COACT: Collaborative Objective AI-Assisted Clinical Team Assessment in Emergency Medicine

Ioan-Sorin Comșa[1], Ivan Moser[1], Tanja Birrenbach[2], Thomas C. Sauter[2], and Per Bergamin[1]
[1]Institute for Research in Open-, Distance- and eLearning,
Swiss Distance University of Applied Sciences, CH-3900 Brig, Switzerland,
[2]Department of Emergency Medicine, Inselspital, Bern University Hospital,
University of Bern, CH-3010 Bern, Switzerland.
E-mails: {ioan-sorin.comsa, ivan.moser, per.bergamin}@ffhs.ch, {tanja.birrenbach, thomas.sauter}@insel.ch

***Abstract*—Effective teamwork is vital in emergency medicine, yet training interprofessional teams is challenging due to resource constraints, time limitations, and inconsistent assessment methods. Traditional performance evaluation relies on subjective ratings, often biased, and lacking objective validation. To address this, we introduce COACT (Collaborative Objective AI-Assisted Clinical Team Assessment), a data-driven framework for the evaluation of team dynamics in emergency medical team training. COACT enables debriefing by ~~leveraging~~ using wearable devices that capture physiological and motion data during training. The system processes raw signals in real time across multiple timescales to detect team dynamics. A Machine Learning (ML) model then analyzes the processed data, identifying synchronization patterns ~~linked~~ related to team coordination and effectiveness. We introduce the COACT index, a novel metric ~~quantifying how closely a team 's~~ that quantifies how closely team patterns align with high-performing teams. The system operates on a continual ML paradigm, enabling self-monitoring, self-training, and self-adaptation as new data is collected. ~~Results show a~~ The results show up to 90% accuracy between the COACT index and subjective team performance ratings, validating the potential of the proposed framework as an objective ~~,~~ AI-driven team assessment tool for emergency medicine training.**

***Index Terms*—Team Performance Assessment, Wearable Signals, Pervasive Computing, Machine Learning, Pattern Analysis.**

I. Introduction

The first minutes after an accident are critical for patient survival, requiring emergency teams to act with precision~~and speed [1].~~ , speed, and coordination [1]. Training is essential but resource-intensive and often lacks realism due to the unpredictability of emergencies [2]. Moreover, assessing team performance remains a challenge, as traditional methods rely ~~heavily~~ on subjective ratings [3]. Although recent advances in wearable technology offer promising alternatives (i.e., audio and eye-tracking sensors), they come with limitations such as noise sensitivity, privacy concerns, and intrusiveness [4]. In contrast, physiological and motion sensors provide a non-intrusive and scalable approach for monitoring stress, cognitive load, and coordination in real time [5]. Devices such as Empatica's EmbracePlus capture multiple biomarkers, offering a rich foundation for objective performance analysis [6]. Yet, a ~~significant gap persists: the absence~~ gap remains: the lack of a structured framework to ~~transform this~~ translate multi-modal ~~data~~ signals into interpretable team performance measures.

This paper introduces COACT (Collaborative Objective AI-Assisted Clinical Team Assessment), a real-time framework for objectively evaluating team performance in emergency medicine training using data from Empatica's EmbracePlus. COACT includes: *a)* a real-time pipeline to extract, process, analyze, and interpret physiological and motion data; *b)* a continual Machine Learning (ML) model that identifies evolving patterns and ensures optimal computations; and *c)* a novel COACT Index (CI) that quantifies team synchronicity and performance. COACT provides a scalable, AI-driven solution for enhancing medical training through wearable-based, data-driven team performance assessment.

II. COACT Framework

Figure 1 illustrates the proposed COACT framework and its ~~interaction entities. Participants~~interactions. The participants, in pairs, join the training scenario, each equipped with an EmbracePlus smartwatch. During training, the COACT framework collects and processes raw data. Afterwards, participants engage in a debriefing session, and team performance data is fed back into COACT, enabling it to adapt and refine patterns using continual ML.

A. Training Scenario

~~In the~~ In interprofessional team training, advanced medical and nursing students are paired to complete a VR simulation moderated by a trainer, allowing them to practice teamwork in emergency scenarios [7]. Unlike traditional simulations with actors, which can be subjective, the VR environment offers objective experiences, providing reliable data for the ML. The VR application features three main avatars: Trainee 1 (nursing student), Trainee 2 (medical student), and a virtual patient experiencing headache followed by an epileptic seizure ~~provoked~~ caused by a cerebral hemorrhage. During the scenario, Trainee 1 begins the consultation by taking the patient's history, with the patient's responses controlled by the moderator/trainer. After five minutes, Trainee 2 joins the session and Trainee 1 provides a structured handover of the relevant clinical information. The trainees proceed to manage the case collaboratively. Physiological and motion data from both trainees are captured from EmbracePlus devices, and synchronized with the timestamps from the VR application. Raw data is continuously transmitted ~~via~~ through Bluetooth to

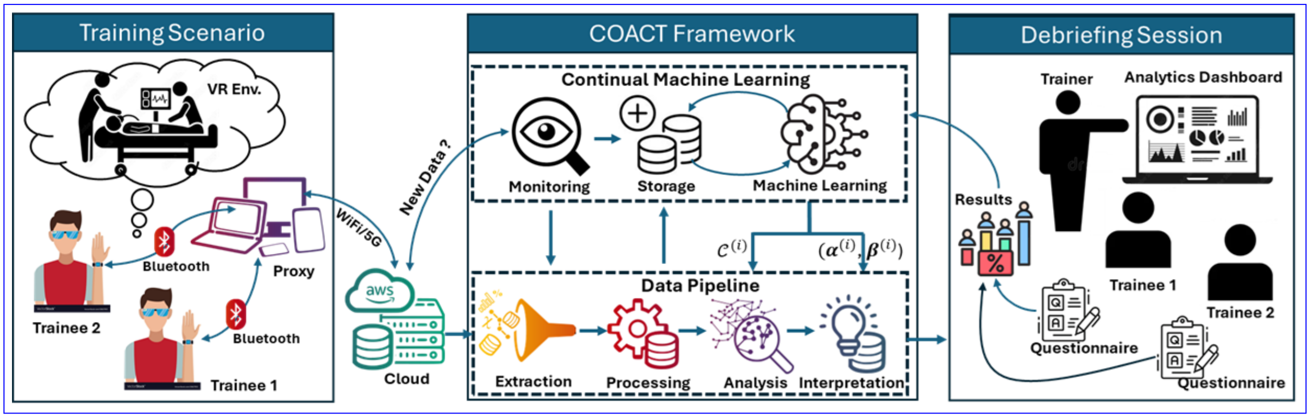

Fig. 1: Proposed COACT System

nearby proxy devices and stored in the cloud via the WiFi/5G interface.

B. Proposed COACT System

The framework ~~from~~ in Fig. 1 processes physiological and motion data from pairs of participants ~~into interpretable metrics for assessing team performance via the CI scores .~~ into interpretable CI scores to assess team performance. Concurrently, continual ML refines these calculations using ~~accumulated data~~ data accumulated from participants over time. The monitoring component detects new pairs, notifying both the data pipeline and continual ML, which periodically retrains based on new data and results from debriefing sessions.

1) Data Pipeline: When the monitoring component detects new raw data from a pair, the pipeline extracts, processes, analyzes, and interprets this data sequentially.

1.a) Extraction: retrieves raw data from cloud-based .avro files to edge storage, converting them into .csv format, including: ~~a~~*i*) motion signals: accelerometer (**A**), steps (**S**); ~~b~~*ii*) physiological signals: electrodermal activity (**E**), pulse rate (**P**), skin temperature (**T**); ~~c~~*iii*) scenario-specific timestamps; and ~~d~~*iv*) Empatica biomarkers processed at 60-second intervals.

1.b) Processing: filters raw data based on scenario-specific timestamps; computes biomarkers at intervals $i \in \{5, 10, 15, \dots\}$s; merges these biomarkers at selected intervals i; and validates processing ~~via~~ by correlation analysis with Empatica biomarkers at $i = 60$s. If we denote by $\mathbf{x}^{(i)} = [\mathbf{a}^{(i)}, \mathbf{s}^{(i)}, \mathbf{e}^{(i)}, \mathbf{p}^{(i)}, \mathbf{t}^{(i)}]$ the processed sample of the biomarker vector derived from raw signals $\mathbf{A}, \mathbf{S}, \mathbf{E}, \mathbf{P}, \mathbf{T}$ ~~at~~ within interval i, the processed dataset is defined as $\mathcal{X}^{(i)} = \{\mathbf{x}_n^{(i)}\}_{n=1}^{N_i}$, with N_i samples per interval i. For pairs $(u_1, u_2) \in \mathcal{U}$, representing ~~the~~ nursing and medical ~~trainee~~ trainees respectively, the datasets become ~~$\mathcal{X}_{u_1}^{(i)} = \{\mathbf{x}_{u1,n}^{(i)}\}_{n=1}^{N_i}$ and $\mathcal{X}_{u_2}^{(i)} = \{\mathbf{x}_{u2,n}^{(i)}\}_{n=1}^{N_i}$, which are synchronized based on~~ $\mathcal{X}_{u_1}^{(i)} = \{\mathbf{x}_{u_1,n}^{(i)}\}_{n=1}^{N_i}$ and $\mathcal{X}_{u_2}^{(i)} = \{\mathbf{x}_{u_2,n}^{(i)}\}_{n=1}^{N_i}$, synchronized by the sample index n.

1.c) Analysis: classifies processed datasets based on centers computed by ~~the~~ continual ML. Let ~~$\mathcal{C}^{(i)} = \{\mathbf{c}_k^{(i)}\}_{k=1}^{K}$ represent these continually~~ $\mathcal{C}^{(i)} = \{\mathbf{c}_k^{(i)}\}_{k=1}^{K_i}$ represent these updated centers with ~~K~~ K_i as the number of centers per interval i. The processed data ~~undergo~~ undergoes the transformation:

$$(\mathcal{X}_{u_1}^{(i)}, \mathcal{X}_{u_2}^{(i)}) \to \mathcal{C}^{(i)} \to (\mathcal{Y}_{u_1}^{(i)}, \mathcal{Y}_{u_2}^{(i)}), \tag{1}$$

where $\mathcal{Y}_u^{(i)} = \{\mathbf{y}_{u,n}^{(i)}\}_{n=1}^{N_i}$ is the transformed dataset for participant u, and each vector $\mathbf{y}_{u,n}^{(i)}$ corresponds to the ~~nearest~~ closest center $\mathbf{c}_k^{(i)}$ (in Euclidean distance) to $\mathbf{x}_{u,n}^{(i)}$.

1.d) Interpretation: computes and visualizes interpretable patterns from analyzed datasets and calculates the COACT index. To simplify interpretation, each biomarker in $\mathbf{y}_{u,n}^{(i)} = [y_{\mathbf{a}}, y_{\mathbf{s}}, y_{\mathbf{e}}, y_{\mathbf{p}}, y_{\mathbf{t}}]$ is discretized using biomarker-specific thresholds. Let $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_m]_{m=1}^{M}$ denote these threshold vectors for accelerometer ($m = 1$), steps ($m = 2$), electrodermal activity ($m = 3$), pulse rate ($m = 4$), and skin temperature ($m = 5$), where each biomarker m has thresholds $\boldsymbol{\alpha}_m = [\alpha_{m,1}, \dots, \alpha_{m,D}]$, and D varies by biomarker. Thus, prior to visualization, the analyzed datasets for each pair $(u_1, u_2) \in \mathcal{U}$ at interval i undergo the transformation:

$$(\mathcal{Y}_{u_1}^{(i)}, \mathcal{Y}_{u_2}^{(i)}) \to \boldsymbol{\alpha}^{(i)} \to (\mathcal{Z}_{u_1}^{(i)}, \mathcal{Z}_{u_2}^{(i)}), \tag{2}$$

where $\mathcal{Z}_u^{(i)} = \{\mathbf{z}_{u,n}^{(i)}\}_{n=1}^{N_i}$ are the discrete datasets used for visualization at each point n of the interval i.

1.e) CI Calculation: Etalon pairs, identified via the TEAM questionnaire [8] during debriefings as teams with optimal cooperation, communication, and coordination, serve as references in the CI calculation. The number of etalon pairs may increase as new data ~~are~~ is ingested by the COACT system. The CI quantifies similarity in motion and physiological patterns between new pairs and established etalon(s). For each etalon pair $(e_1, e_2) \in \mathcal{U}$, similarities between corresponding trainees (u_1 vs. e_1, and u_2 vs. e_2) at interval i and sample n are represented by vector $\mathbf{v}_{u,e,n}^{(i)} = [v_{u,e,n,m}]_{m=1}^{M}$, computed as:

$$v_{u,e,n,m} = \begin{cases} 1, & z_{u,n,m} = z_{e,n,m}, \\ 0, & z_{u,n,m} \neq z_{e,n,m}. \end{cases} \tag{3}$$

The COACT index at interval i is then calculated as:

$$\mathbf{ci}^{(i)} = \cancel{\frac{1}{N \cdot M}} \frac{1}{N_i \cdot M} \sum_{n=1}^{\cancel{N}\, N_i} \sum_{m=1}^{M} \beta_m^{(i)} \cdot v_{u_1,e_1,n,m}^{(i)} \cdot v_{u_2,e_2,n,m}^{(i)}, \tag{4}$$

where $\boldsymbol{\beta}^{(i)} = [\beta_m^{(i)}]_{m=1}^{M}$ denotes biomarker-specific weights. ~~For~~ When multiple etalon pairs ~~, the COACT index is computed with each pair, retaining~~ are available, the CI is computed for each combination, and the maximum value

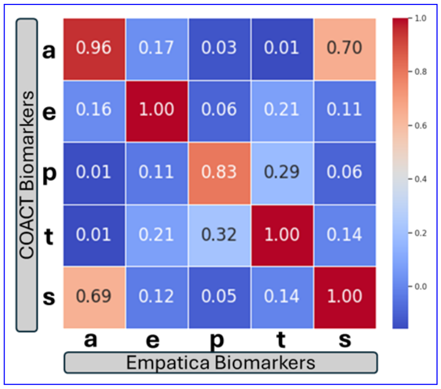

Fig. 2: Pearson Correlation of COACT and Empatica Biomarkers Processed at Interval ~~$i = 60s$~~$i = 60$s

~~obtained~~is retained.

2) Continual ML: For each computation interval i, the data pipeline parameters ($\mathcal{C}^{(i)}$, $\boldsymbol{\alpha}^{(i)}$, and $\boldsymbol{\beta}^{(i)}$) are updated via continual ML, using all accumulated data from the COACT framework during predefined periods. Data centers are recalculated using the unsupervised clustering algorithm (SWAP heuristics), and the optimal number of centers ~~K~~ K_i is determined offline ~~via~~ using the Silhouette index [9]. ~~Threshold and weight vectors are intelligently recalculated, by employing reinforcement learning to identify parameter combinations that yield the highest alignment between~~ To align the computed CI scores and subjective TEAM performance ~~evaluations~~ratings, the threshold and weight vectors are recalculated after each center update using methods like grid search, Bayesian optimization, or reinforcement learning. The Python software implementation of the COACT framework, along with detailed documentation, is ~~publicly~~ available online [9].

C. Debriefing Session

After each training session, a debriefing helps trainees reflect on challenges, technology, and overall experience. Trainers use recorded VR videos to identify communication and procedural gaps, while team biomarker patterns are visualized on an analytics dashboard and compared to ~~top-performing teams (etalons~~the top performing teams (etalon) to calculate ~~the~~ CI scores. ~~At~~ In the end, the trainees consent to ~~data usage~~ the use of data and complete questionnaires assessing acceptance (usability, discomfort, presence, workload, technology acceptance), effectiveness, and feasibility. ~~Additionally, they provide~~ In addition, they complete the TEAM questionnaire [7] to self-assess perceived team performance, with the results integrated into the COACT framework to refine the etalon pairs and ~~improve CI calculationaccuracy.~~ enhance the precision of CI calculation.

III. Experimental Results

A. Methods and Participants

The experimental study involved data from 11 pairs of medical students (ages 24–30) at a Central European University Hospital. Of these, 5 pairs were ~~mixed-gender~~mixed gender, 4 ~~male-only~~male only, and 2 ~~female-only~~female only. Each anonymized participant was assigned a pair ID, trial number, device hand, and role. These identifiers enabled real-time data pairing and processing through the pipeline. Empatica devices continuously recorded raw data from the moment the first student entered until the end of the handover. Each session lasted around 20 minutes, with timestamps aligned to start/stop VR triggers.

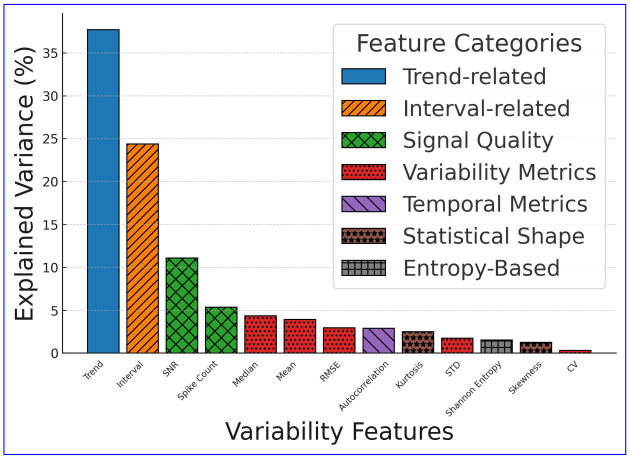

Fig. 3: Explained Variance vs. Variability Features

B. Validation of Data Processing

The data pipeline operates at different intervals $i \in \{10, 20, \ldots, 60\}$s to evaluate the performance of the proposed index $\mathbf{ci}^{(i)}$, bench-marked against TEAM performance ratings. Since Empatica does not disclose its data processing methods nor support intervals other than $i = 60$s, reverse engineering ~~was~~ is applied to derive equivalent processing functions for each sensor and validate them against ~~Empatica's output at 60s~~ the Empatica output at 60s [9]. Accelerometer data ($\mathbf{a}$) is processed by computing the magnitude of movement, followed by the standard deviation within each interval. Step count ($\mathbf{s}$) is aggregated, and temperature ($\mathbf{t}$) is averaged per interval. Electrodermal activity ($\mathbf{e}$) is filtered using a 4th-order low-pass Butterworth filter, with signal amplitudes extracted. Pulse rate ($\mathbf{p}$) is derived from the raw BVP signal ~~through a two-step process:~~ in two steps: ~~*a*~~*i*) ~~*smoothing via Butterworth and moving average filters, and*~~ Butterworth filtering followed by moving average smoothing, and ~~*b*~~*ii*) ~~aligning the~~ alignment of detected diastolic peaks with those provided by the sensor. The obtained tachogram values are then averaged within each interval. Figure 2 presents the Pearson ~~correlation between~~ correlations between the biomarkers processed by COACT and those provided by Empatica at ~~$i = 60$s using~~ $i = 60$s, computed on concatenated data from all participants. Strong diagonal correlations confirm that the COACT module is a viable alternative for analysis at finer time resolutions.

C. Statistical Analysis

~~To assess statistical differences across datasets computed at various time intervals, we extracted multiple variability metrics~~ To assess the statistical difference between datasets at different intervals, multiple metrics are computed for each biomarker. These include: ~~(1)~~ *i)* the computation interval itself~~, (2)~~ ; *ii)* a trend-related metric, measured as the slope of a linear fit to indicate signal direction~~, and (3)~~ ; *iii)* signal quality metrics, such as Signal-to-Noise Ratio (SNR),

TABLE I: COACT Performance Evaluation

Interval [s] i	Samples ~~N~~ N_i	Centers ~~K~~ K_i	acc ~~α_1~~ $\alpha_{1,1}$	steps ~~α_2~~ $\alpha_{2,1}$	eda ~~α_3~~ $\alpha_{3,1}$	pulse ~~α_4~~ $\alpha_{4,1}$	temp ~~α_5~~ $\alpha_{5,1}$	acc ~~β_1~~ $\beta_{1,1}$	steps ~~β_2~~ $\beta_{2,1}$	eda ~~β_3~~ $\beta_{3,1}$	pulse ~~β_4~~ $\beta_{4,1}$	temp ~~β_5~~ $\beta_{5,1}$	Accuracy [%]
10	1443	708	~~34~~ 12	~~89~~ 72	~~29~~ 21	~~53~~ 88	~~23~~ 75	~~0.48~~ 0.15	~~0.06~~ 0.09	~~0.21~~ 0.24	~~0.23~~ 0.52	~~0.07~~ 0.0	90
20	728	368	~~34~~ 11	~~66~~ 72	~~88~~ 25	~~69~~ 85	79	~~0.61~~ 0.18	~~0.14~~ 0.11	~~0.16~~ 0.22	~~0.00~~ 0.44	~~0.09~~ 0.05	~~90~~ 80
30	489	202	~~23~~ 15	~~32~~ 65	~~66~~ 14	~~31~~ 55	~~78~~ 67	~~0.81~~ 0.21	~~0.06~~ 0.12	~~0.07~~ 0.12	~~0.08~~ 0.49	~~0.00~~ 0.06	~~90~~ 80
40	373	121	~~59~~ 51	~~23~~ 52	~~10~~ 32	~~30~~ 38	~~26~~ 72	~~0.53~~ 0.33	~~0.28~~ 0.17	~~0.14~~ 0.02	~~0.03~~ 0.35	~~0.02~~ 0.13	~~90~~ 80
50	298	97	~~33~~ 69	~~27~~ 41	~~26~~ 84	~~42~~ 35	~~72~~ 39	~~0.38~~ 0.44	~~0.18~~ 0.31	~~0.08~~ 0.02	~~0.26~~ 0.15	~~0.09~~ 0.08	~~90~~ 70
60	251	97	~~63~~ 81	~~19~~ 16	~~36~~ 85	~~44~~ 28	~~84~~ 60	~~0.01~~ 0.54	~~0.23~~ 0.26	~~0.17~~ 0.04	~~0.57~~ 0.15	~~0.02~~ 0.01	~~90~~ 70

computed as mean squared over variance, and spike count, which captures abrupt signal changes (defined as values exceeding mean plus one standard deviation)~~. Variability~~ ; *iv)* variability metrics include the mean, median, and Root Mean Square Error (RMSE), with the latter derived from a regression model ~~predicting~~ that predicts the computation interval~~. Temporal~~ ; *v)* temporal dynamics are captured via autocorrelation ~~, while~~ function; *vi)* statistical shape metrics comprise standard deviation, kurtosis, skewness, and coefficient of variation~~. Finally,~~ ; and *vii)* Shannon entropy quantifies signal complexity. These metrics serve as independent variables in a multivariate Ridge regression ~~model~~ used to predict biomarker values. We apply leave-one-out variance analysis to compute the explained variance for each feature. As shown in Fig. 3, ~~the~~ trend and interval metrics contribute the most to differences between time intervals, followed by signal quality and variability metrics. To ensure consistent COACT index performance, the threshold and weight vectors should be calibrated accordingly.

D. Performance Analysis

To evaluate the alignment between the COACT index and subjective TEAM ~~performance ratings, one pair~~ ratings, the pair with the highest rating is designated as ~~an etalon, while~~ the etalon, and the remaining 10 ~~participant~~ pairs are used for classification. Both the TEAM performance ratings and the COACT index scores $\mathbf{ci}^{(i)}$, computed at intervals ~~$i \in \{10, 20, 30, 40, 50, 60\} s$~~ $i \in \{10, 20, 30, 40, 50, 60\}$s, are normalized and binarized into the "Low" (0) and "High" (1) classes using a 0.5 threshold. Table I ~~presents the resulting classification accuracy across all intervals, showing a consistent performance of~~ summarizes the classification performance across different interval durations. The accuracy is highest in short windows (90% ~~, despite a decline~~ at 10s) and gradually decreases with longer intervals, reaching 70% at 50–60s. This trend reflects the reduction in sample size and number of data centers $\mathcal{C}^{(i)}$ as ~~interval duration increases. To maintain this stability, the threshold vectors $\boldsymbol{\alpha}^{(i)}$~~ the interval length increases, highlighting the trade-off between temporal resolution and classification robustness. The threshold and weight vectors $(\boldsymbol{\alpha}^{(i)}, \boldsymbol{\beta}^{(i)})$ are optimized via ~~continual ML using reinforcement learning. Here, $[\boldsymbol{\alpha}^{(i)}, \boldsymbol{\beta}^{(i)}]$ represent the state space. A neural network is trained~~ grid search over 1000 random configurations, retaining the setting with maximum accuracy for each interval i~~to make step-wise decisions that maximize classification accuracy as a reward signal~~. For simplicity and computational efficiency, we use a single threshold ~~per biomarker dimension~~$(D = 1)$ per biomarker, although multiple thresholds could improve precision at the cost of added complexity. ~~The interval-specific optimal states reveal several trends: *a)* Accelerometer importance increases at shorter intervals; *b)* Step count becomes more relevant at longer intervals; *c)* EDA and pulse rate show mid-interval dips in importance with peaks at the extremes; *d)* Skin temperature consistently contributes little across all intervals. Threshold analysis for converting centers $\mathcal{C}^{(i)}$ to interpretable patterns $\mathcal{Z}^{(i)}$ reveals: *a)* Accelerometer~~

Threshold analysis $(\boldsymbol{\alpha}_m)$ shows that accelerometer thresholds increase with ~~longer intervals; *b)* Step thresholds decreasewith longer intervals; *c)* EDA thresholds peak mid-range and drop at extremes; *d)* Pulse thresholds increase with shorter intervals .~~

~~Based on these findings, we conclude the following: *a)* accelerometer is most informative at short intervals, capturing immediate physical responses, while step count gains relevance at longer intervals due to movement accumulation; *b)* pulse and EDA show higher importance at extreme intervals, likely reflecting acute stress (short) and sustained arousal (long), but contribute less at mid-range intervals. Temperature adds minimal value in all intervals due to its slow variability . The learned thresholds mirror these patterns , adjusting sensitivity based on each signal's temporal dynamics~~interval length, step thresholds decrease, EDA thresholds rise sharply beyond 40s, and pulse thresholds drop as intervals lengthen. The biomarker weights $(\boldsymbol{\beta}_m)$ reveal that accelerometer and steps gain importance at longer intervals, while EDA and pulse are more influential at shorter ones; skin temperature remains negligible, with a slight peak at $i = 40$s. These dynamics indicate that accelerometer and steps capture stable, large-scale coordination when aggregated over longer windows, whereas EDA and pulse primarily reflect transient stress and rapid physiological responses that fade with time. The rise in accelerometer and EDA thresholds reflects increased variability requiring stronger signals to establish synchronicity, while the decline in step and pulse thresholds shows compressed variability where even small differences suffice. These patterns highlight a temporal differentiation: longer intervals emphasize sustained coordination but yield lower accuracy because averaging suppresses transient stress responses and micro-coordination shifts. By contrast, short-interval dynamics retain these fine-grained fluctuations, offering richer discriminative features and better alignment with perceived team performance.

E. Implications and Future Directions

The COACT index demonstrates potential as an objective proxy for team performance when properly calibrated. This opens two key research directions for improving emergency medicine training: first, enabling real-time detection of performance weaknesses to adapt task difficulty or personalize learning content; second, enhancing feedback quality by integrating additional sensor data (e.g., kinematics, eye tracking) for deeper insight into motor and cognitive processes. These advancements can support more precise, actionable feedback and predictive interventions. Implementing such systems will require interdisciplinary collaboration and further research on scalability and long-term impact.

IV. Conclusion

This study introduces COACT (Collaborative Objective AI-Assisted Clinical Team Assessment), a novel framework for emergency medicine training. Using Empatica EmbracePlus devices, COACT captures real-time motion and physiological data from participant dyads. A continual ML pipeline processes this data, with clustering algorithms reducing complexity and ~~reinforcement learning~~ grid search optimizing model parameters to assess team performance. The resulting COACT index offers an interpretable, objective alternative to traditional subjective performance evaluations. COACT enables adaptive training and enhances feedback quality through specific performance metrics and actionable insights. ~~With a~~ The results show that processing signals at shorter timescales yields higher concordance with subjective ratings (up to 90~~% concordance between COACT index values and subjective participant ratings , the system shows~~ %), underscoring the importance of fine-grained temporal resolution in capturing team performance. This demonstrates COACT's strong potential for real-time, data-driven assessment and lays the groundwork for further research into objective clinical team evaluation.

Acknowledgment

This research was conducted as part of the INTEAM+ BeLearn project “Enhancing Inter-Professional Team Trainings with a Personalised and Adaptive Learning System”. The authors acknowledge Andrea Neher and Rafael Wespi from the ~~InselSpital~~ Inselspital Bern for their contributions to data collection.

References

[1] M. Cassarino, K. Robinson, R. Quinn, B. Naddy, A. O’Regan, D. Ryan, F. Boland, M. E. Ward, R. McNamara, M. O’Connor, G. McCarthy, and R. Galvin, “Impact of Early Assessment and Intervention by Teams Involving Health and Social Care Professionals in the Emergency Department: A Systematic Review,” *PLOS ONE*, vol. 14, pp. 1–13, 2019.

[2] Y. Okuda, E. O. Bryson, S. DeMaria Jr, L. Jacobson, J. Quinones, B. Shen, and A. I. Levine, “The Utility of Simulation in Medical Education: What Is the Evidence?” *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine*, vol. 76, no. 4, pp. 330–343, 2009.

[3] R. Wespi, T. Birrenbach, S. K. Schauber, T. Manser, T. C. Sauter, and J. E. Kämmer, “Exploring Objective Measures for Assessing Team Performance in Healthcare: An Interview Study,” *Frontiers in Psychology*, vol. 14, 2023.

[4] V. Aukstakalnis, Z. Dambrauskas, K. Stasaitis, L. Darginavicius, P. Dobozinskas, N. Jasinskas, and D. Vaitkaitis, “What Happens in the Shock Room Stays in the Shock Room? A Time-based Audio/Video Audit Framework for Trauma Team Performance Analysis,” *European Journal of Emergency Medicine*, vol. 27, no. 2, pp. 121–124, 2020.

[5] K. Mundnich, B. M. Booth, M. L’Hommedieu, T. Feng, B. Girault, J. L’Hommedieu, M. Wildman, S. Skaaden, A. Nadarajan, J. L. Villatte, T. H. Falk, K. Lerman, E. Ferrara, and S. Narayanan, “TILES-2018, a Longitudinal Physiologic and Behavioral Data Set of Hospital Workers,” *Scientific Data*, vol. 7, no. 354, October 2020.

[6] Empatica Inc., “EmbracePlus: The World’s Most Advanced Smartwatch for Continuous Health Monitoring,” https://www.empatica.com/en-gb/embraceplus, Accessed: 2025-09-12.

[7] R. Wespi, L. Schwendimann, A. Neher, T. Birrenbach, S. K. Schauber, T. Manser, T. C. Sauter, and J. E. Kämmer, “TEAMs go VR — Validating the TEAM in a Virtual Reality (VR) Medical Team Training,” *Advances in Simulation*, vol. 9, no. 38, pp. 1–11, 2024.

[8] S. Cooper, R. Cant, J. Porter, K. Sellick, G. Somers, L. Kinsman, and D. Nestel, “Rating Medical Emergency Teamwork Performance: Development of the Team Emergency Assessment Measure (TEAM),” *Resuscitation*, vol. 81, no. 4, pp. 446–452, 2010.

[9] I.-S. Comșa, I. Moser, and P. Bergamin, “COACT: Collaborative Objective AI-Assisted Clinical Team Assessment Framework,” https://github.com/ioansorincomsa/coact_framework, 2025, Computer Software. Accessed: 2025-09-12.