A Machine Learning Approach to Score Passes Based on Their Offensive Contribution

Kiran Roye

Ossining High School, Ossining NY 10562, USA

1 Introduction

Passing plays a critical role in facilitating high-level ice hockey performance. Effective passing enables teams to transition from defense to offense, evade pressure, and generate scoring opportunities [2,4]. Previous research has largely evaluated passes based on discrete rink regions, often characterizing passes solely by the origin and reception zones [5,3].

This study introduces a machine learning model that characterizes passes based on a broader set of continuous features, including the precise coordinates of the pass origin and reception, pass duration, and estimated velocity. Using these features, the model predicts the expected goals (xG) associated with a shot taken immediately following the reception [6]. By identifying passing patterns that historically lead to high-danger opportunities, the model offers a data-driven tool for optimizing offensive strategy [1].

2 Methodology

2.1 Data Processing

The dataset employed was a proprietary event dataset from Sportlogiq, encompassing 156 games from the Swedish Hockey League (SHL). Each successful pass was assigned a unique ID, with the corresponding reception and subsequent events by the receiving player sharing this identifier. To construct the training dataset, the data was filtered to include only sequences where the reception led directly to a shot attempt.

2.2 Model Development

A supervised learning approach was implemented using a feedforward neural network developed in R with the `keras` package. The target variable was the xG value of a shot immediately following the pass. By restricting the dataset to plays where a reception was followed by a shot, the model directly linked pass quality to shot outcome.

The feature set included both continuous and categorical variables. Continuous features comprised the (x, y) coordinates of the pass origin and reception, the elapsed time between pass and reception, the Euclidean distance of the pass, and the estimated velocity. These features were normalized using z-score standardization. Categorical features, including reception type and manpower situation, were one-hot encoded.

The network architecture consisted of an input normalization layer followed by dense layers of 32, 16, 8, 4, and 2 units respectively, each using ReLU activation, and a final sigmoid-activated output layer predicting xG values in the $[0, 1]$ range. The model was trained using the RMSProp optimizer (learning rate = 0.0005) with mean squared error as the loss function. Training occurred over 32 epochs with a 10% validation split and batch shuffling, promoting generalization and mitigating overfitting.

This modeling approach allows for a nuanced quantification of pass quality based on spatial, temporal, and contextual attributes beyond simple zone-based metrics.

Player-Level Analysis of Passing Expected Goals (PxG)

To evaluate the model's practical utility, it was applied to a broader dataset of completed passes (n = 91,558), not restricted to those immediately leading to a shot. A player-level analysis was conducted to assess the relationship between PxG and assists per game.

Players who had participated in at least five games were selected. For each player, the total PxG generated, the total number of assists, and the number of games played were calculated. PxG per game and assists per game were then derived, and their relationship was evaluated using a Pearson correlation.

3 Results

3.1 Evaluating Efficacy of PxG

The neural network model for expected goals from passes (PxG) demonstrated strong convergence during the training process. Figure 1 illustrates the progression of both MSE and Mean Absolute Error (MAE) metrics over the 32 training epochs for both training and validation datasets. The model achieved optimal performance in the final epoch with training MSE of 0.002119 and MAE of 0.02626, while validation metrics showed slightly superior performance with validation MSE of 0.001713 and validation MAE of 0.02368. This marginal improvement in validation metrics compared to training metrics suggests the model generalized effectively to unseen data without overfitting. The training history visualization reveals a consistent downward trend in both metrics, with the rate of improvement diminishing after approximately the 16th epoch, indicating the model approached its optimal state. The close tracking between training and validation curves throughout the training process further confirms the model's robust generalization capabilities. The low absolute values of the final metrics—particularly the MAE of approximately 0.024 for the validation set—indicate that the model's predictions deviate minimally from the ground truth values.

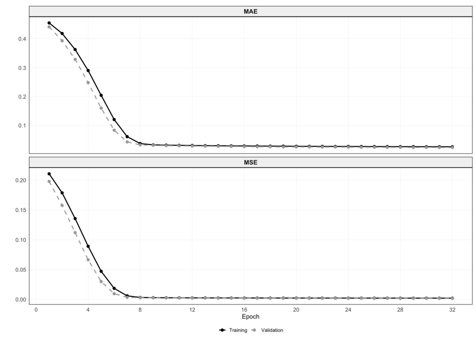

Fig. 1. Training and validation metrics for the xG prediction model over 32 epochs. Solid black lines represent training metrics; dashed gray lines represent validation metrics. Convergence occurs by approximately 20 epochs.

Spatial Characterization of PxG

Further analysis involved projecting a stratified random sample of 2,000 passes onto a scaled hockey rink diagram (Figure 2). Higher PxG values were assigned to passes originating near the perimeter and received near the slot area, corresponding to traditional high-danger scoring zones. The model also identified traditional "cycle" passing patterns around the slot as yielding higher PxG values.

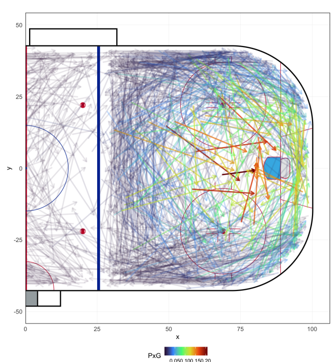

Fig. 2. Visualization of a random sample of passes. Each vector is colored based on its PxG value.

Pass characteristic analysis (Figure 3) revealed that pass velocity exhibited the strongest relationship with PxG ($\rho = 0.21$), though the association was modest. Longitudinal movement showed a weak positive correlation ($\rho = 0.081$), lateral movement showed a slight negative correlation (ρ = -0.097), and pass angle exhibited essentially no correlation with PxG (ρ = -0.013). Although all relationships were statistically significant, most practical effects were limited.

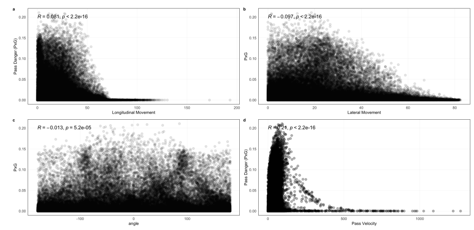

Fig. 3. Scatter plots showing Spearman correlations between PxG and different pass characteristics.

Relationship Between PxG and Assists

Analysis of PxG per game and assists per game showed a strong positive correlation, with a linear model explaining 49% of the variance ($R^2 = 0.49$). This indicates that PxG successfully captures key aspects of playmaking ability. Players with higher PxG values per game consistently produced more assists, reinforcing the utility of PxG for player evaluation and strategic planning. Points in Figure 4 are scaled based on games played, giving more visual weight to larger sample sizes.

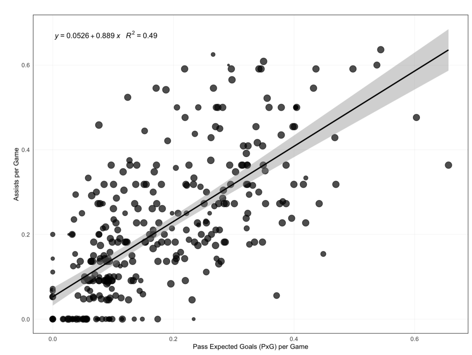

Fig. 4. Relationship between PxG per game and assists per game. Point size reflects games played; line represents the fitted regression.

4 Discussion

This study demonstrates the potential of using machine learning to assess the quality of offensive passes in ice hockey. By considering factors such as pass location, velocity, and time, along with contextual features like manpower situation, the model provides a more comprehensive evaluation of passes than traditional zone-based metrics. The strong correlation between Passing Expected Goals (PxG) and assists per game suggests that PxG effectively captures the value of a player's passing, making it a reliable metric for assessing playmaking ability and identifying high-danger offensive opportunities.

However, there are several areas for improvement. The current model relies on event-level data and does not incorporate the dynamic movements of players during a pass. Including player tracking data would allow for a more granular understanding of how player positioning affects pass effectiveness, potentially improving the model's predictions. Future work could also integrate additional features such as shot types and angles, or explore more sophisticated models like recurrent neural networks, to capture temporal patterns in passing sequences. Overall, while the model provides valuable insights into offensive playmaking, incorporating richer, real-time data would enhance its accuracy and broaden its application to various hockey contexts.

5 Code Access Link

Code: github.com/KingKobra7899/KROYE_LINHAC_2025

References

1. Fenwick, M.: Analytics in Ice Hockey. Hockey Graphs Press (2016)
2. MacKenzie, R., Cushion, C.: Passing and its role in ice hockey offense. International Journal of Sports Science & Coaching **15**(3), 345–357 (2020)
3. Schuckers, M., Curro, R.: Openwar for hockey: Evaluating play sequences. `http://www.statsportsconsulting.com/publications.html` (2017)
4. Schulte, O.: Offensive strategies in professional ice hockey. Journal of Quantitative Analysis in Sports **14**(4), 169–178 (2018)
5. Sporar, J., et al.: Spatial analysis of passing patterns in professional ice hockey. In: Proceedings of the MIT Sloan Sports Analytics Conference (2019)
6. Thomas, A.C.: Expected goals in hockey. `https://war-on-ice.com/xG.html` (2017)