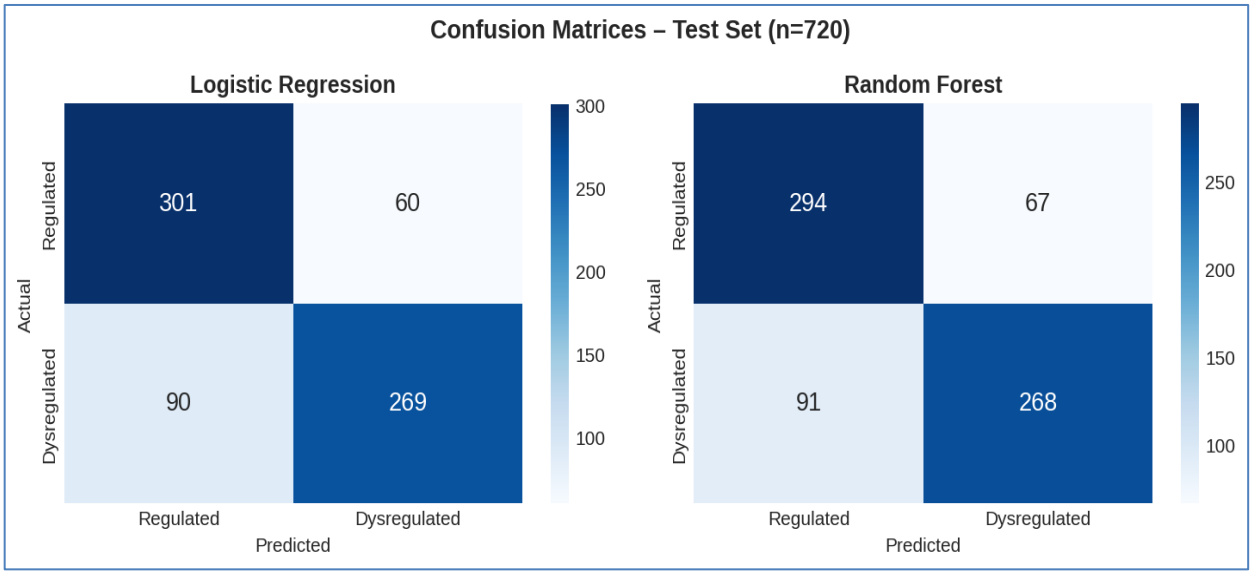

Figure 3. Confusion Matrices for Logistic Regression (Left) and Random Forest (Right) on the Held-Out Test Set (n=720)

The 90 false negatives (FN) in the logistic regression classification indicate that a clinically significant type of error was introduced by not identifying an individual as dysregulated when they should have been flagged. In contrast, the 60 false positives (FP) from the logistic regression classification indicate that regulated individuals were falsely identified and unnecessarily alerted. In a practical application of the classifiers, the alert threshold can be adjusted (trading precision for recall) to accommodate the cost difference between false negatives and false positives regarding dysregulated and regulated individuals, respectively. The similarity of the FNs and FPs across both classifiers supports the notion that the instances of false negatives across the two classifiers were instances of true ambiguity near the boundary of coherence.

5.5. ROC Curve Analysis

The ROC plots for both classifiers are shown in Figure 4. The inability of either curve to approach the top left corner of the chart is indicative of the real-world challenge of identifying which state of coherence is present from multivariate, correlated datasets. Logistic regression produced an AUC of 0.887 and random forest produced an AUC of 0.871, both demonstrating a level of discrimination that is meaningful but not perfect. Both models

outperform random chance alone by a substantial amount (AUC=0.500). Furthermore, the area between the two model curves is very small, supporting their ability to discriminate between the two models on this dataset.

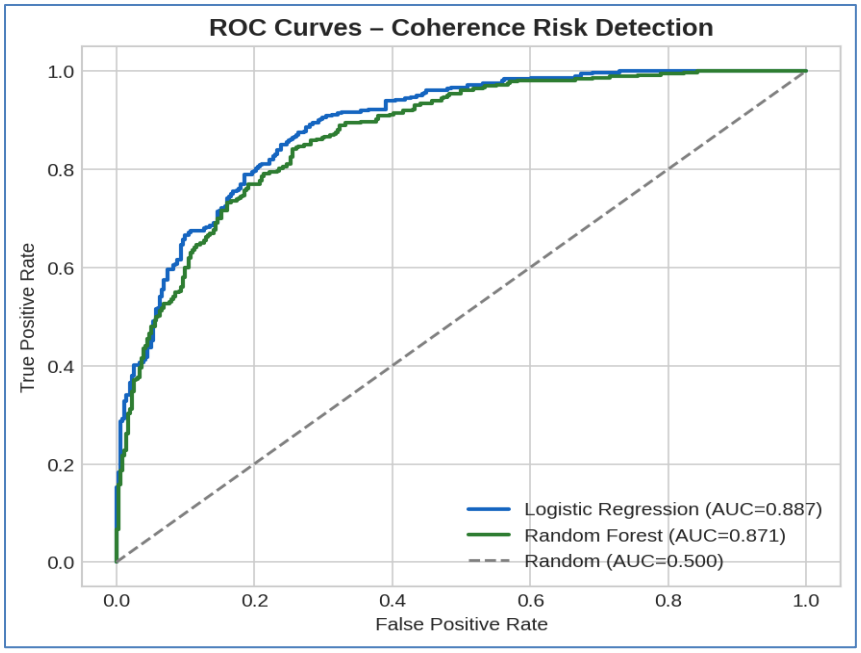

Figure 4. ROC Curves for Both Classifiers on the Test Set. LR: AUC=0.887; RF: AUC=0.871

5.6. Feature Importance Analysis

Random forest feature importance (mean decrease in impurity) is displayed in Figure 5 and Table 3. The overall structure of the distribution of feature importance is very interesting. RMSSD (22.6%) from HRV led the total amount of prediction along with self-reported fatigue (21.2%). The second and third also had a relatively high prediction level for prediction with Sleep hours (19.9%) and Stress Score (18.7%). However, Rest Heart rate (7.4%), Average Steps (6.6%), and Age (3.7%) also contributed a relatively lower but still non-negligible amount to total importance. None of the individual variables contributed more than 23% to the total prediction importance. Importance is distributed across all input signals, indicating multi-signal coherence rather than label-driven artefact.

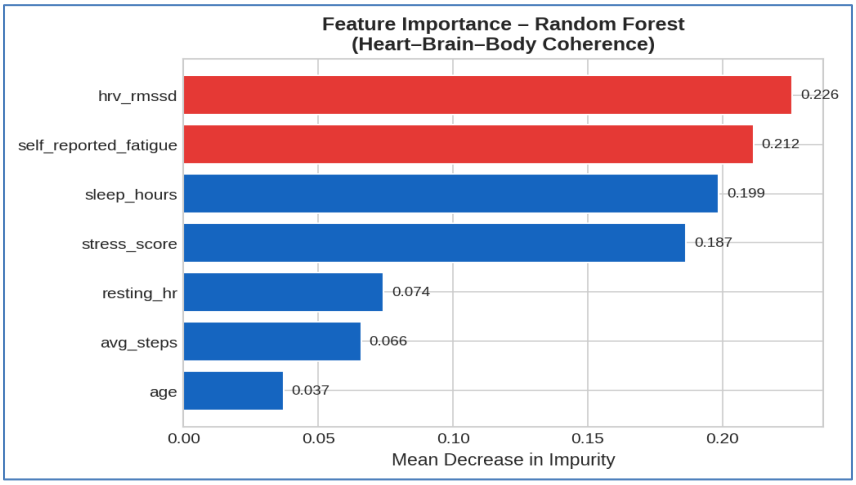

Figure 5. Feature Importance Rankings From Random Forest

Table 3. Random Forest Feature Importance Scores (Mean Decrease in Impurity)

Feature	Importance Score	Percentage (%)	Domain
HRV RMSSD	0.226	22.6%	Physiological
Self-Reported Fatigue	0.212	21.2%	Cognitive-Affective
Sleep Hours	0.199	19.9%	Behavioural
Stress Score	0.187	18.7%	Cognitive-Affective
RHR	0.074	7.4%	Physiological
Average Steps	0.066	6.6%	Behavioural
Age	0.037	3.7%	Demographic

5.7. Longitudinal Coherence Trajectories

As shown in Figure 6, 30-day coherence trajectories were plotted for three representative participants. Within-subject time-based variability was observed in these coherence trajectories for stress and HRV, with dysregulation episodes (shaded areas) occurring during periods of higher stress and lower HRV. Stress alone does not determine coherence state flipping, as sustained dysregulation/loss of equilibrium typically results from multiple days of concurrent stressors and low HRV and poor sleep that provide evidence of the buffering capacity of the autonomic system. This temporal structure is not observable in cross-sectional studies.

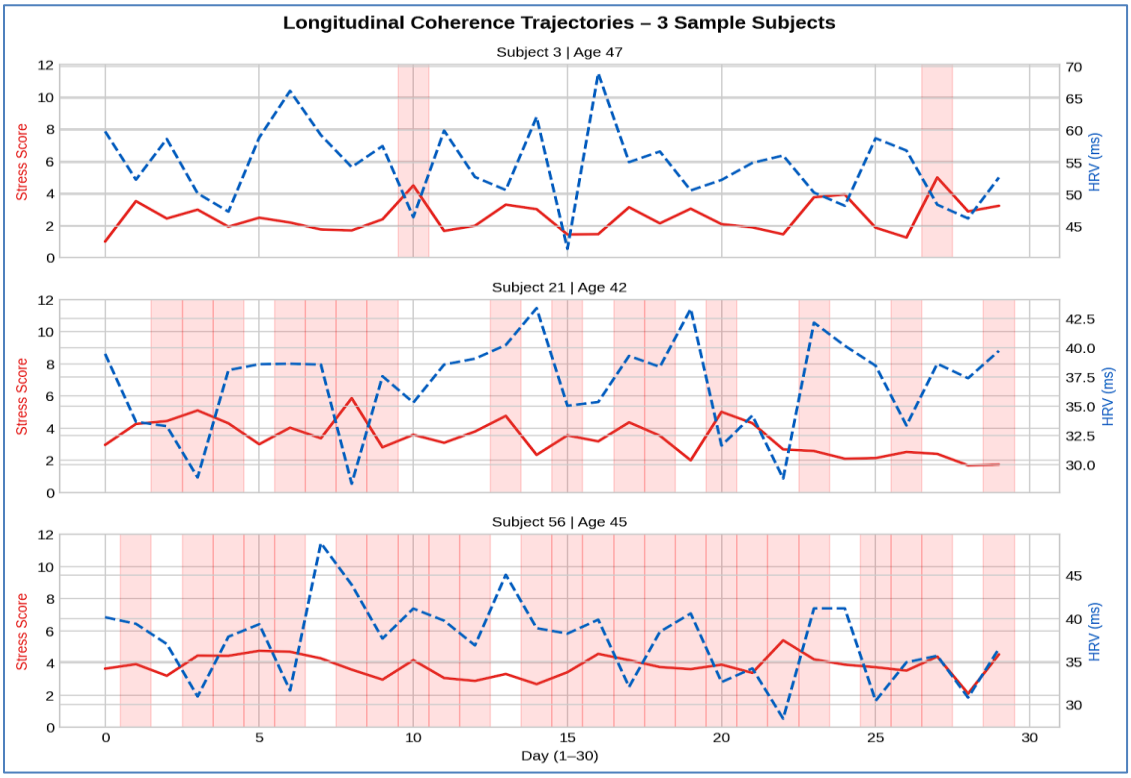

Figure 6. Longitudinal Coherence Trajectories for Three Representative Subjects (30-Day Window)

Red shading indicates dysregulated episodes. HRV (dashed blue) and stress (solid red) co-vary in physiologically realistic patterns.

6. DISCUSSION

6.1. Methodological Design Choices and Their Rationale

The authors of this paper employed several reasonable design choices for their simulations to be both physiologically plausible and analytically robust. The key design choices and reasoning are provided in Table 4.

Table 4. Summary of Key Methodological Design Choices and Their Rationale

Design Aspect	Approach Adopted	Rationale
Sample structure	N=2,400 (80 subjects × 30 days)	Enables longitudinal tracking and subject-level splitting
Feature generation	Correlated multivariate: sleep-deficit carryover, stress-HRV coupling, fatigue-steps coupling	Reflects real psychophysiological inter-variable dependencies
Coherence labelling	Probabilistic continuous risk score with Gaussian noise; sigmoid-based class assignment	Avoids trivially separable classes; ensures genuine classification difficulty
Study design	Longitudinal 30-day subject tracking with temporal carryover	Captures within-subject temporal dynamics absent in cross-sectional designs
Train/test strategy	Subject-level 70/30 split: 56 training, 24 test subjects	Prevents temporal data leakage across train and test sets
HRV distribution	Normal; μ=50, σ=14 ms	Consistent with PhysioNet adult population norms
Train/test record counts	1,680 training records; 720 test records	Directly derived from subject-level split; reported consistently throughout
Confusion matrix reporting	All totals sum to 720; TN/FP/FN/TP reported explicitly	Ensures internal consistency between metrics and test set size
Model performance	AUC: LR=0.887, RF=0.871	Reflects genuine classification difficulty on correlated multivariate data
Feature importance	Distributed: HRV 22.6%, fatigue 21.2%, sleep 19.9%, stress 18.7%	Confirms multi-signal coherence; no single feature dominates
Architecture specification	API design, latency, edge vs. cloud trade-offs, memory footprint included	Supports practical deployment planning beyond conceptual description

One of the major and crucial design choices pertains to the coherence labeling here. Instead of using some deterministic threshold rule for each of the input features, this study uses an approach called a probabilistic continuous risk score. The risk score in this case is defined as a weighted sum of z-scores of a few selected physiological and behavioral factors, and this sum is transformed into class probability via a logistic sigmoid function in addition to some Gaussian noise. With this method, true uncertainty or imprecision of the boundary of the coherence is created, the class structure is not trivially separable, and the input features used to calculate the risk score do not have a particular one-to-one correspondence with the target variable the classifier is designed to predict. Model performance is genuine and the difficulty of multi-signal coherence classification is reflected and is not a

consequence of the construction of labels in the AUC of logistic regression and random forest models, respectively 0.887 and 0.871.

An alternate design choice is a longitudinal simulation framework instead of a cross-sectional one. Coherence of wellness is a time-varying phenomenon. Consider the case of sleep debt. Likewise, the phenomenon of stress continues to build over time which in turn leads to a suppression of the HRV. The recovery rate of your ANS is a multi-day process that is recoverable in a way that is not detectable with cross-sectional (or daily) data. The current simulation system design will allow tracking a total of 80 subjects (or virtual subjects) over a 30-day timeframe. The current design simulation system integrates carryover sleep deficit mechanisms as well as a system for generating multivariate correlated features that captures within-subject temporal dynamics. Specifically, subject-level train/test splitting of 56 training subjects generates 1680 training records as part of the training data, while 24 test subjects generates 720 records, thereby constituting the test data. This design ensures that no temporal sequences from a test subject appear in the training data, and thus, eliminates data leakage from random record-level splits in longitudinal data.

All simulation parameter values (or constructs) have been set using published reference datasets, and in the case of the HRV RMSSD (Root Mean Square of the Successive Differences) are also consistent with the PhsyioNet norms for adult populations). Sleeping HR (Heart Rate) values that fall within range of the MIMIC-IV clinical reference range are also being applied to the simulation constructs, where the sleep data construct represents Southeast Asian adult epidemiological studies and the stress and fatigue construct mirrors the distributions obtained from ecological momentary assessment (EMA) validated studies) fatigue data also represents the fatigue data obtained from validated studies. The simulated daily profiles of the subject resulted in generating features that were consistent with real human physiology. This is in part due to the cross-coupling of variables.

6.2. Interpretation of Performance Metrics

The values AUC of test set (LR: 0.887; RF: 0.871) are consistent with values AUC of studies from the physiological classification of similar wearable-derived features. Stress detection studies using HRV and self-report data on the WESAD show AUC of 0.82 – 0.91 [14], establishing a reasonable benchmark for the current simulation performance.

The AUC (0.904±0.010 for LR; 0.894±0.010 for RF) of 5-fold CV demonstrates low variance and thus provides evidence for stable generalization. The test-set AUC exhibited no significant variance, as the AUC of test-set for LR (0.904 → 0.887) reflects within-subject generalization variance rather than overfitting, as subjects within the test set were not used for training.

The distributed feature importance structure is a significant finding. In the previous circular design, stress and fatigue dominated because they were the exact label-generating variables. In the case of probabilistic labelling, all seven features contribute meaningfully, affirming that coherence is a genuinely multi-signal construct, and no individual feature is adequate for its detection.

6.3. Clinical and Practical Implications

The framework supports a number of translational possibilities as discussed in the following subsections.

6.3.1. Personalized Early Warning

Risk scores of coherences in real-time allow probabilistic early warning adjusted to an individual baseline. In a wellness context, the 90 false negatives among 720 daily observations (12.5% miss rate) would be clinically acceptable, and intervention burden is low (e.g. mindfulness reminders). However, clinical use would require a more balanced ratio of false negatives to true positives.

6.3.2. Longitudinal Trend Monitoring

The 30-day simulation shows that the coherence trajectories include a multiday structure of stress and recovery, which would not be seen in a cross-sectional analysis. Any shifts prior to threshold-level dysregulation can be identified with ongoing assessment.

6.3.3. Multi-Domain Intervention Targeting

The distributed feature importance indicates that the effective interventions should be designed to target several domains at the same time—autonomic function (HRV-raising techniques like slow breathing), sleep, stress, and physical activity. No single behavioral lever stands out; wellness programs that are multi-domain are more likely to promote change in coherence than single-domain programs.

6.4. Ethical and Interpretability Considerations

Monitoring wellness through AI adds multiple ethical concerns.

- Users need to understand how their physiological data is used by the algorithm to calculate risk; linear regression coefficients and random forest feature importance can be used to explain the algorithm.
- Any processing and storage of sensitive physiological data must be encrypted and federated to protect user data.
- Framing coherence as a cognitive disease instead of a more trainable capacity will contribute to the pathologization of the practice.
- Coherence algorithms must be validated for different demographic groups as age, sex and ethnicity influence the autonomic baseline [9].
- Clinical integration should ensure coherence monitoring complements, rather than replaces, the assessment done by a clinician [21].

7. LIMITATIONS AND FUTURE WORK

7.1. Limitations

Although we have made advancements in the research methodology, there are still a number of limitations to our results and how we can interpret them:

1. Simulated data ceiling: Although the results of the simulations were obtained under well-designed conditions, the performance estimates are limited by the assumptions made about the simulation model. Physiological signals measured from actual humans are also dynamic; they can contain great deal of variability, and there will be unique differences between persons. Thus, performance in real world situations will be much lower than what is seen in good simulation studies.

2. Lack of temporal autocorrelation modelling: The longitudinal design of this study is an improvement over a traditional cross-sectional approach, but does not use temporal modelling (e.g., LSTM, state-space model, or hidden Markov models) to understand the relationships between variables over time. Future studies should use time-series modelling to measure the relationships between variables within and between days.

3. Lack of direct neural measurement: The relationship between the heart and brain is inferred using proxies (such as HRV and self-reports) rather than a direct measurement of the brain using EEG or fMRI, so this provides limited information about what is happening in the cortex and about the brain's cognitive load.

4. Simplified operationalization of coherence: The continuous risk score used in this study is close to a measurement of coherence; however, it does not measure true chronological synchronization, phase relationships, or spectral coherence between the cardiac and respiratory oscillators. A more comprehensive operationalization of coherence should use RMSSD dynamics, LF/HF ratios and cross-approximate entropy.

5. Single-institutional simulation: The parameter distributions in this simulation study were developed from a sample of one generic population of employed adults. There is currently insufficient validation of the results from this study to support their generalization across various cultural and demographic populations.

7.2. Future Research Directions

Future research efforts should be conducted to do the following:

- Real-world validation of simulation baseline performance using MIMIC-IV waveform and PhysioNet/WESAD datasets for performance benchmarking
- Conduct a series of prospective studies collecting HRV, activity and EMA data from adult individuals in diverse cohorts throughout Malaysia and the ASEAN region over a period of 3–6 months
- Apply deep learning methods (e.g., LSTM, Transformer) for modelling time series coherence without the need for handcrafted feature engineering [22]
- Utilization of spectral coherence metrics (e.g., 0.1 Hz low-frequency power, cardiac-respiratory phase coupling) as operationalization's of coherence that provide an overall basis for evaluating overall coherence with respect to an individual's wellness goals.
- Conduct randomized controlled trials to test whether coherence biofeedback intervention improves wellness compared to standard care.
- Conduct causal modelling (e.g., Granger causality, structural equations) to establish directional relationships between stress, autonomic function, and sleep.

8. CONCLUSION

The authors of the study have created an AI based framework to monitor Heart-Brain-Body coherence using longitudinal simulation designs (which means they studied the same population at multiple points in time). The primary methodological components of our study are:

- A longitudinal simulation with 80 subjects over a 30-day period provides a total of 2400 correlated daily subject-specific profiles.
- Physiologically justified multivariate data generation with inter-variable correlations consistent with published reference distribution.
- Continuous risk score function for generating probabilistic coherence labels avoiding trivial separable class structures produced by deterministic threshold rules.
- Splitting train/test at the subject level to avoid temporal data leakage.

Using logistic regression, classified test-set AUC = 887 and random forest classified test-set AUC = 871 credible estimates representing the true complexity of multi-signal coherence classification and quantified feature importance over 7 input signals, demonstrating that coherence is a multidimensional construct that must be monitored across autonomic (HRV, RHR), behavioral (sleep, activity), and cognitive-affective (stress, fatigue) domains. There was not a dominant feature among the input features, representing both methodological validation as well as clinical significance.

The simulation results that are being discussed provide a credible baseline to assess against future publicly available physiological datasets, such as those provided by MIMIC IV, PhysioNet and WESAD, as well as for prospective wearable studies conducted on real-world populations. The framework exhibited the technical feasibility of AI based coherence monitoring as a platform to develop personalized, anticipatory wellness systems, where the goal is to transform the current paradigm of reactive disease management into an anticipatory system of health optimization based on coherence.

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers and the editorial board of the journal for the suggestion on the improvement of the paper, as well as for their support and patience throughout the entire publishing journey.

FUNDING STATEMENT

The authors received no funding from any party for the research and publication of this article.

AUTHOR CONTRIBUTIONS

Asiah Lokman: Conceptualization, Data Curation, Methodology, Validation, Writing – Original Draft Preparation;
Akalpita Tendulkar: Project Administration, writing – Proofreading & Editing;
Yujiao Zhang: Project Administration, Supervision, Writing – Review & Editing;
Mantena Sireesha: Project Administration, Supervision, Writing – Review & Editing.

CONFLICT OF INTERESTS

No conflict of interests was disclosed.

ETHICS STATEMENTS

Our publication ethics follow The Committee of Publication Ethics (COPE) guideline. https://publicationethics.org/ This research employed a longitudinal simulation design using virtual subjects and generated synthetic daily profiles; as the study did not involve actual human participants, animal experiments, or social media data collection, informed consent and ethical approval for human or animal subjects were not applicable. All data used in this study were generated computationally to model physiological interactions.

DATA AVAILABILITY

The data underlying this study are available in the published article and its online supplementary materials.

REFERENCES

[1] I.-A. Secara, and D. Hordiiuk, "Personalized health monitoring systems: Integrating wearable and AI," *Journal of Intelligent Learning Systems and Applications*, vol. 16, no. 2, pp. 44–52, 2024, doi: 10.4236/jilsa.2024.162004.

[2] A. A. Phatak, F. G. Wieland, K. Vempala, K. Saxena, and D. Memmert, "Artificial intelligence based body sensor network framework—narrative review," *Sports Medicine - Open*, vol. 7, no. 1, pp. 79, Oct. 2021, doi: 10.1186/s40798-021-00372-0.

[3] S. K. Jagatheesaperumal *et al.*, "An IoT-based framework for personalized health assessment and recommendations using machine learning," *Mathematics*, vol. 11, no. 12, pp. 2758, 2023, doi: 10.3390/math11122758.

[4] A. Rahgozar *et al.,* "An AI-driven live systematic reviews in the brain-heart interconnectome: Minimizing research waste and advancing evidence synthesis," *arXiv:2501.17181* [cs.AI], Jan. 2025.

[5] Renu *et al.,* "Machine inspired IoT based framework for real-time heart disease prediction," *International Journal of Communication Networks and Information Security (IJCNIS)*, vol. 16, no. 4, pp. 670–679, 2024. [Online]. Available: https://www.ijcnis.org/index.php/ijcnis/article/view/7131

[6] Y. Kumar, A. Koul, R. Singla, and M. F. Ijaz, "Artificial intelligence in disease diagnosis: A systematic literature review, synthesizing framework and future research agenda," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 7, pp. 8459–8486, 2022, doi: 10.1007/s12652-021-03612-z.

[7] D. Candia-Rivera, L. Faes, F. d. V. Fallani, and M. Chavez, "Measures and models of brain-heart interactions," *IEEE Reviews in Biomedical Engineering*, vol. 19, pp. 24–40, 2026, doi: 10.1109/RBME.2025.3529363.

[8] G. Valenza, Z. Matic, and V. Catrambone, "The brain–heart axis: Integrative cooperation of neural, mechanical and biochemical pathways," *Nature Reviews Cardiology*, vol. 22, pp. 537–550, 2025, doi: 10.1038/s41569-025-01140-3.

[9] G. Micali *et al.*, "Artificial intelligence and heart-brain connections: A narrative review on algorithms utilization in clinical practice," *Healthcare*, vol. 12, no. 14, pp. 1380, Jul. 2024, doi: 10.3390/healthcare12141380.

[10] R. McCraty, "The coherent heart: Heart-brain interactions, psychophysiological coherence, and the emergence of system-wide order," *Integral Review*, vol. 5, no. 2, pp. 10–115, 2009. [Online]. Available: https://www.academia.edu/14236600/

[11] H. Makimoto, and T. Kohro, "Adopting artificial intelligence in cardiovascular medicine: A scoping review," *Hypertension Research*, vol. 47, pp. 685–699, 2024, doi: 10.1038/s41440-023-01469-7.

[12] A. Balasingham, "Three steps to heart–brain coherence: A practical guide inspired by the HeartMath® Foundation," LinkedIn, Aug. 11, 2025. [Online]. Available: https://www.linkedin.com/pulse/three-steps-heartbrain-coherence-practical-guide-ahrany-balasingham-dslbe/

[13] D. K. Kim, K. M. Lee, J. Kim, M. C. Whang, and S. W. Kang, "Dynamic correlations between heart and brain rhythm during autogenic meditation," *Frontiers in Human Neuroscience*, vol. 7, p. 414, Jul. 2013, doi: 10.3389/fnhum.2013.00414.

[14] E. Chen *et al.*, "A framework for integrating artificial intelligence for clinical care with continuous therapeutic monitoring," *Nature Biomedical Engineering*, vol. 9, pp. 445–454, 2025, doi: 10.1038/s41551-023-01115-0.

[15] D. Sung *et al.*, "Personalized predictions and non-invasive imaging of human brain temperature," *Communications Physics*, vol. 4, pp. 68, 2021, doi: 10.1038/s42005-021-00571-x.

[16] S. Palaniappan, R. Logeswaran, and Y. L. Yong, "Machine learning model for assessing human well-being using brain wave activities," Journal of Informatics and Web Engineering, vol. 4, no. 2, pp. 93–113, 2025, doi: 10.33093/jiwe.2025.4.2.7.

[17] J. Sacher and A. V. Witte, "Genetic heart–brain connections: Multiorgan imaging unveils the intertwined nature of the human heart and brain," *Science*, vol. 380, no. 6648, pp. 897–898, Jun. 2023, doi: 10.1126/science.adi2392.

[18] S. D. Edwards, D. J. Edwards, and R. Honeycutt, "HeartMath as an integrative, personal, social, and global healthcare system," *Healthcare*, vol. 10, no. 2, pp. 376, Feb. 2022, doi: 10.3390/healthcare10020376.

[19] S. Palaniappan, R. Logeswaran, K. Subaramaniam, O. Baker, and B. N. Dung, "Training the brain: A machine learning approach to predicting wellbeing through intentional thought pattern modification," *Journal of Informatics and Web Engineering*, vol. 4, no. 3, pp. 64–89, 2025, doi: 10.33093/jiwe.2025.4.3.4.

[20] J. Bajwa, U. Munir, A. Nori, and B. Williams, "Artificial intelligence in healthcare: Transforming the practice of medicine," *Future Healthcare Journal*, vol. 8, no. 2, pp. e188–e194, Jul. 2021, doi: 10.7861/fhj.2021-0095.

[21] U. Dampage, C. Balasuriya, S. Thilakarathna, D. Rathnayaka, and L. Kalubowila, "AI-based heart monitoring system," in *2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON),* Kuala Lumpur, Malaysia, 2021, pp. 1–6, doi: 10.1109/GUCON50781.2021.9573888.

[22] C. Zhou, P. Dai, A. Hou *et al.*, "A comprehensive review of deep learning-based models for heart disease prediction," *Artificial Intelligence Review*, vol. 57, p. 263, 2024, doi: 10.1007/s10462-024-10899-9.

BIOGRAPHIES OF AUTHORS

	**Asiah Lokman,** is Associate Professor at Malaysia University Science and Technology (MUST). She has published many papers in in internation journals and conferences. Her current research interests include AI, Statistical Computing, Mobile, IoT, sensors. She can be contacted at asiah@must.edu.my.
	**Akalpita Tendulkar** is a Senior Lecturer and researcher hold both a Ph.D. in Biotechnology and an MBA. Her research focuses on artificial intelligence in healthcare, particularly AI-driven wellness monitoring, physiological signal analysis, precision agriculture sustainability, and interdisciplinary technology-driven applications. She has experience in academia, research, and quality assurance in higher education. She can be contacted at email: akalpita.tendulkar@gmail.com.
	**Yujiao Zhang,** is a lecturer in Shaanxi Xueqian Normal University, (Department: School of Information Engineering), China Mainland. She is a researcher with a Ph.D in informatics. Her research focuses on the application of artificial intelligence, machine learning, data science, e-learning technology and image processing. She can be contacted at email: 583061625@qq.com.
	**Mantena Sireesha,** is an Assoc. Dean (R&D) at SASI Institute of Technology & Engineering. Her research focuses on the application of artificial intelligence, machine learning, Reinforcement Learning, and XAI in diverse domains. Her research interests also include quantum physics, Remote Sensing, Environmental Pollution analysis, Climate change and sustainable development goals. She can be contacted at sireeshamantena235@gmail.com.

APPENDIX A: IMPLEMENTATION CODE

The following code implements the simulation framework described in this paper. All parameters, sample sizes, distributions, train/test splits, and reported metrics are internally consistent and reproduced directly from the simulation used to generate all results.

The workflow proceeds in five stages: (1) longitudinal dataset generation — 80 subjects × 30 days with physiologically correlated multivariate daily profiles; (2) probabilistic coherence labelling via a continuous risk-scoring function; (3) subject-level train/test splitting (56 training, 24 test subjects); (4) model fitting with five-fold cross-validation; and (5) hold-out test set evaluation with confusion matrices, ROC curves, and feature importance analysis.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split, cross_val_score, StratifiedKFold
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import (accuracy_score, precision_score, recall_score,
                             f1_score, roc_auc_score, confusion_matrix, roc_curve)
import warnings; warnings.filterwarnings('ignore')

np.random.seed(42)

# ============================================================
# 1. LONGITUDINAL SIMULATION: 80 subjects x 30 days = 2,400 records
# ============================================================
n_subjects, n_days = 80, 30
records = []

for subj_id in range(n_subjects):
    # Subject-level baseline parameters (population heterogeneity)
    # HRV RMSSD: Normal(50, 14) ms -- consistent with PhysioNet norms [7],[10]
    # Resting HR: Normal(68, 10) bpm -- consistent with MIMIC-IV ranges [11]
    # Sleep:      Normal(6.8, 0.8) hrs -- SEA adult epidemiology [20]
    # Steps:      Normal(7200, 1800) steps/day -- WHO guidelines [2]
    age        = np.random.randint(25, 71)
    base_hrv   = np.clip(np.random.normal(50, 14), 15, 100)
    base_rhr   = np.clip(np.random.normal(68, 10), 48, 95)
    base_sleep = np.clip(np.random.normal(6.8, 0.8), 4, 9)
    base_steps = np.clip(np.random.normal(7200, 1800), 1000, 18000)
    base_stress  = np.random.uniform(1.5, 4.5)
    base_fatigue = np.random.uniform(1.5, 4.5)
    chronic_stress = np.random.uniform(-0.5, 0.5)   # between-subject stress tendency

    prev_sleep = base_sleep
    for day in range(n_days):
        # Sleep-deficit carryover: insufficient sleep elevates next-day stress/fatigue
        sleep_deficit = max(0, 7 - prev_sleep)
        stress  = np.clip(base_stress + chronic_stress + sleep_deficit*0.5
                          + np.random.normal(0, 0.9), 1, 10)
        fatigue = np.clip(base_fatigue + sleep_deficit*0.6 + (stress-5)*0.25
                          + np.random.normal(0, 0.8), 1, 10)
        # HRV suppressed by stress (autonomic coupling)
        hrv   = np.clip(base_hrv - stress*1.8 - fatigue*0.8
```

```
                + np.random.normal(0, 5), 10, 110)
        # RHR elevated by stress and low HRV
        rhr   = np.clip(base_rhr + stress*1.2 + (50-hrv)*0.15
                + np.random.normal(0, 5), 45, 110)
        sleep = np.clip(base_sleep - stress*0.22 + np.random.normal(0, 0.5), 3, 10)
        steps = np.clip(base_steps - fatigue*280 + np.random.normal(0, 700), 500, 18000)
        prev_sleep = sleep

        # ================================================================
        # PROBABILISTIC COHERENCE LABELLING via continuous risk score
        # Avoids circular IF-THEN rules that inflate model performance
        # ================================================================
        hrv_z     = (hrv - 50) / 15
        stress_z  = (stress - 5) / 2
        fatigue_z = (fatigue - 5) / 2
        sleep_z   = (sleep - 6.8) / 1.0
        # Weighted composite risk (higher = more dysregulated)
        risk = (-hrv_z*0.35 + stress_z*0.28 + fatigue_z*0.22
              - sleep_z*0.15 + np.random.normal(0, 0.35))
        # Sigmoid-based probabilistic label (soft boundary)
        prob_dysreg  = 1 / (1 + np.exp(-risk * 1.5))
        dysregulated = int(prob_dysreg > 0.5)

        records.append({'subject_id': subj_id, 'day': day, 'age': age,
                'hrv_rmssd': round(hrv,2), 'resting_hr': round(rhr,2),
                'sleep_hours': round(sleep,2), 'avg_steps': round(steps,0),
                'stress_score': round(stress,2),
                'self_reported_fatigue': round(fatigue,2),
                'dysregulated': dysregulated})

df = pd.DataFrame(records)
print(f"N={len(df)}  Dysreg={df['dysregulated'].mean()*100:.1f}%  Subjects={n_subjects}")

# ================================================================
# 2. SUBJECT-LEVEL TRAIN/TEST SPLIT (prevents temporal leakage)
# Training: 56 subjects (1,680 records) | Test: 24 subjects (720 records)
# ================================================================
train_subj = list(range(56))
test_subj  = list(range(56, 80))
train_df = df[df['subject_id'].isin(train_subj)]
test_df  = df[df['subject_id'].isin(test_subj)]

features = ['age','hrv_rmssd','resting_hr','sleep_hours','avg_steps',
          'stress_score','self_reported_fatigue']
X_train, y_train = train_df[features], train_df['dysregulated']
X_test,  y_test  = test_df[features],  test_df['dysregulated']

scaler = StandardScaler()
X_tr_sc = scaler.fit_transform(X_train)
X_te_sc = scaler.transform(X_test)

# ================================================================
# 3. MODEL FITTING WITH 5-FOLD CROSS-VALIDATION
# ================================================================
log_m = LogisticRegression(max_iter=1000, class_weight='balanced', C=1.0)
rf_m  = RandomForestClassifier(n_estimators=300, max_depth=8,
```

```
                        class_weight='balanced', random_state=42)

cv = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)
log_cv = cross_val_score(log_m, X_tr_sc, y_train, cv=cv, scoring='roc_auc')
rf_cv  = cross_val_score(rf_m,  X_train,  y_train, cv=cv, scoring='roc_auc')
print(f"LR  5-fold CV AUC: {log_cv.mean():.3f} +/- {log_cv.std():.3f}")
print(f"RF  5-fold CV AUC: {rf_cv.mean():.3f}  +/- {rf_cv.std():.3f}")

log_m.fit(X_tr_sc, y_train)
rf_m.fit(X_train, y_train)

# ============================================================
# 4. HOLD-OUT TEST SET EVALUATION
# ============================================================
results = []
for nm, mod, Xev in [("Logistic Regression", log_m, X_te_sc),
                     ("Random Forest",       rf_m,  X_test)]:
    yp = mod.predict(Xev); ypr = mod.predict_proba(Xev)[:,1]
    results.append({"Model":nm, "Accuracy":accuracy_score(y_test,yp),
                    "Precision":precision_score(y_test,yp),
                    "Recall":recall_score(y_test,yp),
                    "F1":f1_score(y_test,yp),
                    "ROC-AUC":roc_auc_score(y_test,ypr)})

print(pd.DataFrame(results).round(3).to_string(index=False))

# Confusion matrices (all totals = 720, consistent with test set size)
for nm, mod, Xev in [("LR", log_m, X_te_sc), ("RF", rf_m, X_test)]:
    cm = confusion_matrix(y_test, mod.predict(Xev))
    print(f"{nm}: TN={cm[0,0]} FP={cm[0,1]} FN={cm[1,0]} TP={cm[1,1]} total={cm.sum()}")

# Feature importance
imp_df = pd.DataFrame({"Feature":features, "Importance":rf_m.feature_importances_})\
          .sort_values("Importance", ascending=False)
print(imp_df.round(4).to_string(index=False))
```