Parametrized Dataset Generator for the Classification of Ice Hockey Power Plays

Ken Mangouh Nsiempba*
John Zelek, David Clausi
Vision and Image Processing Group, Systems Design Engineering, University of Waterloo
{kmnsiemp, jzelek, dclausi}@uwaterloo.ca

Abstract

The advent of deep learning tools has significantly enhanced researchers' capabilities in analyzing spatio-temporal data. This type of data analysis holds relevance across various domains.
Improving the ability to identify and cluster patterns within sporting events has profound implications. It can aid in automatic highlight detection and is especially beneficial for coaching, particularly in underfunded and minor leagues. While the insights presented in this paper can be applied to numerous team sports, our focus primarily lies on ice hockey. In this paper, we make three significant contributions: we introduce a simple, parametrized ice hockey formation dataset generator facilitating the development and benchmarking of baseline models; we investigate the impact of noise from the dataset on the accuracy of event classification; and we compare the accuracies of three models: K-Nearest-Neighbors, Graph Networks, and Convolutional Neural Networks.

Introduction

The field of sports analytics leverages statistical, computation and data-driven techniques to gain insights into various aspects of sports performance, strategy and management. For example, we can study a particular pass in ice hockey sequence and players' formation in an offensive zone and use statistical methods to infer the probability of it resulting in a goal. To do so, it is important that we are able to detect those players' formation/sequence so that we acquire enough data about their output. Amassing enough samples about an event can be challenging and thus the purpose of this paper. The overarching goal of this research is to create a synthetic dataset on which models will be trained to detect ice hockey power plays from the players' tracklets. A power play in ice hockey occurs when one team has a numerical advantage due to an opposing player being penalized and sent to the penalty box. The purpose of a power play is to create a temporary imbalance, offering the advantaged team an opportunity to exert offensive pressure and score goals. The disadvantaged team, playing "short-handed," focuses on defense and preventing goals. Detecting such events during footage analysis can enable teams to study their opponents, focusing on the frequency and execution methods of power plays. There are three forms of power plays which will be described [1]:

- The umbrella power play: The umbrella formation involves positioning three players high in the offensive zone: one at the center point and two at the tops of the faceoff circles, creating an "umbrella" shape. This setup emphasizes puck movement along the perimeter and quick shots from the point or top of the circles. It aims to stretch the penalty killers horizontally and create shooting lanes for players in the high slot or net-front areas (see figure 1a).
- The 1-3-1 formation: The 1-3-1 is a dynamic and modern power-play structure that balances players across the offensive zone in a 1-3-1 alignment: one player at the blue line (point), three players across the middle (two wingers and a bumper), and one player stationed in front of the net. This formation offers versatility, allowing for quick puck movement and multiple shooting threats. The bumper player in the slot is crucial for redirecting shots, creating screens, or making short passes to maintain puck possession (see figure 1b).
- The spread formation: The spread, or overload, places an emphasis on puck control and creating space by "overloading" one side of the ice. Typi-

*website: https://botengu.github.io/portfolio/

cally, three players work in close proximity along the boards or in the corner, with one player positioned in front of the net and another at the far point. This setup relies on quick passes and strong puck possession to confuse the penalty killers and open up shooting opportunities (see figure 1c).

Creating a dataset would require scanning through hours of ice hockey footage, detect and track players who are about to be in a power play formation. Furthermore, the task would be made even greater since these events are relatively rare and in existing footage, some players are sometimes not in the field of view or partially occludes. To the best of the authors' knowledge, no dedicated "power play" dataset has been created to train classifiers for their detection. Moreover, the authors recognize that player locations during power plays tends to exhibit relatively small variations. Therefore, creating a dataset that simulates player trajectories with a bit of noise would be highly beneficial for training custom models (as it has been done in other sports [2, 3]). These models would learn to recognize power plays despite randomness in trajectories and positional differences, which make them more realistic. Once trained, such models could be used to classify tracked players in ice hockey footage effectively. Thus further helping in coaching and/or extracting highlights.

The paper makes three significant contributions:

- It introduces a parametrized elementary dataset generator tailored for the study of ice hockey formations.
- It examines the influence of noise from the dataset on the accuracy of event classification.
- It evaluates the effectiveness of three baseline models: K-Nearest-Neighbors (KNNs), Graph Neural Networks (GNNs), and Convolutional Neural Networks (CNNs).

Section will delve into similar works, section will discuss the intricacies of the models which have been used in order to evaluate the datasets and section will discuss the results.

Related works

This research lies at the intersection of two fields: Multi-agent systems and spatio-temporal event recognition.

Event recognition is centered around identifying and understanding human activities or events from video or sensor data. It involves analyzing data to discern patterns and sequences that indicate specific events or actions. On the other hand, multi-agent systems (MAS) delve into the interactions among autonomous agents, exploring how these interactions can be modeled and predicted. This field seeks to understand individual and collective agent behaviors, often encompassing complex decision-making processes.

In the realm of sports, there are compelling reasons to gather and to study spatio-temporal data. The literature highlights various motivations to do so:

- Spatio-temporal data in sports is utilized to predict the outcomes of events or games [4, 5, 6, 7, 8, 9, 10, 11].
- Spatio-temporal data in sports is also employed to study relationships between sport variables [12, 13].
- Spatio-temporal data is also used to establish success metrics [14, 15, 16, 17, 15].
- Furthermore, spatio-temporal data can be used to generate optimal strategies [18, 19].
- Lastly, spatio-temporal data is employed for the analysis and detection of patterns [20, 21, 22, 23, 24, 25, 26, 27].

Those reasons are all further motivation for us to be able to properly train models to detect events (ice hockey power play formation). The next section will discuss the methodology used to generate and evaluate the datasets.

Methods

Dataset generation

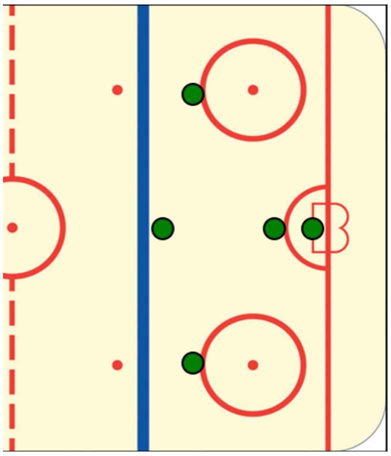
The dataset generation process involves defining three player configurations: the umbrella power play, 1-3-1 power play, and spread power play. These formations are represented through player locations on the ice. We chose these formations because they are the most frequent and sufficiently distinct, making it possible to create a dataset with well-separated classes that increase the likelihood of the classifier accurately distinguishing between them.

Those locations are encoded using the formula :

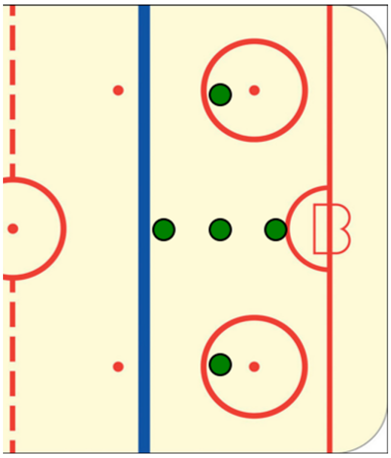
$$\boldsymbol{P_{conf}} = \{(P_{x,i}, P_{y,i}) | i = 1, \dots, 5\} \tag{1}$$

where $(P_{x,i}, P_{y,i})$ are the specific coordinates for player i.

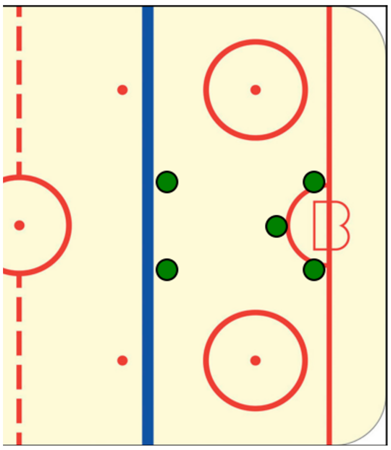
Noise is then added to the players' locations using $\boldsymbol{X}$, a 5×2 random array. The noise is bounded by p_m, our first parameter, which represents a fraction of the rink:



(a) **umbrella**



(b) **1-3-1**



(c) **spread**

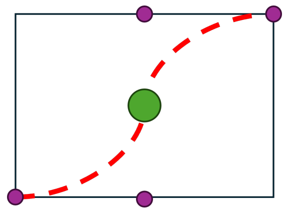
Figure 1: The **umbrella**, the **1-3-1**, and the **spread** power plays. The green dots represent the ideal locations of the players, all located in the offensive zone. Only half of the rink is shown.

$$\boldsymbol{X} \sim Uniform(0,1)^{5\times 2} \tag{2}$$

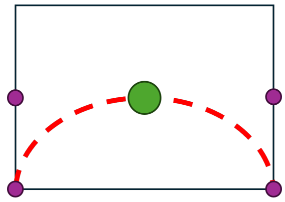
$$\widetilde{\boldsymbol{P}}_{\boldsymbol{conf}} = max(0, min(\boldsymbol{P_{conf}} + \boldsymbol{X} \cdot p_m, 1)) \tag{3}$$

Where $\widetilde{\boldsymbol{P}}_{\boldsymbol{conf}}$ is the modified (noisy) configuration. We clip the array to ensure that the locations are within the rink's dimensions.

To generate the trajectories around the players, we create an imaginary rectangle (of dimensions **w** and **h**) in to limit the extent of the players' trajectories. We apply Bezier curve and generate two curves, one being the first half of the trajectory and the other one being the second half of the trajectory. Equation shows the computation for the first trajectory (2) configurations of figure 2.



(a) Trajectory configuration 1.



(b) Trajectory configuration 2.

Figure 2: The two trajectory configurations. Control points are in purple, and the resulting Bézier curves are the dashed red lines.

$$\mathbf{P1}_{ctl} = \{\widetilde{P}_i - [-\frac{w}{2}, \frac{h}{2}], \widetilde{P}_i - [0, \frac{h}{2}], \widetilde{P}_i\} \tag{4}$$

$$\mathbf{P2}_{ctl} = \{\widetilde{P}_i, \widetilde{P}_i - [0, -\frac{h}{2}], \widetilde{P}_i - [-\frac{w}{2}, \frac{h}{2}]\} \tag{5}$$

$$t_k = \frac{k}{n}, \quad k = 0, 1, 2, \ldots, n \tag{6}$$

$$B1(t_k) = \sum_{i=0}^{2} \binom{2}{i} (1-t_k)^{2-i} t_k^i \, \mathbf{P1}_{ctl}[i] \tag{7}$$

$$B2(t_k) = \sum_{i=0}^{2} \binom{2}{i} (1-t_k)^{2-i} t_k^i \, \mathbf{P2}_{ctl}[i] \tag{8}$$

$$Traj_{pts} = \text{concat}(B1(t_k), B2(t_k)) \tag{9}$$

Where $\mathbf{P1}_{ctl}$ and $\mathbf{P2}_{ctl}$ are the control points for both sides of the trajectories. $B1(t_k)$ and $B2(t_k)$ are bezier curve points for both sides of the trajectories. The concatenation of both elements yields $Traj_{pts}$, the final curve points. Each trajectory is then rotated by a random angle that varies between 0 and 2π.

Architectures of the Models

We initially considered using rules based on location margins from the players' positions to recognize power plays—for example, determining whether the players' positions fall within a specific threshold and applying deductive logic to infer the power play. However, this methodology is less effective when accounting for the type of noise present. The critical factor goes beyond the absolute or relative positions of the players. As the names of some power plays suggest, the players can be viewed as forming a shape, which may be compressed or stretched along various axes (as long as certain rules are respected, e.g., all players remain in the offensive zone). This transforms the problem into one resembling shape classification, for which machine learning provides more effective tools. Machine learning models, particularly those designed for spatiotemporal data such as trajectories, are better equipped to handle noise. They can generalize patterns and ignore irrelevant variations caused by noise, outperforming rule-based methods in these scenarios.

Three machine learning models were used to classify the data:

- K-nearest neighbours: KNNs were used to compare the performance of our models. We set the number of neighbours to 5.
- Convolutional Neural Network: For this network, we treated our data as a grayscale image. Instead of two spatial dimensions, we have one spatial dimension and one temporal dimension. CNNs were chosen for their ability to effectively take neighboring features into account when handling 2D data structures. The output of CNN layer $H^{(l+1)}$ is given by:

$$H^{(l+1)} = \sigma\left(W^{(l)} * H^{(l)} + b^{(l)}\right) \tag{10}$$

  Where $H^{(l)}$ is the input feature map at layer l, $W^{(l)}$ represents the convolutional filter or weight matrix for layer l, $*$ denotes the convolution operation, $b^{(l)}$ is the bias term applied at layer l, σ is the activation function, in this case, ReLU, $H^{(l+1)}$ is the resulting feature map after the convolution operation and activation function are applied.

- Graph Neural Network: The network consisted of two graph convolutional layers that involved message passing. An edge was created between each pair of points. The output of GCN layer is given by $H^{(l)}$:

$$H^{(l+1)} = \sigma\left(\hat{A}H^{(l)}W^{(l)}\right) \tag{11}$$

  Where: X is the input feature matrix, $\hat{A}$ is the normalized adjacency matrix of the graph, $W^{(l)}$ is the weight matrix for the previous GCN layer, σ is an activation function (ReLU in this case), $H^{(l)}$ is the output of the previous GCN layer.

The experimental setup involved creating five datasets by varying the positional margin (p_m) from 0.0 to 0.4 with 0.1 increments. For each dataset, 2000 samples were generated for four classes, and player trajectories were represented by 40 points (n=20). The data was processed by stacking the x and y coordinates into a vector of 80 values per player.

Results

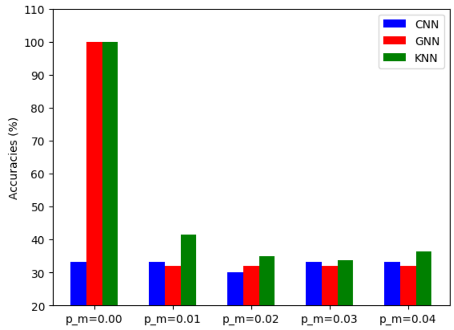
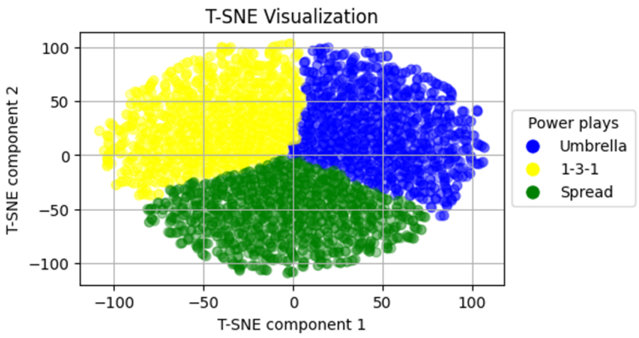


Figure 3: Plots showing the accuracies of all three models across the various datasets

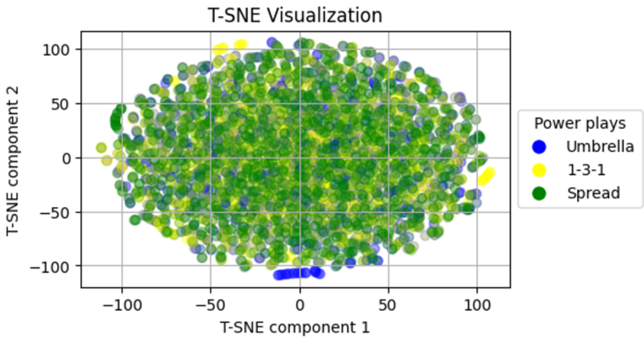
For the KNN, we used a fixed k of 5. For the experiments utilizing both GNN and CNN models, the data was split 80-20% into a training set and a test set, with the training set itself further divided into a 75-25% split for training and validation. We performed cross-validation on the training and validation sets and evaluated the best model on the test set after cross-validation was completed. We used Xavier weight initialization and an Adam optimizer. Each fold ran for 5 epochs with a learning rate of 0.01 and a weight decay of 10^{-5}. Results are shown in figure 3

For $p_m = 0.00$, GNN and KNN both achieved 100% accuracy, indicating their ability to classify perfectly when there were no positional variations in the dataset. CNN performed significantly worse, suggesting it struggled to generalize even without noise. The T-SNE visualization in figure 4a showed that all the power plays were clearly clustered.

For $p_m = 0.01$ to $p_m = 0.04$, the accuracy dropped substantially compared to $p_m = 0.00$ for all models, highlighting the increased difficulty of classification as positional margins introduced variability. KNN consistently outperformed the other models, while CNN and



(a) $p_m = 0$



(b) $p_m = 0.04$

Figure 4: Visualization of the dataset T-SNE embeddings. (a) Embeddings for $p_m = 0$. (b) Embeddings for $p_m = 0.04$.

GNN performed similarly. The T-SNE visualization for $p_m = 0.04$ (figure 4b) showed that there were no clear clusters anymore.

Conclusion

In summary, this work has achieved several objectives. First, it generated synthetic ice hockey power play sequences with parametrizable temporal and spatial features. Second, it demonstrated their compatibility with existing networks, suggesting they serve as a solid foundation for future baselines. And third it compared the classification accuracies from the different baselines.

However, further work is needed to enhance these results. Adding more variations to the dataset might ensure more conclusive findings. The simulation should also better represent actual hockey sequences. For instance, incorporating greater complexity into player paths to produce more realistic trajectories based on real-life data. Introducing a binary feature indicating which player possesses the puck, and adding player roles (forward vs. defender) will be beneficial. Additionally, enabling the scaling of formation shapes should be explored to diversify the data. Moreover, we plan to increase the number of parameters in the deep learning networks to improve the classification accuracies.

Acknowledgements

We would like to acknowledge Stathletes and the IBET fellowship for their support.

References

[1] B. S. Blogs, "Complete idiot's guide to hockey systems: Special teams," 2011, accessed: 2024-12-18. [Online]. Available: https://blueseatblogs.com/2011/10/11/complete-idiots-guide-to-hockey-systems-special-teams/

[2] L. Thorpe, L. Bawden, K. Vendal, J. Bronskill, and R. E. Turner, "SportsNGEN: Sustained Generation of Multi-player Sports Gameplay," Feb. 2024, arXiv:2403.12977 [cs, eess, stat]. [Online]. Available: http://arxiv.org/abs/2403.12977

[3] S. Zheng, Y. Yue, and J. Hobbs, "Generating long-term trajectories using deep hierarchical networks," *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[4] M. A. Alcorn and A. Nguyen, "baller2vec: A Multi-Entity Transformer For Multi-Agent Spatiotemporal Modeling," Sep. 2021, arXiv:2102.03291 [cs]. [Online]. Available: http://arxiv.org/abs/2102.03291

[5] K. Zhao, C. Du, and G. Tan, "Enhancing Basketball Game Outcome Prediction through Fused Graph Convolutional Networks and Random Forest Algorithm," *Entropy*, vol. 25, no. 5, p. 765, May 2023, number: 5 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/1099-4300/25/5/765

[6] D. Ding and H. H. Huang, "A Graph Attention Based Approach for Trajectory Prediction in Multi-agent Sports Games," Dec. 2020, arXiv:2012.10531 [cs]. [Online]. Available: http://arxiv.org/abs/2012.10531

[7] Y. Yuan, X. Weng, Y. Ou, and K. M. Kitani, "AgentFormer: Agent-Aware Transformers for Socio-Temporal Multi-Agent Forecasting," 2021, pp. 9813–9823. [Online]. Available: https://openaccess.thecvf.com/content/ICCV2021/html/Yuan_AgentFormer_Agent-Aware_Transformers_for_Socio-Temporal_Multi-Agent_Forecasting_ICCV_2021_paper.html

[8] H. Kim, H.-J. Choi, C. J. Kim, J. Yoon, and S.-K. Ko, “Ball Trajectory Inference from Multi-Agent Sports Contexts Using Set Transformer and Hierarchical Bi-LSTM,” in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Aug. 2023, pp. 4296–4307, arXiv:2306.08206 [cs]. [Online]. Available: http://arxiv.org/abs/2306.08206

[9] T. Decroos, V. Dzyuba, J. V. Haaren, and J. Davis, “Predicting Soccer Highlights from Spatio-Temporal Match Event Streams,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, Feb. 2017, number: 1. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/10754

[10] J. Dong, Q. Huo, and S. Ferrari, “A Holistic Approach for Role Inference and Action Anticipation in Human Teams,” *ACM Transactions on Intelligent Systems and Technology*, vol. 13, no. 6, pp. 95:1–95:24, Sep. 2022. [Online]. Available: https://dl.acm.org/doi/10.1145/3531230

[11] G. Li, B. Jiang, H. Zhu, Z. Che, and Y. Liu, “Generative attention networks for multi-agent behavioral modeling,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 05, 2020, pp. 7195–7202.

[12] M. J. Lopez, G. J. Matthews, and B. S. Baumer, “How Often Does the Best Team Win? A Unified Approach to Understanding Randomness in North American Sport,” *The Annals of Applied Statistics*, vol. 12, no. 4, pp. 2483–2516, 2018, publisher: Institute of Mathematical Statistics. [Online]. Available: https://www.jstor.org/stable/26666161

[13] N. E. Dona and T. Swartz, “A causal investigation of pace of play in soccer,” *Statistica Applicata - Italian Journal of Applied Statistics*, no. 1, May 2023, number: 1. [Online]. Available: https://www.sa-ijas.org/ojs/index.php/sa-ijas/article/view/195

[14] R. Found, “Goal-based Metrics Better Than Shot-based Metrics at Predicting Hockey Success,” Jan. 2016.

[15] W. Zhou, G. Yu, S. You, and Z. Wang, “An Improved Passing Network for Evaluating Football Team Performance,” *Applied Sciences*, vol. 13, no. 2, p. 845, Jan. 2023, number: 2 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2076-3417/13/2/845

[16] M. Evers, A. Derstroff, S. Leistikow, T. Schneider, L. Mallepree, J. Stampke, M. Leisgang, S. Sprafke, M. Schuhl, N. Krefft, F. Droese, and L. Linsen, “Visual analytics of soccer player performance using objective ratings,” *Information Visualization*, p. 14738716231220539, Jan. 2024, publisher: SAGE Publications. [Online]. Available: https://doi.org/10.1177/14738716231220539

[17] A. Sahasrabudhe and J. Bekkers, “A graph neural network deep-dive into successful counterattacks,” accessed: 2024-02-15. [Online]. Available: https://www.sloansportsconference.com/

[18] Z. Wang, P. Veličković, D. Hennes, N. Tomašev, L. Prince, M. Kaisers, Y. Bachrach, R. Elie, L. K. Wenliang, F. Piccinini, W. Spearman, I. Graham, J. Connor, Y. Yang, A. Recasens, M. Khan, N. Beauguerlange, P. Sprechmann, P. Moreno, N. Heess, M. Bowling, D. Hassabis, and K. Tuyls, “TacticAI: an AI assistant for football tactics,” *Nature Communications*, vol. 15, no. 1, p. 1906, Mar. 2024, publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/s41467-024-45965-x

[19] N. Epasinghege Dona and T. B. Swartz, “Causal Analysis of Tactics in Soccer: The Case of Throw-ins,” *IMA Journal of Management Mathematics*, vol. 35, no. 1, pp. 111–126, Dec. 2023. [Online]. Available: https://academic.oup.com/imaman/article/35/1/111/7304125

[20] D. Raabe, R. Nabben, and D. Memmert, “Graph representations for the analysis of multi-agent spatiotemporal sports data,” *Applied Intelligence*, vol. 53, no. 4, pp. 3783–3803, Feb. 2023. [Online]. Available: https://doi.org/10.1007/s10489-022-03631-z

[21] A. Grunz, D. Memmert, and J. Perl, “Tactical pattern recognition in soccer games by means of special self-organizing maps,” *Human Movement Science*, vol. 31, no. 2, pp. 334–343, Apr. 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167945711000613

[22] G. Anzer, P. Bauer, U. Brefeld, and D. Fassmeyer, “Detection of tactical patterns using semi-supervised graph neural networks.”

[23] R. Marcelino, J. Sampaio, G. Amichay, B. Gonçalves, I. D. Couzin, and M. Nagy, “Collective movement analysis reveals coordination tactics of team players in football matches,” *Chaos, Solitons & Fractals*, vol. 138, p. 109831, Sep. 2020.

[Online]. Available: https://www.sciencedirect.com/science/article/pii/S0960077920302319

[24] M. Garnica-Caparrós and D. Memmert, "Understanding gender differences in professional European football through machine learning interpretability and match actions data," *Scientific Reports*, vol. 11, no. 1, p. 10805, May 2021, publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/s41598-021-90264-w

[25] P. Bauer, G. Anzer, and L. Shaw, "Putting team formations in association football into context," *Journal of Sports Analytics*, vol. 9, no. 1, pp. 39–59, Mar. 2023. [Online]. Available: https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/JSA-220620

[26] F. Korte and M. Lames, "Characterizing different team sports using network analysis," *Current Issues in Sport Science (CISS)*, vol. 3, pp. 005–005, Apr. 2018. [Online]. Available: https://ciss-journal.org/article/view/7563

[27] J. Lopes, L. Mota, S. Mota, J. Torres, R. Silva Moreira, C. Soares, I. Pereira, F. Gouveia, and P. Sobral, *Object and Event Detection Pipeline for Rink Hockey Games*, Apr. 2024.