PACE-Former: Bridging Clinical Safety and Diagnostic Precision in Multi-Center CPET via Systemic Style Adaptation

Cong Wang[1], Bei Xu[2], and Shou-ling Mi*[1]

[1]Zhongshan Hospital, Fudan University, Shanghai, China
[2]BexiMed Co., Ltd., Shanghai, China

Abstract

Background: Cardiopulmonary exercise testing (CPET) is the gold standard for assessing cardiorespiratory fitness. However, its clinical deployment faces three critical challenges: physical heterogeneity across multi-center devices, the non-stationary nature of physiological signals, and safety risks during maximal testing in high-risk patients. Existing AI models focus primarily on offline retrospective analysis, failing to address the urgent clinical need for real-time safety monitoring and prognostic assessment.
Methods: We propose **PACE-Former**, a unified framework utilizing a three-fold decoupling paradigm. (1) **Feature Decoupling:** An input-driven Style Encoder extracts "systemic fingerprints" from resting-phase data, using Conditional Layer Normalization to dynamically calibrate the network against device-specific bias. (2) **Spatiotemporal Decoupling:** A hybrid masking training strategy enables a single model to perform both "online causal inference" (for low false-alarm rates) and "offline global review" (for high precision). (3) **Task Decoupling:** A dual-head architecture jointly outputs Anaerobic Threshold (AT) probability for diagnosis and scalar VO_{2peak} prediction for prognosis, enabling "Virtual Maximal Testing."
Results: Validated on a multi-center cohort using 10-second aggregated data, the model achieved expert-level diagnostic precision in offline mode (Hit Rate within $\pm$20s $> 90\%$). In online mode, it maintained an Early Trigger Rate $< 2\%$ while accurately predicting final VO_{2peak} with $< 5\%$ error at 75% test completion.
Conclusion: PACE-Former successfully bridges the gap between clinical safety and diagnostic precision, offering a robust, generalized solution for intelligent CPET interpretation.

Keywords: Cardiopulmonary Exercise Testing, Anaerobic Threshold, Time Series Forecasting, Domain Generalization, Virtual Maximal Testing

1 Introduction

1.1 Background and Motivation

Cardiopulmonary exercise testing (CPET) provides a holistic assessment of the cardiovascular, respiratory, and muscular systems. The Anaerobic Threshold (AT) and Peak Oxygen Uptake (VO_{2peak}) derived from CPET are critical biomarkers for risk stratification in heart failure, perioperative assessment, and rehabilitation prescription [1, 2].

However, the widespread clinical adoption of CPET AI faces distinct challenges compared to other medical domains. While breakthroughs in medical AI have focused on **Anatomical Structural Recognition** (e.g., lung nodule detection in CT), CPET analysis represents a higher-order challenge of **Physiological Dynamics Inference**. This task involves inherent epistemic uncertainty: AT is a metabolic phase transition occurring within muscle cells, invisible to direct observation. Models must solve a complex inverse problem to infer this moment from noisy, lagged gas exchange signals collected at the mouth [3]. Furthermore, the "ground truth" for AT relies on expert interpretation of multi-dimensional curves, suffering from inherent inter-observer variability ($\approx \pm$30s).

1.2 The Data Challenge: Heterophasic Coupling

Unlike standardized DICOM images, CPET data are multivariate, non-stationary time series characterized by complex dynamics:

- **Heterophasic Coupling:** AT determination relies on the decoupling of linear relationships between $\dot{V}O_2$, $\dot{V}CO_2$, and $\dot{V}_E$. However, due to differences in chemoreceptor sensitivity and gas transport rates, these variables exhibit natural phase lags (e.g., ventilatory compensation $\dot{V}_E$ lags behind metabolic acidosis). Models must align these asynchronous cues.
- **Non-stationary Evolution:** From rest to exhaustion, the statistical distribution (mean, variance) of physiological signals drifts drastically. Models cannot rely on spatial invariance (as in CNNs for images) but must capture transient phase-change features within a dynamically evolving manifold.

1.3 The Clinical Dilemma: Safety vs. Precision

Current single-task models fail to address the contradictory dual needs of clinical workflows:

- **Online Monitoring (Safety First):** For high-risk patients, clinicians need a "Virtual Maximal Test"—predicting VO_{2peak} early (e.g., at 75% load) to terminate the test safely. This requires an extremely low **Early Trigger Rate**; false alarms causing premature termination are unacceptable.

- **Offline Reporting (Precision First):** Retrospective diagnosis requires unbiased temporal localization (Bias ≈ 0) to match expert consensus.

1.4 The Deployment Bottleneck: Systemic Heterogeneity

A major barrier to multi-center deployment is the **Holistic Systemic Fingerprint**. Differences in hardware (Cortex vs. Cosmed), environmental physics (barometric pressure), and protocols (mask dead space) create severe systemic time biases (> 40s) across centers, hindering generalization.

To address these challenges, we introduce **PACE-Former**, a Conformer-based framework that utilizes systemic style adaptation and hybrid task learning to unify real-time safety and offline precision.

2 Methodology

2.1 Data Infrastructure: The CPETx Standard & Universal Adapter

A major barrier to multi-center CPET analysis is the "Data Silo" effect caused by proprietary file formats and inconsistent variable nomenclature. To address this, we established a unified data infrastructure comprising two core components: the **CPETx Standard Schema** and the **Universal Device Adapter**.

2.1.1 The CPETx Standard

We defined a strict schema (see Appendix A) that standardizes three dimensions of CPET data:

1. **Time-Series Data:** Harmonizes breath-by-breath or second-by-second physiological signals (e.g., mapping `VO2/Kg` and `VO2_Kg` to the standard `VO2_kg`).

2. **Metadata:** Unifies subject demographics and environmental calibration parameters (Temperature, Barometric Pressure) crucial for BTPS correction.

3. **Summary Metrics:** Standardizes report-level scalar indices (e.g., VE/VCO_2 Slope, VO_{2peak}) for prognostic benchmarking.

2.1.2 The Universal Adapter Pipeline

We implemented a software middleware following the *Adapter Design Pattern*. As shown in Figure **??** (concept), the pipeline operates in three stages:

1. **Ingestion:** Vendor-specific drivers parse raw files from different manufacturers (Cosmed, Cortex, Vyaire, Schiller).

2. **Harmonization:** Variables are mapped to the `standard_name` defined in our Schema. Units are automatically converted (e.g., $kph \to m/s$, $cal \to J$) to ensure physical consistency.

3. **Validation:** A rigid type-check ensures data integrity (e.g., ensuring $RER > 0$, $SpO_2 \in [0, 100]$) before feeding into the AI model.

This infrastructure decouples the downstream **PACE-Former** model from hardware specifics, enabling true device-agnostic deployment.

2.2 Model Architecture: The PACE-Former

The proposed architecture (Fig. **??**) integrates three key components:

2.2.1 Style-Aware Backbone

We introduce a lightweight 1D-CNN **Style Encoder** that processes the Preload stream. It extracts statistical moments (mean, variance, texture) representing the device and patient baseline. These style embeddings are injected into the main network via **Conditional Layer Normalization (CLN)**. Let x be the feature input. CLN adapts the normalized features using affine parameters (γ, β) generated from the style embedding s:

$$\text{CLN}(x, s) = \frac{x - \mu}{\sigma} \cdot \gamma(s) + \beta(s) \tag{1}$$

This allows the network to perform "adaptive normalization," effectively removing site-specific bias before physiological feature extraction.

The backbone utilizes **Conformer Blocks** [4], combining Convolutional modules (to capture local morphological trends like V-slope deflection) and Self-Attention (to capture long-range heterophasic coupling between metabolic production and ventilatory response).

2.2.2 Dual-Task Heads

- **Time Head (Diagnostic):** Outputs a sequence of probabilities indicating if AT has occurred. We use **Soft-Argmax** during inference to regress sub-bin continuous time indices, overcoming the quantization error of 10s binning.

- **Value Head (Prognostic):** A scalar regression head that predicts the final VO_{2peak} at every time step. This enables the assessment of the "Virtual Maximal Test" capability.

2.3 Hybrid Training Strategy

To unify online and offline capabilities in a single model, we employ **Dynamic Mask Sampling**:

- **Online Mode ($p = 0.5$):** A Causal Mask (upper triangular) is applied to the Self-Attention mechanism. The model can only attend to historical data, optimizing for low latency and monotonic probability progression.
- **Offline Mode ($p = 0.5$):** The mask is removed. The model utilizes bidirectional attention, leveraging recovery phase features (e.g., rapid drop in HR) to refine AT localization.

2.4 Loss Function

The total loss combines classification, regression, and regularization terms:

$$\mathcal{L} = \mathcal{L}_{BCE} + \lambda_1 \mathcal{L}_{Mono} + \lambda_2 \mathcal{L}_{Time} + \lambda_3 \mathcal{L}_{VO2} \quad (2)$$

Where $\mathcal{L}_{Mono}$ enforces monotonic non-decreasing probabilities for the AT event, and $\mathcal{L}_{VO2}$ uses time-weighted MSE (heavier weights near the end) to encourage early convergence of prognostic predictions.

3 Experimental Design

3.1 Dataset and Study Population

This study utilized a multi-center retrospective cohort collected from the Cardiopulmonary Exercise Testing laboratories of three institutions, including Zhongshan Hospital (Fudan University), between January 2023 and December 2024. The study protocol was approved by the Institutional Review Board (IRB No. XXX-202X), and the requirement for informed consent was waived due to the retrospective nature of the analysis.

Raw data were initially screened based on the following inclusion criteria: (1) Standard ramp protocol performed on a cycle ergometer; (2) Complete breath-by-breath gas exchange data recorded; (3) Test termination manifested by symptom limitation. Tests with severe signal artifacts (signal loss $>$ 10% of duration) or total duration $<$ 3 minutes were excluded.

A total of $N = 1,240$ valid CPET sessions were included. To strictly evaluate cross-center generalization, we adopted a Leave-One-Center-Out (LOCO) strategy: data from Center A (Main tertiary hospital, $n = 800$) and Center B ($n = 200$) served as the training and validation domains, while Center C (Community health center, $n = 240$) was held out strictly for external testing.

The cohort covers a wide range of functional capacities, from heart failure patients to healthy volunteers. Detailed demographics and physiological characteristics are summarized in Table 1. The ground truth for Anaerobic Threshold (AT) was determined via a two-round blind review by three senior physiologists.

Table 1: Demographics and Baseline Characteristics of the Study Cohort

Characteristic	**Training Set** (n=1,000)	**Test Set** (n=240)	**P-value**
Demographics			
Age (years)	54.2 ± 12.5	56.1 ± 10.8	0.12
Male Sex, n (%)	620 (62%)	135 (56%)	0.09
BMI (kg/m^2)	24.5 ± 3.2	23.9 ± 2.8	0.04
CPET Metrics			
VO_{2peak} (mL/kg/min)	18.4 ± 5.6	17.8 ± 4.9	0.21
Test Duration (min)	9.2 ± 2.1	8.9 ± 1.8	0.15
Device Manufacturer	Cosmed	Cortex	-

Values are mean ± SD or n (%). P-values via t-test or χ^2 test.

3.2 Dual-Mode Evaluation Metrics

We established distinct metric sets for the two deployment scenarios:

Table 2: Dual-Mode Evaluation Metrics

Mode	**Key Metric**	**Target**
Online	**Early Trigger Rate**	$<$ **2**% (Safety)
	Mean Trigger Delay	$<$ 30s
	VO_2 MAPE @ 75%	$<$ 5% (Prognosis)
	VO_2 Stability	Low Variance
Offline	**Hit Rate @ 20s**	$>$ **90**% (Precision)
	Time Bias	$\approx$ 0s
	Bland-Altman LoA	Clinical limits

4 Results

4.1 Offline Precision: Clinical Equivalence

In the offline diagnostic setting, **PACE-Former** demonstrated high agreement with expert consensus. The cumulative hit-rate curve (Fig. **??**) shows that 92% of predictions fell within a ±20s tolerance (2 bins) of the ground truth. Bland-Altman analysis revealed a mean bias of 0.8s, with limits of agreement narrower than reported inter-observer variability (±30s).

4.2 Online Safety: The Zero-False-Alarm Standard

For real-time monitoring, safety is paramount. Fig. **??** illustrates the distribution of trigger delays. Crucially, the "Early Trigger" region (negative delay) is virtually empty (<1.5%), ensuring the model does not prematurely terminate tests. The mean trigger delay was 18s, which

is physiologically acceptable given the persistence logic required to filter noise.

4.3 Prognostic Value: Virtual Maximal Testing

The VO_{2peak} convergence plot (Fig. **??**) demonstrates the model's prognostic capability. The Mean Absolute Percentage Error (MAPE) drops below 5% once 75% of the test duration is completed. This suggests that for high-risk patients, a sub-maximal test (stopping at $\sim$75% effort) combined with **PACE-Former** can reliably estimate functional capacity without inducing maximal cardiac stress.

4.4 Ablation Study: The Role of Style Adaptation

Removing the Style Encoder resulted in a significant increase in systemic bias (+14s on Center B), confirming that the input-driven adaptation effectively decouples device heterogeneity from physiological features.

5 Discussion

This study presents the first CPET analysis framework to explicitly decouple and optimize for the conflicting requirements of safety and precision. By leveraging meso-scale aggregation (10s bins) and Conditional Layer Normalization, **PACE-Former** overcomes the noise and heterogeneity inherent in multi-center respiratory data.

The physiological significance of the **Conformer backbone** is evident in its ability to handle heterophasic coupling; the attention mechanism naturally aligns the lagged ventilatory response with metabolic events. Furthermore, the **Dual-Head design** validates the feasibility of "Virtual Maximal Testing," potentially transforming CPET protocols for heart failure and perioperative populations.

6 Conclusion

PACE-Former establishes a new technical standard for automated CPET interpretation. It provides a clinically safe, diagnostically precise, and universally applicable tool that requires minimal calibration, paving the way for large-scale, standardized cardiopulmonary phenotyping.

References

[1] Guazzi M, et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice. *Eur Heart J.* 2016;37:2315-2381.

[2] Wasserman K, et al. *Principles of Exercise Testing and Interpretation.* 5th edn. Lippincott Williams & Wilkins; 2012.

[3] Beaver WL, et al. A new method for detecting anaerobic threshold by gas exchange. *J Appl Physiol.* 1986;60:2020-2027.

[4] Gulati A, et al. Conformer: Convolution-augmented Transformer for Speech Recognition. *Interspeech.* 2020.

A The Unified CPET Data Standard

The following tables define the `CPETx_Standard` schema used to harmonize multi-center data. All downstream AI models consume data strictly adhering to this interface.

A.1 Time-Series Schema (Breath-by-Breath / Continuous)

Table 3: Time-Series Variable Definitions

Standard Name	Unit	Type	Description
Temporal Dynamics			
`Time`	mm:ss	string	Elapsed time from test start.
`Phase_Time`	mm:ss	string	Time elapsed within current phase.
`Time_Relative`	s	float	Relative time pointer (0.0 to T).
Gas Exchange & Metabolism			
`VO2`	mL/min	float	Absolute Oxygen Consumption.
`VO2_kg`	mL/kg/min	float	Relative Oxygen Consumption.
`VCO2`	mL/min	float	Carbon Dioxide Production.
`RER`	ratio	float	Respiratory Exchange Ratio (VCO_2/VO_2).
`METS`	MET	float	Metabolic Equivalents.
Ventilation			
`VE`	L/min	float	Minute Ventilation (BTPS).
`Bf`	1/min	float	Breath Frequency.
`VT`	L	float	Tidal Volume.
`PetO2`	mmHg	float	End-tidal PO_2.
`PetCO2`	mmHg	float	End-tidal PCO_2.
`VE_VO2`	ratio	float	Ventilatory Equivalent for O_2.
`VE_VCO2`	ratio	float	Ventilatory Equivalent for CO_2.
Cardiovascular & ECG			
`HR`	bpm	int	Heart Rate.
`VO2_HR`	mL/beat	float	Oxygen Pulse.
`SpO2`	%	float	Oxygen Saturation.
`BP_Syst`	mmHg	int	Systolic Blood Pressure.
`BP_Diast`	mmHg	int	Diastolic Blood Pressure.
`ST_[Lead]`	mV	float	ST-segment deviation (Leads I-V6).
Ergometer & Protocol			
`Power_Load`	W	float	External Work Rate.
`RPM`	r/min	int	Pedaling Cadence.
`Load_Phase`	cat	int	0:Rest, 1:Warmup, 2:Exercise, 3:Recovery.

A.2 Summary Metrics Schema (Report Level)

A.3 Metadata & Calibration Schema

Table 4: Key Prognostic Summary Metrics

Standard Name	Unit	Description
Aerobic Capacity		
Peak_VO2_kg	mL/kg/min	The highest 20-30s average VO_2 achieved.
Peak_VO2_Predicted	mL/min	Predicted value based on Wasserman/Hansen equations.
Peak_METS	MET	Peak functional capacity.
Anaerobic Threshold (AT)		
VO2_at_AT	mL/min	VO_2 at the moment of Anaerobic Threshold.
HR_at_AT	bpm	Heart Rate at AT.
Time_at_AT	mm:ss	Timestamp of the AT event.
Ventilatory Efficiency		
VE_VCO2_Slope	ratio	Linear regression slope of VE vs VCO_2.
OUES	-	Oxygen Uptake Efficiency Slope.
Peak_PetCO2	mmHg	Maximum end-tidal CO_2 pressure.

Table 5: Subject and Environmental Metadata

Category	Fields	Note
Subject	Subject_ID, Age, Gender, Height_cm, Weight_kg	Essential for predicting normative values.
Exam	Exam_ID, Date, Protocol_Name, Ergometer_Type	Defines the physical context of the test.
Env	Pressure_Barometric, Temp_Ambient, RH_Ambient	Used for STPD/BTPS gas correction.