A Gaussian Mixture Model Approach for Characterizing Playing Styles of Ice Hockey Players

Priyansh Gupta[1], Mikael Vernblom[2], Niklas Carlsson[1], and Patrick Lambrix[1]

[1] Linköping University, Sweden
[2] Linköping Hockey Club, Sweden

Abstract. Player categorization based on playing style is a highly important task in professional ice hockey, aiding scouting, player development, and strategic decision-making. Traditional methods often rely on simple metrics like goals or assists, which fail to capture the full complexity of a player's style and contributions. Motivated by the increasing availability of detailed event data and advances in machine learning based modeling techniques, this paper explores a richer, data-driven approach to player categorization. We build on recent work in player vector representations and apply Gaussian Mixture Models (GMMs) to cluster forwards and defenders based on event data from five seasons of the Swedish Hockey League (SHL). Our contributions are threefold: (1) we construct detailed player vectors that summarize a wide range of offensive and defensive skills, (2) we apply GMMs to identify soft clusters of players, allowing for nuanced overlapping playing styles, and (3) we analyze the resulting clusters to interpret distinct player profiles and provide concrete examples. Our results offer a more flexible and realistic view of player roles, reflecting the continuous and multi-dimensional nature of playing styles. The approach helps enhance talent evaluation and roster building, and offers an efficient framework for future analyses across leagues and seasons.

1 Introduction

Player categorization based on playing style is an important task in professional ice hockey, supporting scouting, player development, and strategic decision making. Traditional approaches typically rely on discrete performance metrics, such as goals, assists, or shots, offering only a partial view of a player's overall style and contribution. More recently, increased event data collection and advances in modeling techniques have opened up new possibilities for representing and analyzing player behaviors in more nuanced ways.

In this paper, we build upon recent developments in player vector representations [17] and apply Gaussian Mixture Models (GMMs) to identify clusters of forwards and defenders based on their playing styles. GMMs offer a probabilistic soft clustering approach that is particularly well-suited to model the continuous and overlapping nature of player styles. Unlike hard clustering methods, which assign each player to a single cluster, GMMs allow players to belong to multiple clusters with varying degrees of membership, reflecting the reality that players often exhibit characteristics of multiple styles.

The specific contributions of this paper are threefold. First, we leverage detailed event data from five seasons of the Swedish Hockey League (SHL) to construct player vectors capturing a wide range of offensive and defensive skills. Second, we apply GMMs to these vectors, determining the number of clusters using model selection criteria such as the Bayesian Information Criterion (BIC). Finally, we analyze the resulting clusters to interpret the different playing styles represented among forwards and defenders, and provide examples of players associated with each style.

Organization: Section 2 provides background on finite mixture models, Gaussian mixture models, and model selection methods. Section 3 reviews related work in player evaluation and categorization. Section 4 describes the dataset used in this study. Section 5 outlines our methodology for constructing player vectors and fitting GMMs. Section 6 presents the clustering results and analyzes the identified playing styles, before Section 7 concludes the paper.

2 Background

In this paper, we apply Gaussian Mixture Models (GMMs) to cluster forwards and defenders based on their playing styles. A GMM is a type of Finite Mixture Model (FMM) where each component is a Gaussian distribution. GMMs are particularly well-suited for player data, as different playing styles often overlap and evolve along continuous spectrums, making soft clustering approaches like GMMs more appropriate than hard clustering alternatives.

An FMM models data from a combination of unobserved groups, without knowing in advance which point belongs to which group. With FMMs, each group is associated with its own probability distribution, and the overall dataset is modeled as a weighted sum of these components. Instead of trying to fit just one model to the entire dataset, this allows the FMMs to fit multiple smaller models and combine them [16]. This offers a flexible framework that better captures complex data structures than single-model approaches.

Although various distributions, such as Poisson and Bernoulli, can be used within the FMM framework, the choice of Gaussian distributions in GMMs provides two key advantages: flexibility and interpretability. Gaussian components can model elliptical clusters with different orientations and scales, which is important when different playing styles vary along distinct combinations of performance features. Furthermore, GMMs naturally produce soft assignments of players to clusters, reflecting the intuition that playing styles often exist on a continuum rather than falling into rigid categories.

Formally, an FMM assumes that each observation y comes from one of g different groups (components), each described by its own distribution. These components are mixed using probabilities $\pi_1, \pi_2, ..., \pi_g$, where each π_i is the mixing proportion for the i-th component. These values are all positive and add up to 1. The overall distribution of an FMM can either be a probability density function (PDF), if the data is continuous, or a probability mass function (PMF) in the case of a discrete dataset [2,15]. More specifically, the PDF or PMF of the mixture model is represented as follows:

$$f(y) = \sum_{i=1}^{g} \pi_i f_i(y), \tag{1}$$

where $f_i(y)$ represents the PDF or PMF of the i-th component, π_i represents the mixing proportion of the component, and g is the total number of components [16]. Depending on the type of distribution, the component densities $f_i(y)$ are represented as $f_i(y, \theta_i)$, where θ_i is the vector of unknown parameters for the i-th component density. In the case of a Gaussian distribution, these parameters are the mean and variance $\theta_i = [\mu_i, \sigma_i^2]$, resulting in the following way to represent the PDF or PMF of the mixture model:

$$f(y, \Psi) = \sum_{i=1}^{g} \pi_i f_i(y, \theta_i), \tag{2}$$

where y is the data we want to model, and Ψ is a vector containing all the unknown parameters in the mixture model, such as π_i and θ_i for i-th component. It can be expressed as $\Psi = (\pi_1, ..., \pi_{g-1}, \zeta^T)^T$, where the parameter ζ includes the parameters of the selected distributions for all g components [15]. In the context of this study, y refers to a player vector that characterizes an individual player's style, and the parameters Ψ collectively describe how these playing styles are distributed across the population.

To estimate the values of the parameters of a FFM, i.e., the mixing proportion π and the parameters of each component distribution θ, several approaches exist, but the most commonly used is the Expectation-Maximization (EM) algorithm. EM applies maximum likelihood estimation to fit the FMM. The EM algorithm consists of two main steps: the Expectation (E-step), which calculates the probability that each data point belongs to each component based on the current parameter estimates, and the Maximization (M-step), which updates the parameter estimates using these probabilities to better fit the data. These two steps are repeated iteratively until convergence is achieved [15].

A key challenge in applying FMM is identifying the value of g, i.e., the number of components in the model. Lower values of g may lead to underfitting, while higher values can result in overfitting [15]. To address this, several model selection criteria have been developed, aiming to balance model fit with complexity. Two widely used criteria for model selection are the Akaike Information Criterion (AIC) [1] and the Bayesian Information Criterion (BIC) [27]. Both criteria evaluate models based on the maximized likelihood $\hat{L}$, while introducing a penalty that increases with the number of estimated parameters $|\Psi|$; thus, discouraging overfitting. The AIC provides an estimator of the relative prediction error, calculated as follows

$$\mathrm{AIC} = 2|\Psi| - 2\ln(\hat{L}), \tag{3}$$

which can be used to compare the quality of multiple models fitted to the same dataset. Lower AIC values indicate models that are expected to predict new data more accurately. Similarly, the BIC is given by

$$\mathrm{BIC} = |\Psi| \ln(n) - 2\ln(\hat{L}), \tag{4}$$

where n is the number of observations. BIC imposes stricter penalty on model complexity compared to AIC, making it more conservative when selecting the number of components, specially for large datasets [28].

In practice, both AIC and BIC are calculated for models with different values of g, and the model with the lowest value of the selected criterion is considered optimal. While AIC favors more complex models, BIC generally performs better in identifying an adequate number of components in FMMs, especially in the context of large datasets [15,28].

3 Related Work

Characterizing and comparing players in ice hockey has been done in different ways. The most common approach is to use performance metrics [8]. These range from the traditional metrics such as goals, assists, and points to Corsi and xG (expected goals) which are all well-known in the hockey discourse. To deal with some of the disadvantages of the traditional metrics, other advanced data-driven metrics have been proposed such as extensions for the +/- metric using regularized logistic regression models [14,4]. There is also work on combining metrics, such as in [5] where principal component analysis is used on 18 basic stats. A major critique for traditional metrics has been that context is not taken into account. Therefore, some approaches for player performance metrics take game context into account such as event impacts, e.g., [24,19], and much of the work that models the dynamics of an ice hockey game using Markov games where two opposing sides (i.e., the home team and the away team) try to reach states in which they are rewarded (e.g., scoring a goal) [29,7,20,25,26,11,22,13,9]. We note that the introduction of new metrics may change the way the game is played. For instance, in [6] it was shown that team play transitioned first to taking more shots (high Corsi, shot-based), and then to taking high-quality shots (high expected goals). Player rankings are presented in [23,12,10].

Player categorization is a relatively unexplored field in the context of ice hockey. In earlier work, a player could belong to only one role or category [30,3]. More recent work used soft clustering techniques to categorize players, allowing for a player to belong to different roles with some probability [21,17]. In the latter case, players can be compared based on their membership in different roles.

This work can be seen as a variant of the work of [17]. In that paper, we used player vectors to characterize a player's playing style. The player vectors contain representations of skills that are computed from game event data. Further, we used fuzzy clustering on the vectors to generate five types of defender playing styles and five types of forward playing styles. For these types, we showed typical skill levels and players with similar styles. The data included complete seasons for the three leagues AHL, SHL, and HockeyAllsvenskan for 2021/22 and 2022/23, as well as data from the 2023/24 season up until Jan. 28th, 2024.

4 Data

The dataset used in this research is a proprietary dataset developed by Sportlogiq[3], and consists of event data for all the SHL regular season games for 5 seasons (2019/20 to 2023/24). In total, the dataset consist of 1820 games, 1072 unique players, 16 unique teams, and 6,814,336 events. Among the 1072 unique players, there are 656 forwards, 377 defenders, and 94 goaltenders. We note that in the dataset 55 players have been marked as playing in more than one position.

5 Method

5.1 Player vectors

We use the player vectors introduced in [17], and briefly summarizing how these were constructed here. The player vectors for defenders and forwards were constructed by concatenating the feature vectors for the skills applicable to each of the two player categories. Tables 1 and 2 summarize these skills. More specifically, for each skill, a feature vector is constructed which contains the frequencies of each feature that describes that skill standardized using MinMaxScaler in the scikit-learn library for Python [18]. Further, non-negative matrix factorization (NMF) was applied to each feature vector using the NMF in the scikit-learn library. After this operation, every skill is represented by one feature and these are concatenated into player vectors (i.e., vectors of length 13 for defenders and of length 18 for forwards).

Figs. 1a and 1b use boxplots to show the distributions of the values for the skills for defenders and forwards, respectively, for the dataset containing all seasons. Here, the lower edge of the box represents the lower quartile value (25%) value, the (red) line in the box the median (50%) value, and the upper edge of the box the higher quartile (75%) value. The lower whisker shows the minimum value and the upper whisker the maximum value. Points below the lower whisker or above the upper whisker are outliers.

5.2 Gaussian Mixture Model

To decide on the number of clusters for forwards and for defenders, we used the BIC approach. A full-factorial grid search was performed to identify different configurations of the model; resulting in a total of 14,400 different models being evaluated, each with different values of parameters such as number of components, covariance types, maximum number of iterations, and different initialization methods. Table 3 summarizes the example values used for each parameter.

[3] https://www.sportlogiq.com/hockey/

Table 1: Skills and example actions for defenders [17].

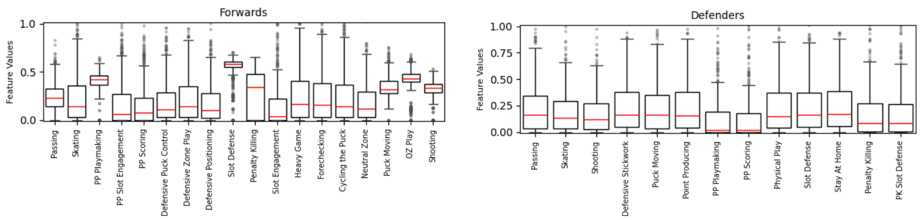
Skills	Actions
Passing	e.g., different types of passes
Skating	e.g., exits, entries, dumps
Shooting	e.g., different types of shots
Defensive Stickwork	e.g., blocked passes, loose puck recoveries
Puck Moving	e.g, some types of passes, dump-in recoveries
Point Producing	e.g., different offensive zone events
Powerplay Playmaking	e.g., powerplay playmaking events
Powerplay Scoring	e.g., powerplay shots and goals
Physical Play	e.g., body checks and defensive plays
Slot Defense	e.g., blocked shots and dump outs
Stay at Home	e.g., different defensive zone events
Penalty Killing	e.g., different penalty killing events related to puck recovery
Penalty Killing Slot Defense	e.g., different penalty killing defensive plays

Table 2: Skills and example actions for forwards [17].

Skills	Actions
Passing	e.g., different types of passes
Skating	e.g., different types of controlled entries
Powerplay Playmaking	e.g., different types of controlled entries and passes in powerplay
Powerplay Slot Engagement	e.g., powerplay actions close to net
Powerplay Scoring	e.g., powerplay shots and goals
Defensive Puck Control	e.g., dump outs and loose puck recoveries
Defensive Zone Play	e.g., different defensive zone actions
Defensive Positioning	e.g., blocked shots and passes
Slot Defense	e.g., rebounds and dump outs
Penalty Killing	e.g., shorthanded defensive plays
Slot Engagement	e.g., offensive actions close to net
Heavy Game	e.g., body checks and defensive plays
Forechecking	e.g., offensive zone loose puck recoveries
Cycling the Puck	e.g., puck protections and receptions
Neutral Zone	e.g., different neutral zone actions
Puck Moving	e.g, some types of passes, entries
Offensive Zone Play	e.g., different offensive zone events
Shooting	e.g., different types of shots

Table 3: Grid Search Parameters for Gaussian Mixture Models.

Parameter	Values
Initialization Method	K-means++, Kmeans, Hybrid hierarchical, Random, Random from data
Number of components	3-11
Covariance type	Spherical, Tied, Diagonal, Full
Convergence Threshold	10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}
Regularization covariance	10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}
Max iterations	100, 200, 300, 400, 500



(a) Forwards.

(b) Defenders.

Fig. 1: Boxplots of the skill value distributions, as calculated across all seasons.

Fig. 2 shows the average of AIC and BIC values based on the number of components applied to the skill vectors for forwards and defenders, respectively. For this paper, we chose to use the same number of clusters as in [17]. Fig. 2 shows a significant decrease between three and five components for both AIC and BIC, suggesting that adding components to the model improves the performance. As the decrease after five components is smaller, we deem it interesting to work with five clusters as in [17]. As the BIC is smallest for 10 components, we will run the algorithms with 10 components as well in the future.

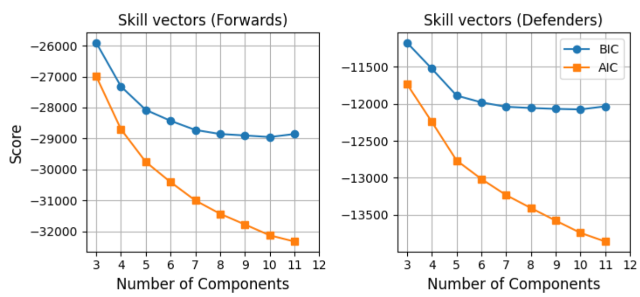


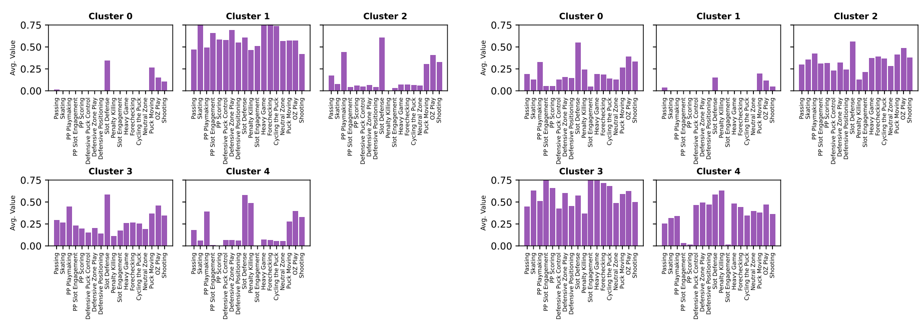
Fig. 2: Average AIC and BIC score by the number of components for forwards (left) and defenders (right) for the dataset containing all seasons.

6 Results

6.1 Forwards

Fig. 3a shows for the five clusters of forwards the average skill values for the ten forwards closest to the centroid of the cluster for the five seasons aggregated. When assigning players to the cluster for which it has highest membership, the players are relatively evenly distributed to the different clusters. To clusters F-all.0, F-all.1, F-all.2, F-all.3, and F-all.4, there are 32, 126, 111, 144, and 120 players assigned, respectively. The forwards in cluster F-all.0 do not excel in any skills, but have there strengths in slot defense, puck moving, OZ play, and shooting. Forwards in cluster F-all.1 are good in many skills and excel in skating, heavy game, forechecking, and cycling the puck. The strengths of forwards in cluster F-all.2 are slot defense, puck moving, OZ play, shooting, and playmaking. Cluster F-all.3 forwards have many skills with strengths in slot defense, puck moving, OZ play, shooting, and PP playmaking. Finally, the forwards in cluster F-all.4 are good at slot defense, penalty killing, OZ play, and PP playmaking, but lack in slot engagement, PP slot engagement, and PP scoring.

We also investigated clusters when we use data for one season. As an example, Fig. 3b shows for the 5 clusters of forwards the average skill values for the ten forwards closest to the centroid of the cluster for season 2019/20. The cluster skill box diagrams for the other seasons are similar. We note that the skill distributions for clusters F-all.0, F-all.1, F,2, and F-all.3 have a clear counterpart in the 2019/20 season (cluster F-all.0 - cluster F-19/20.1, cluster F-all.1 - cluster F-19/20.3, cluster F-all.2 - cluster F-19/20.0, cluster F-all.3 - cluster F-19/20.2). The cluster F-19/20.4 is similar to cluster F-all.4, but the forwards in the cluster have higher skill levels for many of the skills. Example forwards for the clusters are given in Table 4.



(a) All seasons. In the text we call the clusters F-all.0 to F-all.4.

(b) Season 2019/20. In the text we call the clusters F-19/20.0 to F-19/20.4.

Fig. 3: Average skill values of the ten forwards closest to each cluster centroid.

6.2 Defenders

Fig. 4a shows for the 5 clusters of defenders the average skill values for the ten defenders closest to the centroid of the cluster for the 5 seasons aggregated. Cluster D-all.3 has the most players regarding highest membership. Defenders in cluster D-all.0 have low skill levels. The players in this cluster all played few games in the SHL. D-all.1 defenders are highly skilled and excel in defensive skills such as slot defense, stay at home, penalty killing, PK slot defense, physical play, and defensive stickwork. They do not seem to play powerplay. D-all.3 defenders have average skill levels. Defenders in cluster D-all.4 are highly skilled and in

Table 4: Forwards closest to the centroids for the clusters. F-all.0 - F-all.4 for the 5 seasons aggregated. F-19/20.0 - F-19/20.4 for season 2019/20.

Cluster F-all.0 (132 players)	**Cluster F-all.1** (126 players)	**Cluster F-all.2** (111 players)	**Cluster F-all.3** (144 players)	**Cluster F-all.4** (120 players)
Mateusz Szurowski	Simon Ryfors	Jacob Micflikier	Marco Kasper	Mikael Frycklund
Melvin Fernström	Kalle Östman	Peter Holland	Dick Axelsson	Mikkel Boedker
Johan Lundgren	Linus Fröberg	Markus Nenonen	Filip Cederqvist	Juuso Ikonen
William Magnusson	Sebastian Strandberg	Marcus Paulsson	Markus Modigs	Joonas Nattinen
Linus Lööf	Andreas Wingerli	Adam Johnson	Tuomas Kiiskinen	Petrus Palmu
Cluster F-19/20.0 (56 players)	**Cluster F-19/20.1** (49 players)	**Cluster F-19/20.2** (55 players)	**Cluster F-19/20.3** (65 players)	**Cluster F-19/20.4** (50 players)
Rok Ticar	Jesper Kandergård	Jesper Frödén	Emil Pettersson	Joakim Andersson
Gustav Possler	Melker Eriksson	Tuomas Kiiskinen	Brendan Shinnimin	Johan Johnsson
Marcus Paulsson	Alexander Ljungkrantz	Olle Lycksell	Johan Sundström	Adam Pettersson
Viktor Lodin	Linus Hedman	Dominik Bokk	Greg Scott	Axel Wemmenborn
Michael Latta	Samuel Solem	Henrik Törnqvist	Ted Brithén	John Dahlstrom

comparison to D-all.1 defenders excel in offensive skills such as passing, shooting, puck moving, point producing, and they also play powerplay. The players in cluster D-all.2 played few games. In contrast to the players in D-all.0. these players were junior players and some long-time injured players (e.g., Mattias Bäckman in 2019/20).

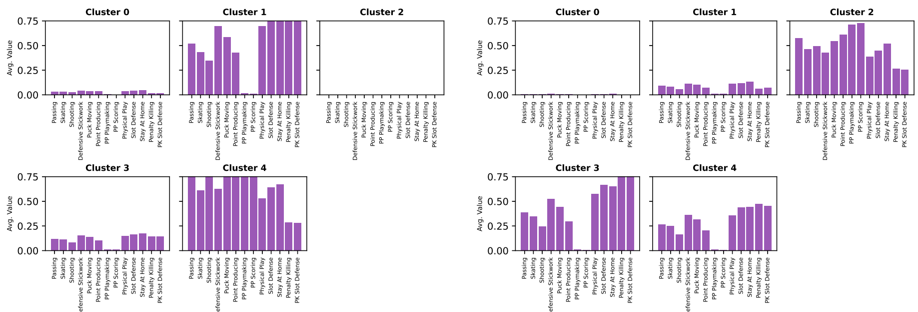
When investigating clusters for different seasons, we observed that clusters similar to D-all.2, D-all.3, and D-all.4 appeared. However, there were no clusters similar to D-all.0. For D-all.1, there was an equivalent cluster for the seasons 2019/20, 2020/21, and 2021/22. For the seasons 2022/23 and 2023/24 cluster D-all.1 could be seen as an aggregation for 2 clusters for the season. Fig. 4b shows for the 5 clusters of defenders the average skill values for the ten defenders closest to the centroid of the cluster for season 2019/20. Example defenders for the clusters using data from all season, and from the 2019/20 season, respectively, are given in Table 5.

Players can change cluster during their career. For instance, Rasmus Rissanen belonged mainly to cluster D-19/20.4 while in 2023/24 he belonged mainly to the cluster that matches D-19/20.3. In this case, his best skills are still in the defensive work, but he has raised the skill level of most of his skills.

7 Conclusion

In this paper, we presented a Gaussian Mixture Model (GMM) approach for characterizing playing styles among ice hockey defenders and forwards. Our method provides a data-driven framework for identifying distinct player types based on skill profiles, offering new insights into player evaluation and team composition.

In future work, we plan to use data from AHL and HockeyAllsvenskan as well (as in [17]) and investigate whether the playing styles are the same or different in the different leagues. Further, we will use the algorithms from [17] on the data used in this paper and compare the different techniques. Such comparisons



(a) All seasons. In the text we call the clusters D-all.0 to D-all.4.

(b) Season 2019/20. In the text we call the clusters D-19/20.0 to D-19/20.4.

Fig. 4: Average skill values of the ten defenders closest to each cluster centroid.

Table 5: Defenders closest to the centroids for the clusters. D-all.0 - D-all.4 for the 5 seasons aggregated. D-19/20.0 - D-19/20.4 for season 2019/20.

Cluster D-all.0 (60 players)	**Cluster D-all.1** (42 players)	**Cluster D-all.2** (67 players)	**Cluster D-all.3** (127 players)	**Cluster D-all.4** (57 players)
Lukas Klok	Anton Mylläri	Albin Thyni Johansson	Daniel Brickley	Matt Caito
Jordan Murray	Oscar Engsund	Nils Strandberg Sarén	Ville Pokka	Oskar Nilsson
Axel Landén	Jonathan Sigalet	Oskar Hassel	Julius Bergman	Lucas Ekeståhl-Jonsson
Theodor Johnsson	Tim Erixon	Jakob Bondesson	Joonas Lyytinen	Kristian Näkyvä
Elias Rosen	Arvid Lundberg	Gustav Berglund	Anton Strålman	Joel Nyström
Cluster D-19/20.0 (18 players)	**Cluster D-19/20.1** (41 players)	**Cluster D-19/20.2** (40 players)	**Cluster D-19/20.3** (29 players)	**Cluster D-19/20.4** (19 players)
Gustav Berglund	Filip Johansson	Jesper Sellgren	Oscar Engsund	Daniel Glad
Jakob Bondesson	Jonas Junland	Marcus Högström	Emil Wahlberg	Jesper Pettersson
Emil Andrae	Lucas Nordsäter	Jonathon Blum	Jonathan Sigalet	Nichlas Torp
Albin Thyni Johansson	Patrik Norén	Oskar Nilsson	Arvid Lundberg	Johan Ivarsson
Christian Lindberg	Julius Bergman	Johan Fransson	Hampus Larsson	Niklas Arell

are expected to provide insights into the relative strengths and weaknesses of different unsupervised learning techniques for player style characterization.

Overall, our findings contribute to the growing research area on quantitative analysis of player behavior, and we hope they will provide tools and foundation for further research into improved player development, scouting, and strategic decision-making in professional ice hockey.

References

1. Akaike, H.: A new look at the statistical model identification. IEEE Transactions on Automatic Control **19**(6), 716–723 (1974). https://doi.org/10.1109/TAC.1974.1100705
2. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer (2006)
3. Chan, T.C., Cho, J.A., Novati, D.C.: Quantifying the contribution of NHL player types to team performance. Interfaces **42**(2), 131–145 (2012). https://doi.org/10.1287/inte.1110.0612
4. Gramacy, R.B., Jensen, S.T., Taddy, M.: Estimating player contribution in hockey with regularized logistic regression. Journal of Quantitative Analysis in Sports **9**(1), 97–111 (2013). https://doi.org/10.1515/jqas-2012-0001
5. Gu, W., Foster, K., Shang, J., Wei, L.: A game-predicting expert system using big data and machine learning. Expert Systems with Applications **130**, 293–305 (2019). https://doi.org/10.1016/j.eswa.2019.04.025
6. Johansson, U., Wilderoth, E., Sattari, A.: How analytics is changing ice hockey. In: Lambrix, P., Carlsson, N., Vernblom, M. (eds.) Proceedings of the Linköping Hockey Analytics Conference LINHAC 2022 Research Track. Linköping Electronic Conference Proceedings, vol. 191, pp. 49–59 (2022). https://doi.org/10.3384/ecp191006
7. Kaplan, E.H., Mongeon, K., Ryan, J.T.: A Markov Model for Hockey: Manpower Differential and Win Probability Added. INFOR Information Systems and Operational Research **52**(2), 39–50 (2014). https://doi.org/10.3138/infor.52.2.39
8. Lambrix, P., Carlsson, N.: Data-driven player performance metrics in ice hockey (2025), forthcoming book chapter.
9. Lambrix, P., Carlsson, N., Säfvenberg, R.: Goal-based performance metrics for ice hockey accounting for goal importance. Journal of Sports Analytics **11** (2025). https://doi.org/10.1177/22150218241290987
10. Lehmus Persson, T., Kozlica, H., Carlsson, N., Lambrix, P.: Prediction of tiers in the ranking of ice hockey players. In: Brefeld, U., Davis, J., Van Haaren, J., Zimmermann, A. (eds.) Machine Learning and Data Mining for Sports Analytics. MLSA 2020. Communications in Computer and Information Science, vol. 1324, pp. 89–100 (2020). https://doi.org/10.1007/978-3-030-64912-8_8
11. Liu, G., Schulte, O.: Deep reinforcement learning in ice hockey for context-aware player evaluation. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. pp. 3442–3448 (2018). https://doi.org/10.24963/ijcai.2018/478
12. Liu, Y., Schulte, O., Li, C.: Model Trees for Identifying Exceptional Players in the NHL and NBA Drafts. In: Brefeld, U., Davis, J., Van Haaren, J., Zimmermann, A. (eds.) Machine Learning and Data Mining for Sports Analytics. MLSA 2018. Lecture Notes in Computer Science, vol. 11330, pp. 93–105 (2019). https://doi.org/10.1007/978-3-030-17274-9_8

13. Ljung, D., Carlsson, N., Lambrix, P.: Player pairs valuation in ice hockey. In: Brefeld, U., Davis, J., Van Haaren, J., Zimmermann, A. (eds.) Machine Learning and Data Mining for Sports Analytics. MLSA 2018. Lecture Notes in Computer Science, vol. 11330, pp. 82–92 (2019). https://doi.org/10.1007/978-3-030-17274-9_7
14. Macdonald, B.: A Regression-Based Adjusted Plus-Minus Statistic for NHL Players. Journal of Quantitative Analysis in Sports **7**(3) (2011). https://doi.org/10.2202/1559-0410.1284
15. McLachlan, G., Peel, D.: Finite Mixture Models. Wiley Series in Probability and Statistics, Wiley (2000)
16. McLachlan, G.J., Lee, S.X., Rathnayake, S.I.: Finite mixture models. Annual Review of Statistics and Its Application **6**, 355–378 (2019). https://doi.org/10.1146/annurev-statistics-030718-104035
17. Olivestam, A., Rosendahl, A., Wilderoth, E., Carlsson, N., Lambrix, P.: Characterizing playing styles for ice hockey players. In: Brecht, T., Carlsson, N., Vernblom, M., Lambrix, P. (eds.) Proceedings of the Linköping Hockey Analytics Conference LINHAC 2023 Research Track. Linköping Electronic Conference Proceedings, vol. 209, p. 39–50 (2024). https://doi.org/10.3384/ecp209004
18. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
19. Pettigrew, S.: Assessing the offensive productivity of NHL players using in-game win probabilities. In: MIT Sloan Sports Analytics Conference (2015)
20. Routley, K., Schulte, O.: A Markov Game Model for Valuing Player Actions in Ice Hockey. In: Meila, M., Heskes, T. (eds.) Uncertainty in Artificial Intelligence. pp. 782–791 (2015)
21. Säfvenberg, R., Carlsson, N., Lambrix, P.: Identifying player roles in ice hockey. In: Brefeld, U., Davis, J., Van Haaren, J., Zimmermann, A. (eds.) Machine Learning and Data Mining for Sports Analytics. MLSA 2023. Communications in Computer and Information Science, vol. 2035, pp. 131–143 (2024). https://doi.org/10.1007/978-3-031-53833-9_11
22. Sans Fuentes, C., Carlsson, N., Lambrix, P.: Player impact measures for scoring in ice hockey. In: MathSport International 2019 Conference. pp. 307–317 (2019)
23. Schuckers, M.: Draft by Numbers: Using Data and Analytics to Improve National Hockey League (NHL) Player Selection. In: MIT Sloan Sports Analytics Conference (2016)
24. Schuckers, M., Curro, J.: Total Hockey Rating (THoR): A comprehensive statistical rating of National Hockey League forwards and defensemen based upon all on-ice events. In: MIT Sloan Sports Analytics Conference (2013)
25. Schulte, O., Khademi, M., Gholami, S., Zhao, Z., Javan, M., Desaulniers, P.: A Markov Game model for valuing actions, locations, and team performance in ice hockey. Data Mining and Knowledge Discovery **31**(6), 1735–1757 (2017). https://doi.org/10.1007/s10618-017-0496-z
26. Schulte, O., Zhao, Z., Javan, M., Desaulniers, P.: Apples-to-apples: Clustering and Ranking NHL Players Using Location Information and Scoring Impact. In: MIT Sloan Sports Analytics Conference (2017)
27. Schwarz, G.: Estimating the Dimension of a Model. The Annals of Statistics **6**(2), 461 – 464 (1978). https://doi.org/10.1214/aos/1176344136

28. Steele, R.J., Raftery, A.E.: Performance of bayesian model selection criteria for gaussian mixture models. University of Washington, Department of Statistics, Technical Report (570) (2009), https://stat.uw.edu/research/tech-reports/performance-bayesian-model-selection-criteria-gaussian-mixture-models
29. Thomas, A., Ventura, S.L., Jensen, S., Ma, S.: Competing Process Hazard Function Models for Player Ratings in Ice Hockey. The Annals of Applied Statistics **7**(3), 1497–1524 (2013). https://doi.org/10.1214/13-AOAS646
30. Vincent, C.B., Eastman, B.: Defining the style of play in the nhl: An application of cluster analysis. Journal of Quantitative Analysis in Sports **5**(1) (2009). https://doi.org/10.2202/1559-0410.1133