Check for updates

Special Issue: Proceedings of the 70th HFES International Annual Meeting

Proceedings of the Human Factors and Ergonomics Society Annual Meeting
1–4
© The Author(s) 2026

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/10711813261499640
journals.sagepub.com/home/pro


Beyond Dyads: Machine Learning for Team Performance Prediction in Multi-Member Teams

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

Abstract

Team operations, in which multiple members collaborate to accomplish tasks together, play a pivotal role in many safety-critical domains, such as military training, naval operations, and medical care. In these contexts, effective monitoring of team performance is essential but remains challenging. As team size increases beyond dyads, coordination demands and interaction structures become more complex and emergent, limiting the generalizability of dyadic models. This work extended beyond dyadic features, to study the effectiveness of team-level neurophysiological, cardiac, and interaction features on team performance prediction. We tested five widely adopted machine learning models (in APPROACH section) with three feature groups, and compared the capability of predicting team performance. Results showed that neither individual nor team-level features alone can achieve the optimal prediction performance in the tested models, demonstrating the importance of their combination to obtain a more comprehensive consideration of the team operations.

Keywords

team performance, multi-modal, pyschophysiological data

Introduction

Team operations, in which multiple members collaborate to accomplish tasks together, play a pivotal role in many safety-critical domains. In these contexts, effective monitoring of team performance is essential but remains challenging due to traditional approaches are poorly suited for real-time assessment. Neurophysiological sensing offers a complementary alternative by enabling real-time, unobtrusive monitoring of individual and team states, supporting continuous assessment and prediction of team performance. Recent work has leveraged machine learning to model team performance using psychophysiological and behavioral data from both individual and team-level (Bisbey & Salas, 2019; Strang et al., 2011). However, most of the existing works at team-level study only dyadic teams (Li et al., 2025; Zahmat Doost et al., 2025). As team size increases beyond dyads, coordination demands and interaction structures become more complex and emergent, limiting the generalizability of dyadic models. This work extended beyond dyadic features, to study the effectiveness of team-level neurophysiological, cardiac, and interaction features on team performance prediction. We tested five widely adopted machine learning models (in APPROACH section) with three feature groups, naming Individual, Team, and Combined, and compared the capability of predicting team performance. Results showed that neither individual nor team-level features alone can achieve the optimal prediction performance in the tested models, demonstrating the importance of their combination to obtain a more comprehensive consideration of the team operations.

Background

Machine learning approaches to predicting team performance employ diverse feature architectures spanning individual and team-level metrics (Atweh et al., 2022).

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

Corresponding Author:
Mengtao Lyu, Georgia Institute of Technology, Atlanta, GA 30332, USA.
Email: mengtao.lyu@gatech.edu

Individual-level features typically include physiological measures, reflecting individual's cognitive status. These physiological measures are also adopted to understand the overall status at the team's level. For instance, Hałgas et al. (2023) reviewed evidence showing that team coordination dynamics can be reflected in physiological measures such as heart rate variability (HRV), inter-beat interval (IBI), and respiratory sinus arrhythmia (RSA) as team-level indicators of joint arousal levels and cognitive load. Team-level features also emphasize interaction patterns and collective dynamics. Peifer et al. (2021) collective communication framework captures verbal coordination through speaking patterns and linguistic markers. Gorman et al. (2019) introduced information entropy as a time series approach to quantify system variation and adaptation in teams. These team-level collective and interaction features capture emergent properties beyond individual-level indicators. Evidence also supports that incorporating team-level synchrony metrics can enhance the power of team performance prediction in machine learning models (Rico-González et al., 2023; Yadav et al., 2025).

Approach

In this study, we collected the training data of 16 groups of Fire Support Team (FiST) from Marine Corps training base. Each team is composed of five members, including the Lead, Joint Terminal Attack Controller (JTAC), Fire Support Officer (FSO), Forward Observer/ Mission (FOM), and Forward Observer Assistant (FOA). The simulations mirrored fire suppression operations across three critical phases: preparation, planning, and execution. These scenarios required members to dynamically form sub-teams to manage parallel and sequential subtasks, adding significant complexity to both task execution and interpersonal interactions. During the training sessions, all team members' physiological and behavioral data were collected, including cardiac, respiratory, electroencephalogram (EEG), Functional near-infrared spectroscopy (fNIRS), and their communication logs.

Physiological signals were synchronized to a global reference timeline and downsampled to 1 Hz to generate an aligned raw feature set. For the team-level features, a layered dynamics model, including neuro, autonomic, and communicative adaptation layers, was developed to derive entropy and average mutual information (AMI) (Gorman et al., 2019; Yin et al., 2025). Entropy serves to quantify the variety and reorganization of system states, with higher values indicating a shift in coordination patterns, while AMI provides a measure of influence by quantifying the degree of information exchange and interaction strength among team members. Individual-level features comprised cardiac measures (IBI and HRV), respiratory parameters (rate and amplitude), and EEG frequency bands (alpha, beta, theta, and delta. These features were subsequently processed using a sliding window approach to create an enhanced dataset, with 43 individual-level and 23 team-level metrics, for model training. Specifically, we extracted the mean, standard deviation (STD), root mean square of successive differences (RMSSD), and slope (trend) from the 1 Hz sensor data to capture underlying physiological and behavioral patterns.

Team operational performance was evaluated by training instructors using a four-tier rating system, where Class 1 denoted failure and Class 4 represented full mission success. To assess the incremental predictive power of team-level metrics, we compared classification performance across three feature sets: (1) individual-level only, (2) team-level only, and (3) a combination of both. We evaluated five distinct models: Logistic Regression, Random Forest, eXtreme Gradient Boosting (XGBoost), Multi-Layer Perceptron (MLP), and Support Vector Machine (SVM). Analysis was conducted using sliding time windows ranging from 30 to 180 seconds, analyzed in 30-second increments. The 16 training sessions yielded over 320 windowed samples for the 180-second window. We employed a 5-fold Group K-Fold cross-validation strategy to prevent data leakage between segments of the same session. Model performance was quantified using Accuracy and F1 scores.

Outcome

Among all the tested models and time window configurations, the SVM model with combined features using a 150-second time window demonstrated the best overall performance (accuracy = 0.79, F1 score=0.80, shown in Figure 1), closely followed by the MLP model, which achieved similar performance (accuracy = 0.80, F1 score=0.76) under the same time window. Across all models, the optimal prediction performance was consistently observed with 120 or 150-second windows when using combined features. This optimal window duration may reflect a balance between capturing comprehensive patterns (longer duration) and preserving sufficient temporal resolution (shorter duration) for effective performance prediction. The comparative analysis across different feature groups revealed distinct performance patterns. Team features alone yielded relatively poor prediction performance, with the best-performing MLP model achieving only 0.418 accuracy and 0.421 F1 score. Individual features performed substantially better, with SVM reaching 0.732 accuracy and 0.769 F1 score. Notably, the combined feature approach achieved superior performance compared to using either team features or individual features alone, demonstrating that integrating team-level and individual-level features provides complementary value for prediction accuracy. This pattern was consistent across all tested machine learning models, with combined features showing accuracy improvements of 6-15% over individual features alone.

Feature importance analysis further supported the value of team-level features, though their contribution patterns

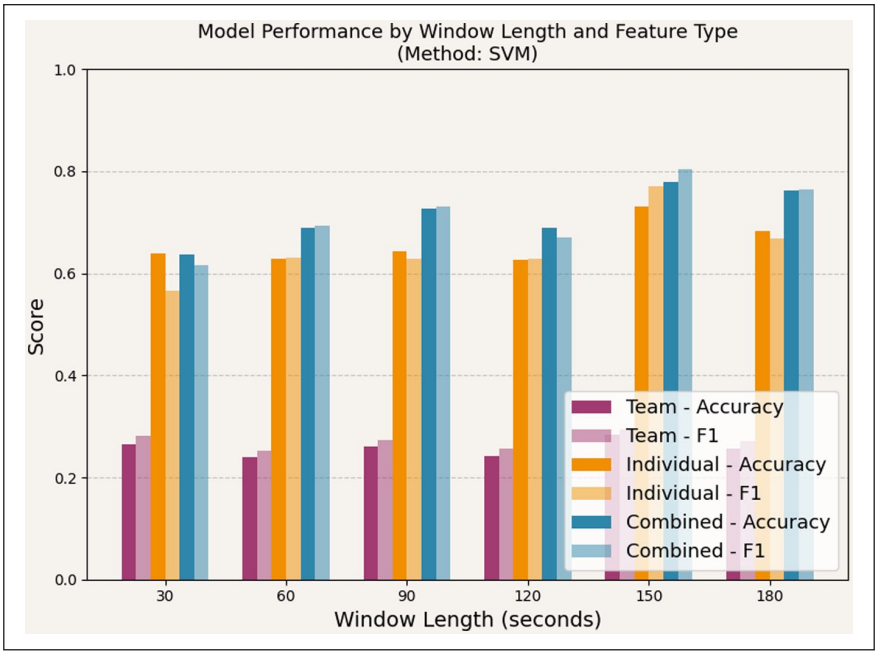

Figure 1. Classification results across individual (SVM).

varied by analysis method. While team features were absent from the top 20 features when using regression-based coefficients, SHAP (SHapley Additive exPlanations) analysis revealed that the team-level mean and standard deviation of EEG beta frequency band emerged as the 6th and 20th most contributing features, showing their non-linear contribution to model performance.

Conclusion

Our findings suggest that integrating team-level physiological features can complement prediction performance compared to using individual-level features alone, extending previous conclusions from dyadic teams to teams with more members and more complex interactions (Li et al. 2025; Zahmat Doost et al. 2025). Although the team-level features adopted in this work are primitive and cannot effectively predict team performance when used alone, the performance improvement achieved by integrating team-level and individual-level features is consistent across models. The SHAP-based feature importance analysis reveals that team-level features contribute to model performance comprehensively in a non-linear manner. These results demonstrate that team-level features can be an effective complementary to performance prediction. This finding underscores that effective measurement of team performance necessitates considering both individual cognitive processes and collective team processes (e.g., frameworks such as Interactive Team Cognition), as both levels provide essential context for understanding the multifaceted nature of team operations. Future work will explore more sophisticated team-level features that better capture interaction patterns in multi-member teams. Overall, this empirical study demonstrates that optimal team performance prediction requires comprehensive analysis from both individual and team perspectives, as neither individual-level nor team-level features alone can achieve the best outcomes.

ORCID iDs

Mengtao Lyu https://orcid.org/0000-0003-1505-1970
Mengyao Li https://orcid.org/0000-0002-0819-4693
Nancy Cooke https://orcid.org/0000-0003-0408-5796
Lixiao Huang https://orcid.org/0000-0002-5543-8742
Xiaoyun Yin https://orcid.org/0000-0002-8629-5836
Elmira Zahmat Doost https://orcid.org/0000-0002-6948-0044

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by The Defense Advanced Research Projects Agency (DARPA) via The Objective Prediction of Team Effectiveness via Models of Performance Outcomes (OP TEMPO) program (HR001124S0006-OPTEMPO).

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Atweh, J. A., Marie Moacdieh, N., & Riggs, S. L. (2022). Identifying individual-, team-, and organizational-level factors that affect team performance in complex domains based

on recent literature. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *66*(1), 1795–1799. https://doi.org/10.1177/1071181322661213

Bisbey, T. M., & Salas, E. (2019). Team dynamics and processes in the workplace. In T. M. Bisbey & E. Salas, *Oxford research encyclopedia of psychology*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190236557.013.13

Gorman, J. C., Demir, M., Cooke, N. J., & Grimm, D. A. (2019). Evaluating sociotechnical dynamics in a simulated remotely-piloted aircraft system: A layered dynamics approach. *Ergonomics*, *62*(5), 629–643. https://doi.org/10.1080/00140139.2018.1557750

Hałgas, E. A., Van Eijndhoven, K. H. J., Gevers, J. M. P., Wiltshire, T. J., Westerink, J. H. D. M., & Rispens, S. (2023). A review of using wearable technology to assess team functioning and performance. *Small Group Research*, *54*(1), 41–76. https://doi.org/10.1177/10464964221125717

Li, M., Rojas, E., Bracken, B., Cooke, N., Desrochers, P., Gorman, J., Huang, L., Kilcullen, M., Rosen, M., Winder, A., Yadav, A., Yin, X., Zahmat Doost, E., & Zhou, S. (2025). Physiological synchrony as a predictor of team performance: A machine learning approach. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *69*(1), 333–337. https://doi.org/10.1177/10711813251357937

Peifer, C., Pollak, A., Flak, O., Pyszka, A., Nisar, M. A., Irshad, M. T., Grzegorzek, M., Kordyaka, B., & Kożusznik, B. (2021). The symphony of team flow in virtual teams. Using artificial intelligence for its recognition and promotion. *Frontiers in Psychology*, *12*, 697093. https://doi.org/10.3389/fpsyg.2021.697093

Rico-González, M., Pino-Ortega, J., Méndez, A., Clemente, F., & Baca, A. (2023). Machine learning application in soccer: A systematic review. *Biology of Sport*, *40*(1), 249–263. https://doi.org/10.5114/biolsport.2023.112970

Strang, A., Funke, G. J., Knott, B. A., & Warm, J. S. (2011). Physio-behavioral synchronicity as an index of processes supporting team performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *55*(1), 1447–1451. https://doi.org/10.1177/1071181311551301

Yadav, G. A., Bracken, B. K., Cooke, N. J., Desrochers, P., Gorman, J. C., Grimm, D. A. P., Huang, L., Kilcullen, M., Li, M., Rojas, E., Rosen, M., Scalia, M. J., Winder, A., Yin, X., Doost, E. Z., & Zhou, S. (2025). Bio-behavioral team dynamics measurement system: Multimodal sensing, dynamical systems modeling, and machine learning pipelines to predict and characterize team performance. *Nonlinear Dynamics, Psychology & Life Sciences*, *29*(4), 529.

Yin, X., Grimm, D. A. P., Zhou, S., Pisors, J., Yang Teo, A. Z., Lijenstolpe, M., Doost, E. Z., Cooke, N. J., & Gorman, J. C. (2025). Observing resilient performance in space-based teaming mission using digital technologies. *Applied Ergonomics*, *124*, 104409. https://doi.org/10.1016/j.apergo.2024.104409

Zahmat Doost, E., Gorman, J. C., Yin, X., Zhou, S., Yadav, G. A., Cooke, N., Huang, L., Bracken, B. K., Winder, A., Desrochers, P., Li, M., Rojas, E., Rosen, M., & Kilcullen, M. (2025). BioTDMS: A multi-sensor and data-driven system for real-time team performance assessment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *69*(1), 1255–1259. https://doi.org/10.1177/10711813251369877