Proceedings of the Human Factors and Ergonomics Society Annual Meeting
2025, Vol. 69(1) 333–337

DOI: 10.1177/10711813251357937
journals.sagepub.com/home/pro
Sage

Physiological Synchrony as a Predictor of Team Performance: A Machine Learning Approach

Mengyao Li[1], Emanuel Rojas[1], Bethany Bracken[2], Nancy Cooke[3], Phillip Desrochers[2], Jamie Gorman[3], Lixiao Huang[3], Molly Kilcullen[4], Michael Rosen[4], Aaron Winder[2], Arya Yadav[3], Xiaoyun Yin[3], Elmira Zahmat Doost[3], and Shiwen Zhou[3]

Abstract

As human-AI collaboration becomes more common, robust metrics for evaluating team performance are essential. This study examines whether physiological synchrony can predict team effectiveness using machine learning (ML) models. Dyads completed a mission-planning task while multimodal physiological data (e.g., HRV, respiration, fNIRS) and communication were collected. We extracted both individual and interactional-level features (e.g., synchrony, coherence) and trained ML models to classify team performance. Logistic Regression and SVM models achieved up to 96% accuracy when including interactional synchrony features, outperforming models using individual data alone. Key predictors included breathing synchrony and oxygenated hemoglobin coherence. These findings demonstrate the added value of modeling physiological coupling to understand emergent team dynamics. Future work will expand the sample size and incorporate team-level recurrence and entropy metrics to better capture collective coordination. This approach offers a pathway toward real-time performance monitoring and adaptive interventions in high-stakes collaborative environments.

Keywords

machine learning, physiological synchrony, team performance

Introduction

Teams play a critical role in diverse domains, including military decision-making, aviation crew coordination, and medical care. As artificial intelligence (AI) advances in autonomy, it is increasingly positioned as a potential teammate, highlighting the importance of effective human-autonomy teaming. To analyze and enhance team performance, robust measurement methods are essential. Machine learning (ML) offers a promising approach by leveraging bio-physiological sensor data to predict team performance. Despite the recognition of team cognition as an emergent process (Cooke et al., 2013), existing measurement methods rely heavily on individual-level assessments, overlooking the interactional and team-level dynamics that shape performance. This paper identifies three critical levels of data for feature engineering in ML models: individual-level aggregation, interaction-level features (e.g., similarity, synchrony), and team-level patterns (e.g., recurrence, entropy). In this preliminary study, we focus on individual, and interaction-level features, leaving team-level patterns for future work. We develop and compare ML models incorporating individual and interactional-level features, demonstrating their complementary value in predicting team performance and advancing ML-based approaches for human and hybrid human-AI teams.

Background

Team performance is inherently interdependent, shaped not only by individual capabilities but also by interactional and emergent team-level processes. Traditional ML-based approaches to team performance modeling primarily focus on individual physiological and behavioral data, such as EEG, heart rate variability (HRV), and eye tracking. While these measures are useful for assessing cognitive workload

[1]Georgia Institute of Technology, Atlanta, USA
[2]Charles River Analytics, Cambridge, MA, USA
[3]Arizona State University, Tempe, USA
[4]Johns Hopkins University, Baltimore, MD, USA

Corresponding Author:
Mengyao Li, Georgia Institute of Technology, 654 Cherry Street NW, Atlanta, GA 30332, USA.
Email: mengyao.li@gatech.edu

and stress at the individual level, they fail to capture the coordinated physiological responses that defines effective teamwork (Brannick & Prince, 1997). A critical limitation in existing approaches is the lack of interdependent data modeling, which fails to account for how team members synchronize in response to shared tasks and stressors (Brannick & Prince, 1997). This limitation arises because team cognition is not merely an aggregation of individual states but emerges from interaction patterns (Cooke et al., 2013). To address this, we explore multi-level feature modeling. Research has demonstrated that synchrony in physiological signals, such as respiration and cardiac activity, is associated with improved communication, shared situational awareness, and higher collective performance (Gorman et al., 2012; Liu et al., 2021). To bridge this gap, it is essential to consider data at the individual, interactional, and team levels when applying ML techniques to understand and measure team performance. At the individual level, team activation is inferred by aggregating individual neuro-physiological and behavioral data (Kivikangas et al., 2014). At the interaction level, features such as similarity, mimicry, and synchrony serve as strong predictors of social dynamics and performance (Gonzales et al., 2010; Valdesolo & DeSteno, 2011). For instance, Paromita et al. (2023) used conversational turn similarity to predict micro-behaviors, while Kusmakar et al. (2020) developed a model of player interaction networks to predict scoring opportunities. Liu et al. (2021) introduced a multi-brain network model to link neural synchrony with team collaboration. At the team level, dynamic patterns over time, such as recurrence rate and entropy, have been analyzed using methods like recurrence analysis (Knight et al., 2021). For example, Gorman et al. (2012) modeled team communication dynamics, while Strang et al. (2014) used physio-behavioral data to assess coupling during collective tasks. By systematically evaluating the contribution of these multi-level features to ML-based team performance prediction, we aim to demonstrate the added value of physiological synchrony metrics and provide empirical support for ITC and emergent team cognition theories.

Method

We adapted the Noncombatant Evacuation Operation (NEO) task, a team-based collaboration exercise originally developed by the U.S. Navy, for two participants. In this task, dyads collaboratively planned a rescue mission under constrained resources, including weapons, personnel, and time. Each participant had unique mission-critical knowledge: one handled weapons and personnel, while the other focused on environmental and intelligence data. Verbal communication was necessary for coordination. To examine team reorganization, we introduced two perturbations: (1) an assembly task requiring participants to build a Lego set and (2) a sudden reduction in the mission deadline. Teams developed an action plan covering routing, extraction strategies, injury management, and safe return logistics, which was subsequently evaluated on a 100-point performance scale. Throughout the task, multimodal physiological and behavioral data were collected, including fNIRS, EEG, ECG, PPG, respiration, alongside communication logs. Preliminary data from two teams (of a planned sample of 35 team samples) are presented here. Full analysis of 35 teams will be presented at the conference.

Physiological data were first preprocessed using filtering and normalization techniques to reduce noise and ensure comparability across subjects. Given the preliminary nature of this study and the small dataset, we focused on individual (e.g., HRV time-domain and frequency-domain metrics; $n=28$) and interaction-level features (e.g., breathing synchrony, RSP coupling, and PPG phase coherence; $n=15$), excluding team-level dynamics such as recurrence and entropy, which require larger samples to capture robust patterns. Features were derived using a sliding window approach to balance temporal resolution with signal stability: long windows (5 min) were used for HRV to capture meaningful variability in autonomic nervous system activity, as HRV metrics require several minutes of continuous data to be statistically reliable. In contrast, short windows (5 s) were applied to other signals to detect finer-grained interactional dynamics. Features with zero variance, excessive missing values, or high correlation ($r>.7$) were removed to reduce redundancy, reducing the set from 43 to 22. See Appendix Table A1 for a complete list of selected features.

Team performance was scored on a 100-point rubric by trained raters based on mission completeness, coordination efficiency, and decision rationale. We used a median split (mean score = 67.5) to label teams as high (above mean) or low (below mean) performers. Four machine learning models—K-Nearest Neighbors (KNN), Multi-Layer Perceptron (MLP), Logistic Regression, and Support Vector Machine (SVM)—were trained and evaluated using two feature sets: (1) Individual-level features and (2) Combined individual and interactional features. Due to the small sample size and temporal nature of physiological data, we employed a stratified temporal 80 to 20 train-test split. We first identify all temporal indices for each class separately, creating class-specific ordered sequences. Each class undergoes temporal splitting using the same test ratio, ensuring balanced representation across partitions. In the full dataset, we plan to implement leave-one-team-out cross-validation (LOTO-CV) to robustly assess model performance across diverse team compositions and task responses. Additionally, grid search was used for hyperparameter tuning, and regularization techniques were applied. Model performance was assessed using accuracy, precision, recall, and F1-score. For interpretability, Logistic Regression's feature coefficients were analyzed to identify the most influential predictors of team performance. Interactional features, such as team breathing synchrony and O2HB sync index, emerged as key contributors.

Table 1. Classification Results Across Individual and Combined Feature Sets.

Level of Feature	Accuracy	Precision	Recall	F1
Individual				
KNN	0.79	0.82	0.80	0.79
MLP	0.79	0.81	0.79	0.79
Logistic	0.86	0.88	0.87	0.86
SVM	0.87	0.89	0.87	0.86
Individual + Interactional				
KNN	0.66	0.67	0.66	0.65
MLP	0.74	0.74	0.74	0.74
Logistic	0.96	0.96	0.96	0.96
SVM	0.95	0.96	0.95	0.95

Outcome

As shown in Table 1, models trained on individual-level physiological features demonstrated strong performance, with Logistic Regression (Accuracy = 0.86, F1-score = 0.86) and SVM (Accuracy = 0.87, F1-score = 0.86) outperforming KNN and MLP, both of which achieved approximately 79% accuracy and F1-scores. Incorporating interactional features (e.g., team breathing synchrony, physiological coupling, and cross-correlations) led to substantial performance gains in Logistic Regression and SVM, achieving 96% and 95% accuracy, respectively, a 10-percentage point improvement over models using individual features alone. Notably, Logistic Regression demonstrated the most pronounced improvement, reaching near-perfect classification performance across all metrics (Precision, Recall, F1-score = 0.96). Conversely, KNN performance deteriorated (Accuracy = 0.66, F1-score = 0.65), suggesting sensitivity to the increased feature dimensionality. Logistic Regression's feature importance analysis identified interactional synchrony metrics as the strongest predictors of team performance. O2HB synchrony and RSP phase coherence emerged as the top important features, followed by team breathing synchrony and coupling stability. These findings indicate that team-level physiological synchronization serves as a more robust predictor of team performance than individual physiological states alone (Figure 1).

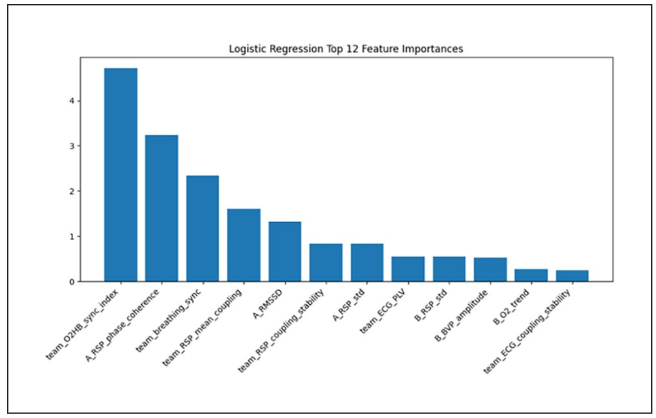

Figure 1. Logistic regression results for top 12 features. Interactional-level features are top important features.

Conclusion

Our findings suggest that although individual physiological features alone can effectively predict team performance, incorporating interactional-level synchrony metrics enhances predictive power, particularly for Logistic Regression and SVM models. The strong performance of interactional features aligns with prior research on physiological coupling in team settings, indicating that synchrony measures capture meaningful team dynamics that contribute to performance outcomes. While the current models demonstrate high accuracy, the limited dataset poses a significant risk of overfitting. Future work will validate these findings using a larger sample ($n = 35$) and explore model robustness across different team compositions and tasks. Additionally, we will integrate team-level features, such as recurrence and entropy measures, to better capture emergent team dynamics and refine performance predictions. These higher-order patterns may provide deeper insights into the collective coordination mechanisms that drive team effectiveness. Overall, this study provides empirical evidence supporting the integration of physiological interactional features for team performance prediction, with implications for real-time monitoring and adaptive interventions in collaborative settings.

Appendix

Table A1. Selected Features and Definition.

Feature	Sensor	Level	Definition
A RSP Std	RSP	Individual	Participant A's standard deviation of the respiratory rate, indicating breathing variability.
A RSP phase coherence			Participant A's stability of breathing rhythm measured through Hilbert phase coherence.
A_BVP Amplitude	PPG		Participant A's peak-to-trough amplitude of blood volume pulse signal.
B BVP Amplitude			Participant B's peak-to-trough amplitude of blood volume pulse signal.
A_O2_trend	fNIR		Participant A's mean difference between oxygenated (O2HB) and deoxygenated (HHB) hemoglobin.
B_O2_trend			Participant Bs mean difference between oxygenated (O2HB) and deoxygenated (HHB) hemoglobin.
A_RMSSD	ECG (Time-domain)		Participant A's Root Mean Square of Successive Differences between normal heartbeats—measures heart rate variability in time domain
A_pNN50			Participant A's proportion of successive NN intervals larger than 50 ms.
B_pNN50			Participant B's proportion of successive NN intervals larger than 50 ms.
A_SDNN			Participant A's standard deviation of normal beat (NN) intervals. Encompasses both short-term high frequency and long-term low frequency.
B_HF	ECG (Frequency-domain)		Participant B's high Frequency power (0.15–0.4 Hz) of heart rate variability. Reflects sympathetic activity.
Team_breathing_sync	RSP	Interaction	Maximum correlation between two team members' breathing patterns.
Team_RSP_mean_coupling			Average dynamic coupling strength between breathing patterns, which captures how strongly one person's breathing influences the other.
Team_RSP_coupling_stability			Standard deviation of coupling strength, which captures how consistent the respiratory coupling is overtime.
Team_PPG_max_xcorr	PPG		Highest correlation values between two PPG signals within the time windows.
Team_PPG_xcorr_lag			Time lag at maximum cross-correlation—indicates leader-follower dynamics
Team_O2HB_sync_index	fNIR		Correlation between oxygenated hemoglobin (O2HB) levels in team members' brain, which captures potential cognitive overload or stress.
Team_RR_correlation	ECG		Correlation between team members' RR (beat-to-beat) intervals
Team_RR_sync_index			Proportion of time partners' normalized RR intervals share the same sign
Team_ECG_PLV			Phase Locking Value between breathing signals—measures breathing synchronization.
Team_ECG_mean_coupling			Mean correlation of respiratory signals across time windows.
Team_ECG_ coupling stability			Standard deviation of correlation strength
Team_ECG_sync			Maximum cross-correlation between partners' breathing signals

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is funded by DARPA OP TEMPO FP00041063.

ORCID iD

Emanuel Rojas https://orcid.org/0009-0003-9979-0013

References

Brannick, M. T., & Prince, C. (1997). An overview of team performance measurement. In M. T. Brannick, E. Salas, & C. W. Prince (Eds.), *Team performance assessment and measurement* (pp. 15–28). Psychology Press.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. *Cognitive Science*, *37*(2), 255–285.

Gonzales, A. L., Hancock, J. T., & Pennebaker, J. W. (2010). Language style matching as a predictor of social dynamics in small groups. *Communication Research*, *37*(1), 3–19.

Gorman, J. C., Cooke, N. J., Amazeen, P. G., & Fouse, S. (2012). Measuring patterns in team interaction sequences using a discrete recurrence approach. *Human Factors*, *54*(4), 503–517.

Kivikangas, J. M., Kätsyri, J., Järvelä, S., & Ravaja, N. (2014). Gender differences in emotional responses to cooperative and competitive game play. *PLOS ONE*, *9*(7), Article e100318.

Knight, S., Lavan, N., Torre, I., & McGettigan, C. (2021). The influence of perceived vocal traits on trusting behaviours in an economic game. *Quarterly Journal of Experimental Psychology*, *74*(10), 1747–1754.

Kusmakar, S., Shelyag, S., Zhu, Y., Dwyer, D., Gastin, P., & Angelova, M. (2020). Machine learning enabled team performance analysis in the dynamical environment of soccer. *IEEE Access*, *8*, 90266–90279.

Liu, T., Duan, L., Dai, R., Pelowski, M., & Zhu, C. (2021). Teamwork, team-brain: Exploring synchrony and team interdependence in a nine-person drumming task via multiparticipant hyperscanning and inter-brain network topology with fNIRS. *NeuroImage*, *237*, Article 118147.

Paromita, P., Khader, A., Begerowski, S., Bell, S. T., & Chaspari, T. (2023). Linguistic and vocal markers of microbehaviors between team members during analog space exploration missions. *IEEE Pervasive Computing*, *22*(2), 7–18.

Strang, A. J., Funke, G. J., Russell, S. M., Dukes, A. W., & Middendorf, M. S. (2014). Physio-behavioral coupling in a cooperative team task: Contributors and relations. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 145–158.

Valdesolo, P., & DeSteno, D. (2011). Synchrony and the social tuning of compassion. *Emotion*, *11*(2), 262–266.