Deep Learning-based Classification of Patients with Postural Orthostatic Tachycardia Syndrome using Wearable ECG and Accelerometer Data

Hyunjun Choi[1,†], Nicholas Matsumoto[1,†], Xi Li[1], Debbie Teodorescu[2], Anxhela Kote[2], Min-Jing Yang[2], Xiao Liu[2], Miguel E. Hernandez[1,†], Jason H. Moore[1*], Graciela Gonzalez Hernandez[1*], Peng-Sheng Chen[2*]

[1]*Department of Computational Biomedicine, Center for Artificial Intelligence Research and Education, Cedars Sinai Medical Center,*
700 N. San Vicente Blvd., Pacific Design Center, Suite G-541H, West Hollywood, USA
[2]*Department of Cardiology, Smidt Heart Institute, Cedars-Sinai Medical Center,*
8700 Beverly Blvd., Davis 1016, Los Angeles CA 90048
[†]*E-mail: AISupport@csmc.edu*
www.cedars-sinai.org

Postural Orthostatic Tachycardia Syndrome (POTS) is a chronic autonomic disorder characterized by chronic ($>$ 3 months) orthostatic intolerance and an increase in heart rate (HR) of $\geq$ 30 beats per minute (bpm) without orthostatic hypotension. Traditional diagnostic approaches, such as the active standing or tilt-table test, are typically conducted under controlled clinical conditions, limiting their ability to capture the natural variability of symptoms and the intricate physiological responses occurring in daily life. These tests may cause patient discomfort, dizziness, nausea, or syncope. Furthermore, they are time-consuming and cannot be used as a screening tool for POTS. To address these limitations, this study explored wearable devices that continuously collect physiological data-specifically, electrocardiogram (ECG) and accelerometer (ACC)-derived metrics-from POTS patients and healthy controls during routine daily activities. Physiological features around posture-change events identified in the data were processed and used to train and test a baseline deep learning model. The model demonstrated promising performance in accurately differentiating POTS patients from healthy controls in a relatively small cohort (66 from POTS patients and 20 from controls), indicating its potential as a feasibility study for clinical decision support. Future studies involving larger and more diverse samples under varying clinical conditions would be necessary to enhance the robustness and viability of our diagnostic model.

.

Keywords: Postural Orthostatic Tachycardia Syndrome, POTS, Wearable devices, Diagnostic devices, Deep learning, Ambulatory monitoring, Real-world data, Daily activities, Autonomic dysfunction

1. Introduction

Postural Orthostatic Tachycardia Syndrome (POTS) is a chronic autonomic disorder characterized by chronic ($>$ 3 months) orthostatic intolerance and an increase in heart rate (HR) of

*Co-senior authors

≥30 beats per minute (bpm) without orthostatic hypotension during orthostatic tests.[1] It is frequently associated with dizziness, fatigue, palpitations, syncope, presyncope, and multiple other symptoms. POTS was estimated to have a prevalence of 0.2%,[1] but its incidence dramatically increased during the COVID-19 pandemic[2,3] and persists to this date. It is estimated that 2% to 14% of COVID-19 survivors develop POTS, and 9% to 61% experience POTS-like symptoms.[3] POTS can also occur after receiving mRNA vaccines for COVID-19.[4,5] The combined effects of COVID-19 infection and vaccination exposures have substantially broadened the at-risk population. Currently, the diagnosis of POTS primarily relies on clinical evaluations like the tilt-table test or active standing tests, conducted in controlled laboratory or clinical settings.[6] However, these traditional methods have notable limitations, as they are typically performed under artificial testing conditions, often failing to accurately represent the genuine variability of symptoms and the complex physiological responses patients experience daily.[7,8] The abrupt postural changes and prolonged upright positioning required during the tilt-table test may cause considerable patient discomfort, including dizziness, nausea, headache, and in some cases, temporary loss of consciousness. These tests are time-consuming and cannot be used as a screening tool in the general population. It is therefore highly desirable to develop alternative diagnostic approaches that are less burdensome and more comfortable for patients.[9,10]

To overcome these diagnostic constraints, the present study aims to evaluate the diagnostic potential of data derived from wearable technology. This technology collects physiological data from patients while they are engaged in their routine daily activities. This data serves as the foundation for developing and evaluating deep learning models intended to aid clinical decision support in POTS by detecting physiological changes occurring naturally in daily life with wearable technology.

This study introduces a novel approach that automatically identifies posture-change points and leverages ECG- and ACC-derived features from these event windows in combination with deep learning architectures to classify POTS.

2. Methods

The study includes 66 POTS patient measurements and 20 control measurements. Participants were provided with wearable monitoring devices, specifically the Faros 180 ECG monitor, capable of continuously collecting ECG and ACC data.[6] Based on ECG and ACC data, we extracted features that identify posture-change events (referred to as *changepoints* or *CPs*) that reflect physiological characteristics —including Heart Rate Change (HRC), Skin Sympathetic Nerve Activity (SKNA), and movement intensity. These features were organized into longitudinal datasets reflecting physiological state changes of each participant over time, and were used as input data for artificial intelligence (AI) models. Fig. 1 provides an overview of the proposed data-driven diagnostic framework, highlighting key stages: accelerometer-based orientation-invariant transformation (OIT)[11,12] feature extraction, changepoint detection, ECG and ACC feature extraction, sequential data preparation, and model-based classification and visualization.

2.1. *Study Population and Data Collection*

This study analyzed data from 66 POTS and 20 control measurements and was conducted with approval of the Institutional Review Board (IRB) at Cedars-Sinai Medical Center. All participants provided informed written consent prior to participation. Physiological data were recorded using a wearable biosignal monitor (Faros 180, Bittium Corp., Finland), placed on the participants' chest.[13] This device continuously captured ECG signals sampled at 1 kHz and triaxial ACC data sampled at 25 Hz along three perpendicular axes (X, Y, Z), enabling the assessment of body posture and movement over at least 24 hours. Participants were advised to perform their usual daily activities without any restrictions to ensure that the collected physiological data accurately reflected their typical daily routines.

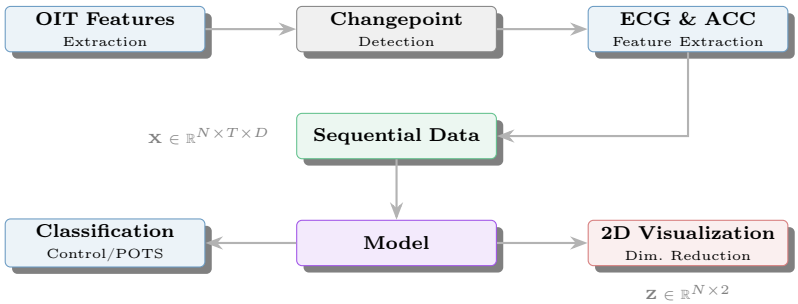

Fig. 1. Overview of our diagnostic framework. **N**: total number of subjects, **T**: sequence length, **D**: dimensionality of combined ECG and ACC features, **X**: input data representing sequential ECG and ACC features, and **Z**: 2D visualization coordinates after dimensionality reduction. The model performs end-to-end learning for both classification and representation learning.

2.2. *Data Preprocessing and Feature Extraction*

The data processing pipeline in this study can be divided into four stages: (1) noise reduction of ECG and accelerometer signals, (2) orientation-invariant transformation of accelerometer signals, (3) detection of posture-change events, and (4) extraction of ECG and ACC features around detected events.

The data collected from wearable sensors was preprocessed using already established algorithms[6,11,12,14,15] to extract features from physiological signals and to reduce noise and enhance signal quality. ECG- and ACC-based features were extracted after preprocessing, as summarized in Table 1.

For accelerometer signals, preprocessing involved applying a moving average filter with a 31-sample window to remove high-frequency noise. Subsequently, orientation-invariant ACC features were derived using the heuristic orientation-invariant transformation (Heuristic OIT).[11,12] The Heuristic OIT transforms raw three-dimensional accelerometer data into nine orientation-robust descriptors, including vector magnitudes, temporal difference norms (first- and second-order), angles between successive vectors, and angular changes computed from vector cross products. This transformation ensures robustness against variations in sensor orientation. Thus, employing Heuristic OIT features as input to the PELT (Pruned Exact Linear

Time) algorithm,[14] which identifies changepoints based on statistical shifts within time-series data, facilitates a more reliable detection of movement-related changepoints by minimizing interference from sensor orientation changes or arbitrary rotations. For ECG signals, noise reduction involved applying a 0.5–150 Hz bandpass filter to mitigate both low-frequency baseline drift and high-frequency artifacts, following the method described by Lee et al. (2022).[6]

Feature Category	Key Characteristics
ECG-based Features	
Statistical	Amplitude statistics, autocorrelation, trend analysis, Hjorth parameters
Spectral	Power spectrum, frequency entropy, wavelet decomposition
Morphological	Zero crossings, peak-trough metrics, waveform shape
Physiological	Heart rate change, SKNA burst
ACC-based Features	
Movement Intensity	Magnitude, jerk, SMA, RMS, energy metrics
Postural	Stability measures, dynamic acceleration, OIT statistics
Spatial	Rotation vector, trajectory curvature, movement patterns
Temporal	Decay time, spectral entropy, windowed analysis

Table 1: Summary of Feature Categories

Based on these Heuristic OIT features, postural transition-related changepoints were identified using the PELT algorithm.[14] This method detects changepoints by optimizing a predefined cost function that balances segmental statistical fidelity with model complexity through a penalty term. Analysis windows, varying from tens of seconds to several minutes, were subsequently defined around each detected changepoint, and physiological response characteristics were extracted from segments meeting established signal quality standards. The comprehensive methodological process employed for changepoint detection is outlined in Algorithm 1.

Measure	**Formula**	**Description**
SNR	$10 \log_{10}(P_{\text{signal}}/P_{\text{noise}})$	Signal-to-noise ratio (power ratio in dB)
TM	$\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2 \sum(y_i-\bar{y})^2}}$	Morphological similarity (template x_i, beat y_i)
RMSSD	$\sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(RR_{i+1}-RR_i)^2}$	RR variability (successive RR intervals)
SDNN	$\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(RR_i-\overline{RR})^2}$	RR variability (standard deviation)
PE	$-\sum_i p_i \ln(p_i)$	Complexity entropy (p_i: ordinal probability)

Table 2: ECG Signal Quality and Variability Metrics

Orphanidou et al. (2014)[15] introduced a signal quality index (SQI) designed to quantitatively evaluate ECG signals obtained from wearable sensors, aiming to ensure the reliable extraction of heart rates. Their approach combined physiological plausibility assessments with template matching techniques to determine the reliability of ECG signal segments. Inspired by this method, our study evaluated ECG segments within patient-specific datasets using signal quality metrics commonly employed in ECG research, including signal-to-noise ratio (SNR),[16] template matching (TM),[15] and permutation entropy (PE).[17] Additionally, the root

mean square of successive differences (RMSSD)[18,19] and the standard deviation of NN intervals (SDNN)[18,19] were also utilized. The ECG signal quality and physiological metrics used for evaluating ECG segments in this study are summarized in Table 2 and were calculated following the procedure outlined in Algorithm 2. Algorithm 2 filters dataset D of ECG segments $s_{j,k}$ from subjects P_j using threshold Θ to produce D_{filtered}.

Algorithm 1 Change Point Detection via OIT[11]-PELT[14]

Require: $\boldsymbol{A} \in \mathbb{R}^{3\times M}$ ▷ Triaxial ACC data
Require: W ▷ Moving average window size
Ensure: $\mathcal{C} = \{\tau_1, \ldots, \tau_K\}$ ▷ Change points
I: $\tilde{\boldsymbol{A}} = \mathcal{F}_{\text{MA}}(\boldsymbol{A}, W)$ ▷ Smoothing with window W
II: $\boldsymbol{F}_{\text{OIT}} = f_{\text{OIT}}(\tilde{\boldsymbol{A}})$ ▷ OIT feature extraction
III: $\mathcal{C} = f_{\text{PELT}}(\boldsymbol{F}_{\text{OIT}})$ ▷ PELT detection
return $\mathcal{C}$

For these selected ECG segments, statistical metrics, Hjorth parameters, autocorrelation, linear trends, spectral characteristics, wavelet-based features, and heart rate changes were subsequently extracted. Additionally, features related to SKNA derived from ECG signals[6] were calculated. These included metrics such as burst frequency, amplitude, duration, mean area, threshold, maximum burst amplitude, and baseline statistics.

Algorithm 2 Multi-Criteria ECG Quality Assessment

Require: $\mathcal{D} = \{P_j\}_{j=1}^{N}$, $P_j = \{s_{j,k}\}_{k=1}^{M_j} \subset \mathbb{R}^T$
Require: $\Theta = \{\theta_{\text{SNR}}, \theta_{\text{corr}}, \theta^{\text{min}}_{\text{RMSSD}}, \theta^{\text{max}}_{\text{RMSSD}}, \theta^{\text{min}}_{\text{SDNN}}, \theta^{\text{max}}_{\text{SDNN}}, \theta_{\text{PE}}\}$
Ensure: $\mathcal{D}_{\text{filtered}} = \{P_j^{\text{filtered}}\}_{j=1}^{N}$
for all $P_j \in \mathcal{D}$ **do**
  $P_j^{\text{filtered}} \leftarrow \emptyset$
  for all $s_{j,k} \in P_j$ **do**
    $\pi_1(s_{j,k}) := \mathbb{I}[\text{SNR}(s_{j,k}) > \theta_{\text{SNR}}]$
    $\pi_2(s_{j,k}) := \mathbb{I}[\rho_{\text{template}}(s_{j,k}) > \theta_{\text{corr}}]$
    $\pi_{3a}(s_{j,k}) := \mathbb{I}[\text{RMSSD}(s_{j,k}) \in [\theta^{\text{min}}_{\text{RMSSD}}, \theta^{\text{max}}_{\text{RMSSD}}]]$
    $\pi_{3b}(s_{j,k}) := \mathbb{I}[\text{SDNN}(s_{j,k}) \in [\theta^{\text{min}}_{\text{SDNN}}, \theta^{\text{max}}_{\text{SDNN}}]]$
    $\pi_3(s_{j,k}) := \pi_{3a}(s_{j,k}) \wedge \pi_{3b}(s_{j,k})$ ▷ Both HRV metrics valid
    $\pi_4(s_{j,k}) := \mathbb{I}[H_{\text{perm}}(s_{j,k}) < \theta_{\text{PE}}]$
    if $\prod_{i=1}^{4} \pi_i(s_{j,k}) = 0$ **then**
      $P_j^{\text{filtered}} \leftarrow P_j^{\text{filtered}} \cup \{s_{j,k}\}$
    end if
  end for
end for
return $\mathcal{D}_{\text{filtered}}$

ACC-derived features were extracted, encompassing movement intensity, motion metrics, movement patterns, acceleration magnitude, jerk magnitude, signal magnitude area (SMA), zero-crossing rate (ZCR), energy, decay time, spectral entropy, rotation vector, wavelet-

based descriptors, maximum movement intensity, and overall statistics derived from Heuristic OIT.[11,12] Algorithm 3 performs multimodal feature extraction around change points $\mathcal{C} = \{\tau_i\}_{i=1}^{N}$ using multiple window sizes $\mathcal{W} = \{w_1, w_2, ..., w_K\}$. For each change point τ_i, it creates pre-change ($W^- = [\tau_i - w, \tau_i)$) and post-change ($W^+ = [\tau_i, \tau_i + w]$) windows from 3-axis accelerometer data $\tilde{\boldsymbol{A}} \in \mathbb{R}^{3\times T}$, OIT features $\boldsymbol{F}_{\text{OIT}} \in \mathbb{R}^{9\times T}$, and ECG signals $\boldsymbol{e} \in \mathbb{R}^{T}$. The algorithm applies the feature extraction functions Φ_{ACC}, Φ_{OIT}, and Φ_{ECG} to extract respective features, and then combines these features into comprehensive feature vectors $\mathcal{F}$.

Algorithm 3 ECG & ACC Feature Extraction

Require: $\tilde{\boldsymbol{A}} \in \mathbb{R}^{3\times T}$, $\boldsymbol{F}_{\text{OIT}} \in \mathbb{R}^{9\times T}$, $\boldsymbol{e} \in \mathbb{R}^{T}$ ▷ ACC, OIT, ECG signals
Require: $\mathcal{C} = \{\tau_i\}_{i=1}^{N}$, $\mathcal{W} = \{w_1, w_2, ..., w_K\}$ ▷ Change points, window sizes
Ensure: $\mathcal{F} = \{\boldsymbol{f}_{i,w}^{-}, \boldsymbol{f}_{i,w}^{+}\}_{i,w}$ ▷ Multimodal features
for all $(\tau_i, w) \in \mathcal{C} \times \mathcal{W}$ **do**
  $W^- = [\tau_i - w, \tau_i)$, $W^+ = [\tau_i, \tau_i + w]$
  for all $W \in \{W^-, W^+\}$ **do**
    ACC features: $\boldsymbol{f}_{\text{ACC}} = \Phi_{\text{ACC}}(\tilde{\boldsymbol{A}}|_W)$
    OIT features: $\boldsymbol{f}_{\text{OIT}} = \Phi_{\text{OIT}}(\boldsymbol{F}_{\text{OIT}}|_W)$
    ECG features: $\boldsymbol{f}_{\text{ECG}} = \Phi_{\text{ECG}}(\boldsymbol{e}|_W)$
    $\boldsymbol{f}_W = [\boldsymbol{f}_{\text{ACC}}, \boldsymbol{f}_{\text{OIT}}, \boldsymbol{f}_{\text{ECG}}]$ ▷ Concatenate all features
  end for
end for
return $\mathcal{F}$

2.3. *Model Development and Training*

Recent studies have investigated the application of various artificial intelligence methodologies to ACC and ECG data across multiple tasks. These approaches have demonstrated their utility for anomaly detection and activity classification.[11,12,20–28] In line with these advancements, this study implemented and evaluated two baseline deep learning models to classify patients with POTS from healthy controls using wearable-derived ECG and accelerometer (ACC) features.

The first model was a Transformer-based classifier[29] employing the Attention with Linear Biases (ALiBi) mechanism.[30] This model consisted of six transformer layers, each layer having a hidden dimension of 128, eight attention heads, and a dropout rate of 0.3. Gradient checkpointing was applied for memory optimization.

The second model was a hybrid architecture combining Convolutional Neural Networks (CNN), Bidirectional Long Short-Term Memory (BiLSTM), and self-attention mechanisms (CNN-BiLSTM-Attention).[31,32] The CNN layer contained 64 filters with kernel size 3, followed by two layers of Bidirectional LSTM units, each with a hidden size of 128. A self-attention layer was included to emphasize important temporal segments. The packed sequence method was employed to handle variable sequence lengths.

All models were optimized using Adam with a learning rate of 1×10^{-4}, weight decay of 1×10^{-4}, and coefficients set as $\beta_1 = 0.9$, $\beta_2 = 0.999$. To mitigate issues related to class imbalance, Focal Loss was applied with parameters $\gamma = 2.0$ and $\alpha = 0.25$. Training proceeded

using batches of size 8 for up to 100 epochs.

Data augmentation strategies—including jittering (σ=0.01), scaling (σ=0.1), and permutation (n=4 segments)—were applied to improve model generalization. Additionally, gradient clipping with a threshold of 1.0 and a ReduceLROnPlateau learning rate scheduler were employed to ensure stable model training.

To evaluate model performance, a 10-fold stratified cross-validation was performed. The ratio of POTS patients and control subjects was kept consistent within each fold.

During training in each fold, outliers were identified and handled. Criteria for handling outliers were determined exclusively from each fold's training data and consistently applied to validation data. The standardization process followed the same method.

Performance metrics, including balanced accuracy, sensitivity, specificity, precision, recall, F1-score, and AUC, were recorded at each training epoch. Confusion matrices for the best-performing model in each fold were saved. Results from all folds were summarized using means and standard deviations.

In the holdout evaluation, the model was trained using approximately 70% of the total dataset, which included data from both POTS patients and control subjects. Additionally, around 10% of the dataset was separately allocated for validation purposes. Finally, model performance was assessed using a test dataset comprising roughly 20% of the total data, which was not utilized at any stage during training or validation. All data splits were performed at the session level.

Preprocessing parameters, including standardization and criteria for outlier handling determined during training, were consistently applied to the test dataset.

The trained models classified each sample in the test set as either POTS or control. Additionally, learned feature representations were visualized using dimensionality reduction techniques, enabling comparison between the distributions of test set and those of the training and validation sets.

3. Results

3.1. *Comparative Feature Analysis*

Physiological data collected from wearable sensors were analyzed to investigate potential differences between individuals diagnosed with POTS and healthy control participants. The analyzed features included the previously described ECG and accelerometer-derived metrics, comprising statistical, spectral, morphological, physiological, movement intensity, postural, spatial, and temporal characteristics.

Welch's t-tests revealed that a total of 1169 features showed statistically significant differences between groups after applying the Benjamini-Hochberg false discovery rate correction (FDR, $q < 0.05$). Notably significant features with an FDR-adjusted q-value less than 0.001 included SKNA burst duration, heart rate changes, logarithmic-scale complexity measures, SMA computed both 10 minutes before events and within 10-minute intervals, ECG amplitude maxima over 1-minute intervals, and variance-based principal component characteristics measured around event intervals ($\pm$10 minutes). Detailed statistical results for these significant features—including exact q-values and 95% confidence intervals—are summarized in Table 3.

Boxplots of these significant features listed in Table 3 are presented in Figure 2.

Overall, the analysis demonstrated that physiological features extracted within specific temporal windows around each detected changepoint exhibited statistically significant differences between POTS patients and healthy control participants. The visualization and statistical assessments further suggested that these features could have potential utility in distinguishing between the two groups. These analytical findings provided valuable direction for selecting meaningful features and designing input features for the development of AI-based classification models.

Feature	**FDR q-value**	**95% CI**
ECG max (1 min)	<0.001	[0.253, 0.284]
SMA (10 min)	<0.001	[0.329, 0.361]
EVR PC2 (10 min)	<0.001	[0.371, 0.403]
SKNA burst dur. (5 min)	<0.001	[0.231, 0.262]
SKNA burst dur. (10 min)	<0.001	[0.226, 0.257]
log Hjorth compl. (10 min)	<0.001	[0.198, 0.229]
log Hjorth compl. (5 min)	<0.001	[0.192, 0.224]
log Hjorth compl. (1 min)	<0.001	[0.181, 0.212]
log Hjorth compl. (30 sec)	<0.001	[0.168, 0.199]
HR change (10 min)	<0.001	[0.113, 0.144]
HR change (5 min)	<0.001	[0.098, 0.130]
HR change (1 min)	<0.001	[0.080, 0.112]

Table 3: Detailed Statistical Results for All Features (CP = Changepoint; EVR = Explained Variance Ratio; dur. = duration; compl. = complexity; HR = Heart Rate)

Note: Welch's t-test was used for all comparisons. FDR = False Discovery Rate adjusted q-value; CI = 95% Confidence Interval for effect sizes calculated using Hedges' g. All listed features are significant at FDR-adjusted $q < 0.001$.

3.2. *Classification Performance*

The generalization performance of each model was evaluated using 10-fold stratified cross-validation, and the results are summarized in Table 4.

The Transformer model achieved a balanced accuracy of 0.825 ± 0.116, sensitivity of 0.800 ± 0.135, and precision of 0.940 ± 0.079. This model also demonstrated specificity of 0.850 ± 0.198, an F1-score of 0.841 ± 0.093, and an AUC of 0.813 ± 0.181.

The CNN-BiLSTM-Attention (CNN-BiLSTM) model exhibited higher scores compared to the Transformer across all reported metrics, except for precision. Specifically, the model achieved a balanced accuracy of 0.900 ± 0.084 and sensitivity of 0.950 ± 0.062, indicating a reduced likelihood of false negatives. Furthermore, this model had specificity of 0.850 ± 0.142 and precision of 0.935 ± 0.062. Additionally, the CNN-BiLSTM-Attention model recorded an F1-score of 0.939 ± 0.052 and an AUC of 0.900 ± 0.103.

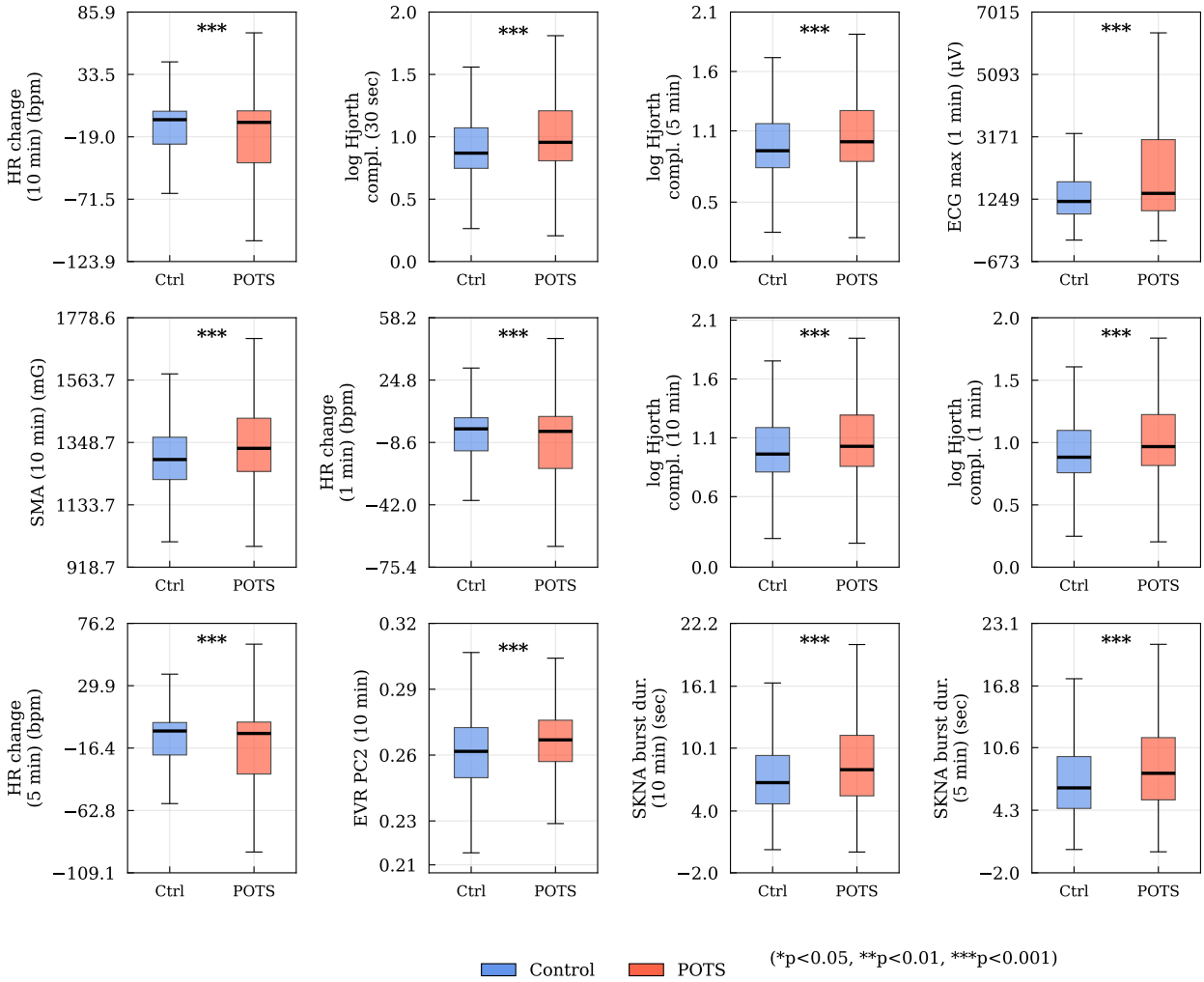

Fig. 2. Comparison of physiological variables between Control and POTS groups

Metric	**Model Architecture**	
	Transformer	CNN-BiLSTM
Balanced Accuracy	0.825 ± 0.116	**0.900 ± 0.084**
Sensitivity	0.800 ± 0.135	**0.950 ± 0.062**
Specificity	0.850 ± 0.198	**0.850 ± 0.142**
Precision	**0.940 ± 0.079**	0.935 ± 0.062
F1 Score	0.841 ± 0.093	**0.939 ± 0.052**
AUC-ROC	0.813 ± 0.181	**0.900 ± 0.103**

Table 4: 10-Fold Cross-validation performance metrics (mean ± 95% confidence interval).

To evaluate the practical performance of the final models trained on the complete dataset, validation was performed using a holdout test set.

The CNN-BiLSTM model exhibited superior overall validation performance compared to

the Transformer model, achieving higher balanced accuracy (86.77% vs. 85.01%), sensitivity (93.55% vs. 90.32%), specificity (96.67% vs. 94.93%), and F1-score (95.08% vs. 93.33%), except for specificity, which was equal (80% vs. 80%). The complete comparison of these validation metrics is provided in Table 5, indicating that the CNN-BiLSTM model may be more effective for classifying this particular dataset than the Transformer model.

Metric	Transformer	CNN-BiLSTM
Balanced Accuracy	85.01%	**86.77%**
Sensitivity	90.32%	**93.55%**
Specificity	80.00%	80.00%
Precision	94.93%	**96.67%**
F1 Score	93.33%	**95.08%**

Table 5: Validation performance metrics

Feature space visualization (Fig. 4) further illustrates the capability of the CNN-BiLSTM-Attention model in distinguishing between POTS patients and healthy controls. Dimensionality reduction techniques, including PCA, t-SNE, and UMAP, applied to the high-dimensional latent representations transformed by this model, revealed separation between the two groups in the two-dimensional feature space, indicating effective discrimination. Additionally, the test set data from POTS patients closely align with the distribution of the existing POTS group, suggesting the model's robustness and potential to generalize effectively to new patient data.

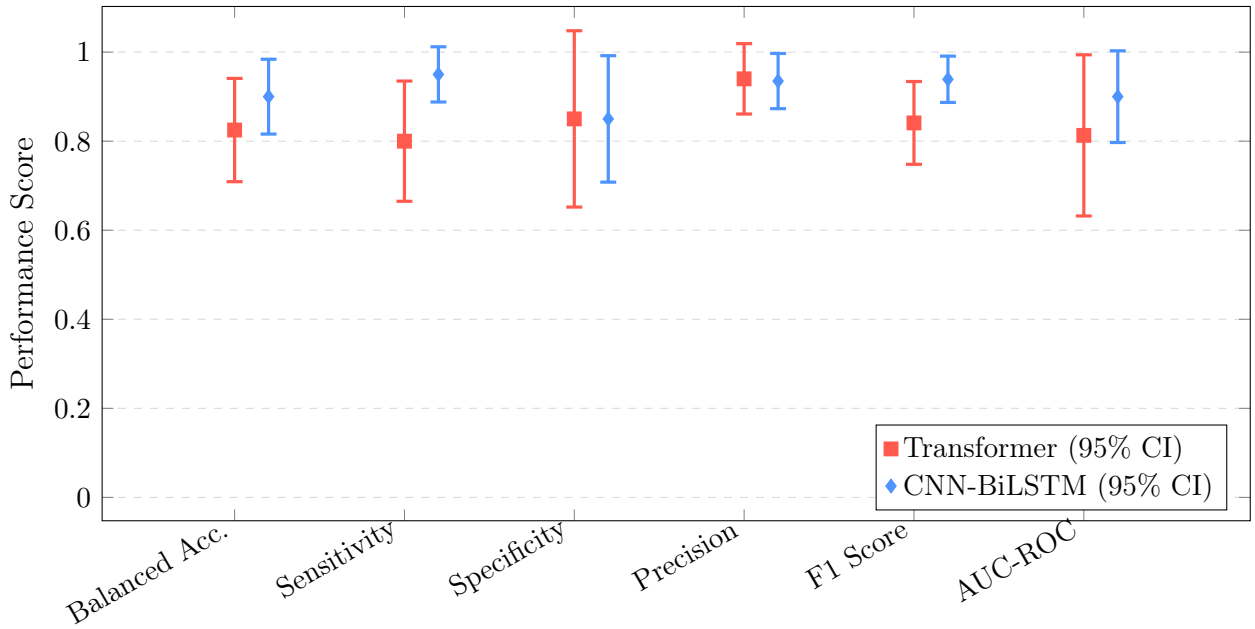

Fig. 3. 10-Fold Cross Validation Results with 95% Confidence Intervals: Model performance comparison with mean values and confidence intervals.

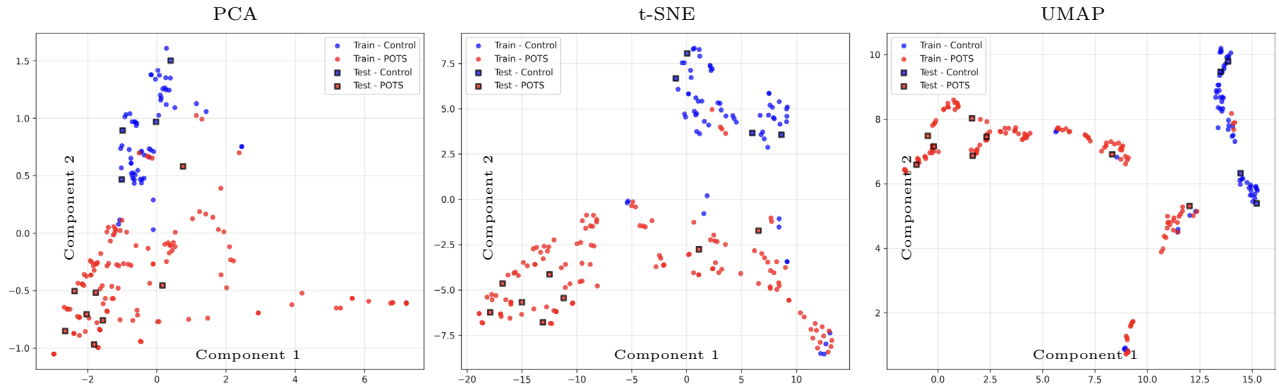

Fig. 4. Feature Space Visualization - CNN-BiLSTM-Attention Model

4. Discussion

The findings of this study suggest potential clinical implications for enhancing the diagnostic approach to POTS. In contrast, the wearable-based methodology employed in this study facilitates continuous and unobtrusive monitoring of subtle yet potentially clinically relevant changes in physiological signals, including ECG and accelerometer-derived features, as they occur naturally in real-world settings. The deep learning models trained on wearable sensor data demonstrated promising performance in differentiating between POTS patients and healthy controls. Nonetheless, to establish the robustness of these findings and evaluate their clinical applicability, it is essential to perform additional validation studies that include expanded patient cohorts, varied clinical settings, and more comprehensive control-group datasets.

Despite the promising results, several considerations must be acknowledged regarding this study. First, the dataset used in this research may not fully represent the broad diversity of clinical conditions and patient populations, suggesting that further validation with larger and more representative samples is required to enhance the generalizability and clinical applicability of the diagnostic models. Second, this study focused primarily on specific physiological signals; therefore, integrating additional physiological parameters or external environmental factors in future research may further improve model generalizability and practical clinical applicability. In addition, the POTS cohort included both patients taking treatment medications and those who were not. Therefore, it remains necessary in future research to clarify explicitly whether the developed classifier is detecting disease-specific signals or instead patterns driven by medication use or other confounding factors.

Additionally, the physiological parameters analyzed in this research—including features derived from ECG and ACC data—may hold diagnostic promise for conditions beyond POTS, indicating potential utility across multiple clinical scenarios. Prior studies have shown ECG and HRV parameters to be valuable in evaluating autonomic dysfunction related to heart failure, cardiac arrhythmias, and diabetes mellitus.[33,34] Likewise, accelerometry-derived metrics have been successfully utilized in assessing motor abnormalities in neurological diseases such as Parkinson's disease and in assessing fall risk among older adults.[35,36] Future research may focus on investigating the potential of these physiological metrics in additional disease contexts, thereby supporting the creation of broadly applicable diagnostic technologies targeting diverse neurological and autonomic conditions.

5. Conclusion

This study investigated the feasibility of using deep learning techniques with features extracted from wearable-derived ECG and ACC data to differentiate patients diagnosed with POTS from healthy controls. The proposed wearable-based approach effectively captured subtle yet clinically significant physiological variations and symptom patterns occurring naturally in daily life. It demonstrates its potential to address key limitations of conventional diagnostic tests such as the tilt-table test. Two deep learning architectures, specifically the Transformer and hybrid CNN-BiLSTM models, demonstrate promising performance that suggests potential applicability in real-world clinical scenarios. The proposed integration of wearable-derived physiological signal analysis and deep learning techniques holds promise for future patient-centered diagnostics and improved clinical management for individuals with POTS.

6. Acknowledgments

This study was supported in part by institutional funds from the Cedars-Sinai Center for AI Research and Education (CAIRE), the American Heart Association grant 23IPA1052289, the United States National Institutes of Health grants R01 HL139829 and OT2OD028190, and the Burns & Allen Chair in Cardiology Research of the Cedars-Sinai Medical Center. In addition, Graciela Gonzalez-Hernandez was funded in part by the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health, under Award Number R01AI164481.

References

1. R. S. Sheldon, B. P. Grubb II, B. Olshansky, W.-K. Shen, H. Calkins, M. Brignole, S. R. Raj, A. D. Krahn, C. A. Morillo, J. M. Stewart et al., 2015 heart rhythm society expert consensus statement on the diagnosis and treatment of postural tachycardia syndrome, inappropriate sinus tachycardia, and vasovagal syncope, Heart rhythm **12**, e41 (2015).
2. D. Dulal, A. Maraey, H. Elsharnoby, P. Chacko and B. Grubb, Impact of covid-19 pandemic on the incidence and prevalence of postural orthostatic tachycardia syndrome, European Heart Journal-Quality of Care and Clinical Outcomes , p. qcae111 (2025).
3. C. K. Ormiston, I. Świątkiewicz and P. R. Taub, Postural orthostatic tachycardia syndrome as a sequela of covid-19, Heart rhythm **19**, 1880 (2022).
4. A. C. Kwan, J. E. Ebinger, J. Wei, C. N. Le, J. R. Oft, R. Zabner, D. Teodorescu, P. G. Botting, J. Navarrette, D. Ouyang et al., Apparent risks of postural orthostatic tachycardia syndrome diagnoses after covid-19 vaccination and sars-cov-2 infection, Nature cardiovascular research **1**, 1187 (2022).
5. D. L. Teodorescu, A. Kote, J. N. Reaso, C. Rosenberg, X. Liu, A. C. Kwan, S. Cheng and P.-S. Chen, Postural orthostatic tachycardia syndrome after covid-19 vaccination, Heart Rhythm **21**, 74 (2024).
6. A. Lee, X. Liu, C. Rosenberg, S. Borle, D. Hwang, L. S. Chen, X. Li, N. B. Merz and P.-S. Chen, Skin sympathetic nerve activity in patients with chronic orthostatic intolerance, Heart Rhythm **19**, 1141 (2022).
7. W. B. Plash, A. Diedrich, I. Biaggioni, E. M. Garland, S. Y. Paranjape, B. K. Black, W. D. Dupont and S. R. Raj, Diagnosing postural tachycardia syndrome: comparison of tilt testing compared with standing haemodynamics, Clinical science **124**, 109 (2013).

8. H. Lee, P. A. Low and H. A. Kim, Patients with orthostatic intolerance: relationship to autonomic function tests results and reproducibility of symptoms on tilt, Scientific reports **7**, p. 5706 (2017).
9. L. Xu, X. Cao, R. Wang, Y. Duan, Y. Yang, J. Hou, J. Wang, B. Chen, X. Xue, B. Zhang et al., Clinical features of patients undergoing the head-up tilt test and its safety and efficacy in diagnosing vasovagal syncope in 4,873 patients, Frontiers in Cardiovascular Medicine **8**, p. 781157 (2022).
10. T. Furukawa, Role of head-up tilt table testing in patients with syncope or transient loss of consciousness, Journal of arrhythmia **33**, 568 (2017).
11. A. Yurtman and B. Barshan, Activity recognition invariant to sensor orientation with wearable motion sensors, Sensors **17**, p. 1838 (2017).
12. A. Yurtman, Activity recognition invariant to position and orientation of wearable motion sensor units, PhD thesis, Bilkent Universitesi (Turkey)2019.
13. X. Liu, C. Rosenberg, J. Ricafrente, M. E. Leier, H. Dinh, T. H. Everett and P.-S. Chen, Using an ambulatory electrocardiogram monitor to record skin sympathetic nerve activity, Heart rhythm **19**, 330 (2022).
14. R. Killick, P. Fearnhead and I. A. Eckley, Optimal detection of changepoints with a linear computational cost, Journal of the American Statistical Association **107**, 1590 (2012).
15. C. Orphanidou, T. Bonnici, P. Charlton, D. Clifton, D. Vallance and L. Tarassenko, Signal-quality indices for the electrocardiogram and photoplethysmogram: Derivation and applications to wireless monitoring, IEEE journal of biomedical and health informatics **19**, 832 (2014).
16. M. Elgendi, Optimal signal quality index for photoplethysmogram signals, Bioengineering **3**, p. 21 (2016).
17. M. Henry and G. Judge, Permutation entropy and information recovery in nonlinear dynamic economic time series, Econometrics **7**, p. 10 (2019).
18. T. Pham, Z. J. Lau, S. A. Chen and D. Makowski, Heart rate variability in psychology: A review of hrv indices and an analysis tutorial, Sensors **21**, p. 3998 (2021).
19. M. G. Frasch, Comprehensive hrv estimation pipeline in python using neurokit2: Application to sleep physiology, MethodsX **9**, p. 101782 (2022).
20. Y. Zheng, Z. Liu, R. Mo, Z. Chen, W.-s. Zheng and R. Wang, Task-oriented self-supervised learning for anomaly detection in electroencephalography, in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2022.
21. A. Jiang, C. Huang, Q. Cao, Y. Xu, Z. Zeng, K. Chen, Y. Zhang and Y. Wang, Anomaly detection in electrocardiograms: Advancing clinical diagnosis through self-supervised learning, arXiv preprint arXiv:2404.04935 (2024).
22. R. Hribar and D. Torkar, Explainable anomaly detection of 12-lead ecg signals using denoising autoencoder, in Intelligent Secure Trustable Things, (Springer, 2024) pp. 127–140.
23. R. Liu, A. A. Ramli, H. Zhang, E. Henricson and X. Liu, An overview of human activity recognition using wearable sensors: Healthcare and artificial intelligence, in International Conference on Internet of Things, 2021.
24. A. R. Javed, M. U. Sarwar, S. Khan, C. Iwendi, M. Mittal and N. Kumar, Analyzing the effectiveness and contribution of each axis of tri-axial accelerometer sensor for accurate activity recognition, Sensors **20**, p. 2216 (2020).
25. J. Park, D.-W. Kim and J. Lee, Calanet: Cheap all-layer aggregation for human activity recognition, Advances in Neural Information Processing Systems **37**, 69419 (2024).
26. E. J. Huang and J.-P. Onnela, Augmented movelet method for activity classification using smartphone gyroscope and accelerometer data, Sensors **20**, p. 3706 (2020).
27. M. Kos, M. Bogdan, N. W. Glynn and J. Harezlak, Classification of human physical activity based on raw accelerometry data via spherical coordinate transformation, Statistics in medicine **39**, 2901 (2020).

28. M. A. Alsheikh, A. Selim, D. Niyato, L. Doyle, S. Lin and H.-P. Tan, Deep activity recognition models with triaxial accelerometers., in AAAI workshop: artificial intelligence applied to assistive technologies and smart environments, 2016.
29. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser and I. Polosukhin, Attention is all you need, Advances in neural information processing systems **30** (2017).
30. O. Press, N. A. Smith and M. Lewis, Train short, test long: Attention with linear biases enables input length extrapolation, arXiv preprint arXiv:2108.12409 (2021).
31. J. P. Chiu and E. Nichols, Named entity recognition with bidirectional lstm-cnns, Transactions of the association for computational linguistics **4**, 357 (2016).
32. X. Yin, Z. Liu, D. Liu and X. Ren, A novel cnn-based bi-lstm parallel model with attention mechanism for human activity recognition with noisy data, Scientific Reports **12**, p. 7878 (2022).
33. U. Acharya, L. Min and P. Joseph, Hrv analysis using correlation dimension and dfa innovations technol, Biol. Med.(ITBM-RBM) **23**, 333 (2002).
34. F. Shaffer and J. P. Ginsberg, An overview of heart rate variability metrics and norms, Frontiers in public health **5**, p. 258 (2017).
35. S. Patel, K. Lorincz, R. Hughes, N. Huggins, J. Growdon, D. Standaert, M. Akay, J. Dy, M. Welsh and P. Bonato, Monitoring motor fluctuations in patients with parkinson's disease using wearable sensors, IEEE transactions on information technology in biomedicine **13**, 864 (2009).
36. A. Weiss, T. Herman, N. Giladi and J. M. Hausdorff, Objective assessment of fall risk in parkinson's disease using a body-fixed sensor worn for 3 days, PloS one **9**, p. e96675 (2014).