Bachelor Computer Science & Economics

Analysis of the

Penalty Corner in Hockey

Eva Knoester

Dr. A.J. Knobbe Dr. AW de Leeuw

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

04/06/2025

Contents

1 Abstract

Sports data science is increasingly applied to tactical analysis in elite sports, yet field hockey remains underrepresented in this domain. This thesis contributes to closing that gap by focusing on one of the sport's most decisive moments: the penalty corner. Using video footage from the Dutch national field hockey team, this study combines computer vision and subgroup discovery to extract and analyze player movements, formations, and ball trajectories during these set plays.

Through a multi-stage pipeline of field detection, player tracking using YOLO and Deep SORT, and ball tracking via motion-based filtering and trajectory prediction, structured positional data was obtained from static top-down camera recordings. Tactical features were then derived at both the individual and team level across key moments in the penalty corner. These include attacker compactness, defender dispersion, line-stopper positioning, and flyer angles. The resulting dataset was analyzed using SubDisc, a subgroup discovery tool, to extract interpretable tactical rules correlated with successful or failed penalty corner outcomes.

The analysis revealed that dynamic defensive play, particularly proactive flyer runs in a slightly diagonal line and compact late-stage formations, was most associated with penalty corner success. On the attacking side, structured movement towards the goal, fast ball injection delivered just right of the central axis of the circle, and coordinated compact movement were key to scoring. Field element detection achieved near perfect accuracy, whereas ball tracking still leaves room for improvement, potentially through future deep learning approaches. Although the detection and tracking pipelines occasionally faced challenges such as occlusion and player misidentification, the overall methodology provides a reasonable automated and interpretable framework for tactical analysis in field hockey. This research provides a foundation for future applications of data science in hockey, both to support coaching decisions and to guide the development of smarter, video-based training tools.

2 Introduction

2.1 Background

Hockey is a sport characterized by a wide range of tactical perspectives and strategies. Although these approaches may vary, they all share the same goal: to win the game by scoring more goals than the opponent. Understanding and analyzing such tactical decisions can offer valuable insight into both offensive and defensive game strategies. As noted by Garganta [Gar09], identifying patterns in team behavior can support coaches in making more informed evidence-based decisions during matches.

Despite its growing popularity, field hockey still lacks the data, knowledge, and financial resources commonly found in other professional sports. As a result, the sport has seen relatively little progress in terms of data-driven tactical and behavioral analysis. Hence, this study addresses this gap by focusing on one of the most impactful moments in the game, namely the penalty corner.

The penalty corner is a strategic scoring opportunity, accounting for approximately 34% of all goals scored by the Dutch national team during their last 248 matches [dH24]. This illustrates the impact a well-executed penalty corner can have on professional hockey matches. Investigating the different approaches in a penalty corner could help identify certain patterns. By optimizing these patterns, the penalty corner execution could be significantly enhanced to increase scoring success.

For professional football, for instance, studies have demonstrated that winning football teams generate more scoring chances than their opponent [Hea19]. This suggests that optimizing the penalty corner could improve the performance and success of the Dutch hockey team.

This research investigates how various tactical approaches during a penalty corner influence the likelihood of scoring. Using match footage provided by the Royal Dutch Hockey Federation (KNHB), this study applies computer vision and machine learning to extract and engineer features from video data. These features include both attacking strategies and the defensive responses of the opposing team. By modeling interactions during a penalty corner, this study aims valuate the success of various tactical approaches. The goal is to determine which strategies are most effective for both attacking and defending teams.

2.2 Research Scope

This study focuses on analyzing penalty corners executed by the Dutch national field hockey team, using video footage provided by the KNHB. The scope is limited to corners captured from a top-down static camera angle, allowing for precise tracking of player and ball movements. The analysis includes both attacking and defensive strategies, considering initial formations, player roles and movements, ball trajectory and much more information.

Only penalty corners with a clear outcome are included in the dataset. Cases involving defensive fouls that lead to a re-take or ambiguous results have been excluded to maintain clarity and improve model reliability. The flowchart 1 below shows the possible outcomes of the penalty corner based on the rules of [(FI23]. This flowchart also demonstrates which outcomes are included and which cases are excluded to maintain a binary and robust target.

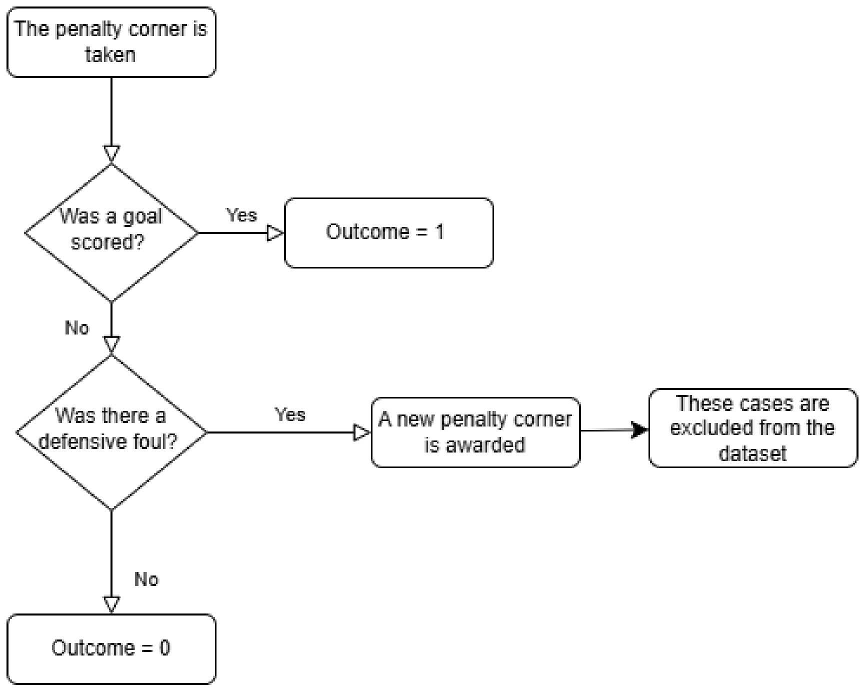


Figure 1: Flowchart showing all possible outcomes of a penalty corner and which cases are excluded for this research

2.3 Research goals

The primary goal of this research is to determine which approach leads to the most successful penalty corner outcomes, for both attacking and defending teams. This will be achieved by developing a machine learning model trained on features extracted from video data. In collaboration with the KNHB (Dutch Hockey Federation), a research question has been thought out. The research question that this paper will try to answer is:

"What is the most effective approach to maximize scoring, the aim of the attacking team, and minimize goals conceded, the objective of the defending team, during a penalty corner in field hockey?"

This question can be subdivided into sub-questions such as, *"Which individual player features have the most impact when trying to optimize the penalty corner?"*, *"What is the most effective offensive and defensive team behavior during the penalty corner?"* and *"What are the worst offensive and defensive actions during the penalty corner?"*.

The answers to these sub-questions will be compared with existing insights from hockey strategy literature to identify patterns and provide practical recommendations for coaches. Ultimately, these insights will help structure the research and contribute to identifying the most effective penalty corner strategies.

2.4 Structure

This thesis is built up in multiple sections, starting with Chapter 2 that reviews related work in the areas of sports analytics, computer vision, and machine learning. The focus will be on prior research related to background subtraction, player and ball tracking. In Chapter 3, we introduce the dataset provided by the KNHB and describe how the data was collected by different computer vision techniques. This section explains how the raw video material was transformed into structured input suitable for analysis. Chapter 4 outlines the design of the machine learning pipeline, including data preprocessing, feature engineering, and the classification approach. In Chapter 5, we present the experimental setup, evaluation metrics, and results, followed by a discussion of the model's performance in collecting data and identifying effective strategies for both attacking and defending teams. Finally, Chapter 6 summarizes the main findings, discusses the practical implications for coaches and analysts, and suggests directions for future research in tactical decision-making and data-driven hockey optimization.

3 Related Work

3.1 Background subtraction

Background subtraction plays a crucial role in segmenting moving objects from static backgrounds, enabling accurate object detection and tracking in dynamic environments [TKBM99]. There have been many developments in different kind of approaches to subtract the foreground scenes from the background. A basic background subtraction approach operates on the assumption that motion occurs when the color or intensity of a pixel changes significantly over time [BJE+09]. While individual pixel changes are used as indicators of movement, further processing is typically required to group these pixels together to detect objects. However, since individual pixel changes do not necessarily correspond to complete objects, additional processing, such as optical flow and connected component analysis, is required to group related pixels into coherent object-level segments. This can be summarized in the following equation [SHJ15]:

$$X_t(s) = \begin{cases} 1 & \text{if } d(I_{s,t}, B_s) > \tau, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Here, X_t represents the movement label at time t (also referred to as the motion mask), d denotes the distance between $I_{s,t}$, the video frame at time t for pixel s, and B_s, the background model at pixel s; τ is a predefined threshold. The primary distinction among background subtraction (BS) methods lies in how the background B is modeled and the specific distance metric d that is used. The following paragraphs present an overview of various BS techniques and their respective approaches.

By comparing the 'non-moving' pixels with the pixels 'in-motion', one can filter out the irrelevant background model against the foreground objects of interest to primarily focus on these. Unfortunately, a video often contains noise which complicates the method of using a global threshold. Therefore, various methods have been developed to improve the robustness of background subtraction under changing lighting conditions, object occlusions, and noise. A couple of the most well-known algorithms have been evaluated by Sobral and Vacavant [SV13]. They concluded that their best methods were quite time and memory consuming and therefore suggested that Gaussian Mixture Model scored relatively well with some others. Furthermore, they argued that background subtraction is mainly for static background, which the videos of this research contain and thus benefit from.

Higham et al. [Hig16] conducted a comparative analysis of background subtraction algorithms and applied Particle Swarm Optimization (PSO) to determine the optimal algorithm for video processing in EuroHockey 2015. Their study evaluated multiple background subtraction techniques, including frame differencing, Gaussian Mixture Models (GMM), Dominant Color subtraction and statistical background models. Their research indicated that the Temporal Median and the Gaussian Mixture Model outperformed the other models, especially a combination with Frame Difference is recommended. Ahmed et al. [AAAJ21] on the other hand, he demonstrated the effectiveness of integrating Deep Learning-based object detection with traditional background subtraction techniques.

In summary, background subtraction remains a fundamental component of sports video analysis, with ongoing advancements in adaptive modeling, optimization techniques, and deep learning integration. Therefore, different approaches have been applied in this paper to investigate which models delivered the best result based on the different kind of relevant foreground

3.2 Player detection and tracking

Traditional Methods for Player Detection
Player detection is a fundamental task in sports analytics and computer vision, enabling automated tracking, event recognition, and performance analysis. The primary challenge in player detection lies in occlusions, varying poses, lighting changes, and motion blur, all of which can affect the robustness of detection algorithms. Various approaches have been developed over time, ranging from traditional image processing techniques to modern deep learning-based object detection frameworks.

Early approaches to player detection relied on background subtraction, color segmentation, and edge detection to separate players from the field. For instance, Darrel et al. [DGHW00] proposed a multi-modal approach integrating skin detection, face detection, and disparity maps from stereo vision to identify and track individuals. While effective in controlled settings, their method struggled with lighting variations due to its reliance on a predefined color model.

McKenna et al. [McK00] introduced a multi-level detection system, where players were detected using chromaticity based segmentation and gradient information.They demonstrated that color histograms and silhouette extraction could effectively differentiate people from the background, even in complex environments. Another notable contribution came from Harville [Har04] who developed a plan-view mapping system to detect multiple people using height maps and occupancy grids. While effective for surveillance applications, these methods often struggled with false positives when detecting non-human objects with similar dimensions.

Another commonly used approach in early research is Histogram of Oriented Gradients (HOG) combined with Support Vector Machines (SVM). Introduced by Dalal and Triggs [DT05], this approach demonstrated its effectiveness in pedestrian detection. Mun õz Salinas et al. [MSAGS06] further improved detection by incorporating stereo vision and background modeling, showing that depth information significantly enhanced accuracy. Although HOG+SVM was a foundational step in object detection, it was later surpassed by deep learning methods such as YOLO (You Only Look Once) and Faster R-CNN, which provided faster and more robust detection under real-world conditions. This shift was essential for applications requiring real-time tracking, such as player detection in sports analytics.

Deep Learning-Based Player Detection
Deep learning has revolutionized computer vision by enabling automatic feature extraction and hierarchical learning for object detection and tracking. Traditional methods relied on handcrafted features such as HOG (Histogram of Oriented Gradients), SIFT (Scale-Invariant Feature Transform), and Haar-like features. However, these approaches struggled with lighting changes, occlusions, and complex backgrounds [OW13]. With the rise of Convolutional Neural Networks (CNNs), deep learning models have surpassed these traditional approaches by learning feature representations

directly from raw images. CNNs are highly adaptable and generalizable, making them ideal for real-time player detection in sports analytics [GLO16].

The use of deep learning for player detection has therefore been widely explored in recent years. Ouyang and Wang [OW13] introduced a joint deep learning framework for pedestrian detection, which optimized feature extraction, occlusion handling, and classification. This approach significantly improved detection rates by allowing interactions between different feature components. Similarly, Guo et al. [GLO16] reviewed how deep learning approaches such as Autoencoders, Restricted Boltzmann Machines (RBMs), and CNNs have been applied to computer vision tasks, emphasizing their superiority in object detection. Zhao et al. [Zha19] provided an extensive overview of deep learning-based detection models like R-CNN, Faster R-CNN, SSD (Single Shot MultiBox Detector), and YOLO (You Only Look Once), highlighting their improvements in accuracy and efficiency.

One of the most widely used deep learning models for real-time player detection is YOLO. Unlike traditional region-based models that require multiple stages, YOLO performs detection in a single forward pass, making it significantly faster. Ahmed et al. [AAAJ21] demonstrated how YOLOv3 combined with Deep SORT tracking achieved real-time multi-player tracking in different environments. In this setup, YOLO detects players, while SORT (Simple Online and Realtime Tracker) ensures identity consistency across frames.

YOLO and Deep SORT

The combination of YOLO and Deep SORT has proven effective in various tracking applications. Punn et al. [PSAR21] used YOLOv3 with Deep SORT to monitor social distancing violations, where YOLO detected people in video streams, and Deep SORT assigned a unique ID to each person, tracking their movement. Similarly, Meimetis et al. [MDPH23] modified Deep SORT to track vehicles and pedestrians in real-time traffic monitoring.

In sports analytics, Buric et al. [BIKP19] showed how integrating YOLO with tracking-by-detection methods like SORT and Deep SORT significantly improves player tracking accuracy. Their research, which focused on handball player tracking, found that Deep SORT enhances tracking performance by handling motion prediction and re-identifying players after occlusions.

This combination of YOLO and Deep SORT is directly applicable to this research in tracking field hockey players during penalty execution. By using YOLO for player detection and Deep SORT for maintaining consistent player identities, this study ensures accurate motion tracking, player positioning assessment, and movement analysis. Even under challenging conditions, such as rapid direction changes, occlusions, and varying backgrounds, deep learning-based tracking ensures accurate evaluation of penalty execution and strategic play in field hockey.

Despite these advancements, no automated video-based tracking system currently exists for field hockey penalty analysis. Hockey penalties are still analyzed manually, making the process time-consuming, subjective, and inconsistent. This research aims to bridge that gap by applying YOLO and Deep SORT for player tracking, enabling automated movement analysis to distinguish team strategies during penalty execution. This system ensures accurate, unbiased tracking, provides objective performance metrics, and offers scalability for broader sports analytics applications. By leveraging YOLO for player detection and Deep SORT for identity tracking, this study is able to handle rapid movements, occlusions, and varying backgrounds. This system will enhance performance evaluation, coaching strategies, and field hockey analysis.

3.3 Ball detection and tracking

Tracking fast-moving objects and predicting movement is crucial in sports analytics. Balls, players, and other objects move quickly, change direction frequently, and get occluded. To track and predict their movement, researchers have developed various methods based on speed, angle, and trajectory analysis.

Early methods for detecting fast motion in sports relied on frame differencing, optical flow, and trajectory based models. Yu et al. [YLXT07] introduced a trajectory based ball detection system for soccer. Their method tracked candidate ball positions across frames, helping to eliminate false positives. The study showed that motion consistency over time is key to distinguishing true movement from background noise or occlusions. Similarly, Zhang et al. [ZXT10] applied stereovision and velocity estimation to track a fast-moving table tennis ball. Their system predicted landing and striking points based on the ball's speed, making it useful for dynamic sports like field hockey.

For ball detection in handball, Buric et al. [BIKP19] compared YOLO and Mask R-CNN for fast-moving object tracking. Their results showed that YOLO is faster, while Mask R-CNN provides more precise localization. The study emphasized the importance of combining speed and acceleration data with visual features like color and shape for better accuracy. Predicting movement is just as important as tracking it. Chakraborty and Meher [CM13] developed a trajectory interpolation method to estimate missing ball positions in basketball videos. Their approach helped track balls even when they were temporarily hidden or occluded.

This research applies similar methods to field hockey penalty tracking. By using motion modeling, speed estimation, color and size thresholds, and angle tracking, it ensures accurate and automated analysis. This approach helps overcome the limitations of manual video review. Future improvements could include deep learning models to further enhance trajectory prediction and shot classification.

4 Methodology

In this chapter, we describe the methodology used to automatically analyze and classify offensive and defensive strategies during field hockey penalty corners. The analysis is based on video data provided by the KNHB, recorded from a static top-down camera angle as shown in Picture 2. From these recordings, raw video material is transformed into structured data through a series of computer vision and feature extraction techniques. Each penalty corner was manually reviewed to determine its outcome based on the referee's decision and the final result. The dataset originally contained around 200 penalty corners, but some were excluded due to ambiguous outcomes. In certain cases, defensive rule violations led to a retaken corner, making it difficult to assign a clear result. Since model training is more effective with binary targets, these instances were omitted. After this filtering process, the final dataset consists of 97 corners, each labeled as either a goal (target = 1) or no goal (target = 0).

Picture 2: Top Static View of the Penalty corner

These videos are used to create a dataset that provides valuable information for Machine Learning models to recognize patterns. The data is extracted using various visual approaches and models, which will be discussed in this chapter. The methodology is divided into three main parts. First, we describe the preprocessing of video data and the detection pipelines used to extract field elements, ball movement, and individual players. This includes the implementation of various computer vision techniques such as object detection, background subtraction, and edge detection. Second, we outline how tactical features are derived from the detected elements, covering both player-level and team-level metrics for offensive and defensive play. Lastly, we present the subgroup discovery framework used to identify statistically significant patterns associated with penalty corner success or failure. This process enables the extraction of interpretable tactical rules that support the final recommendations. This entire process is visualized in Figure 3.

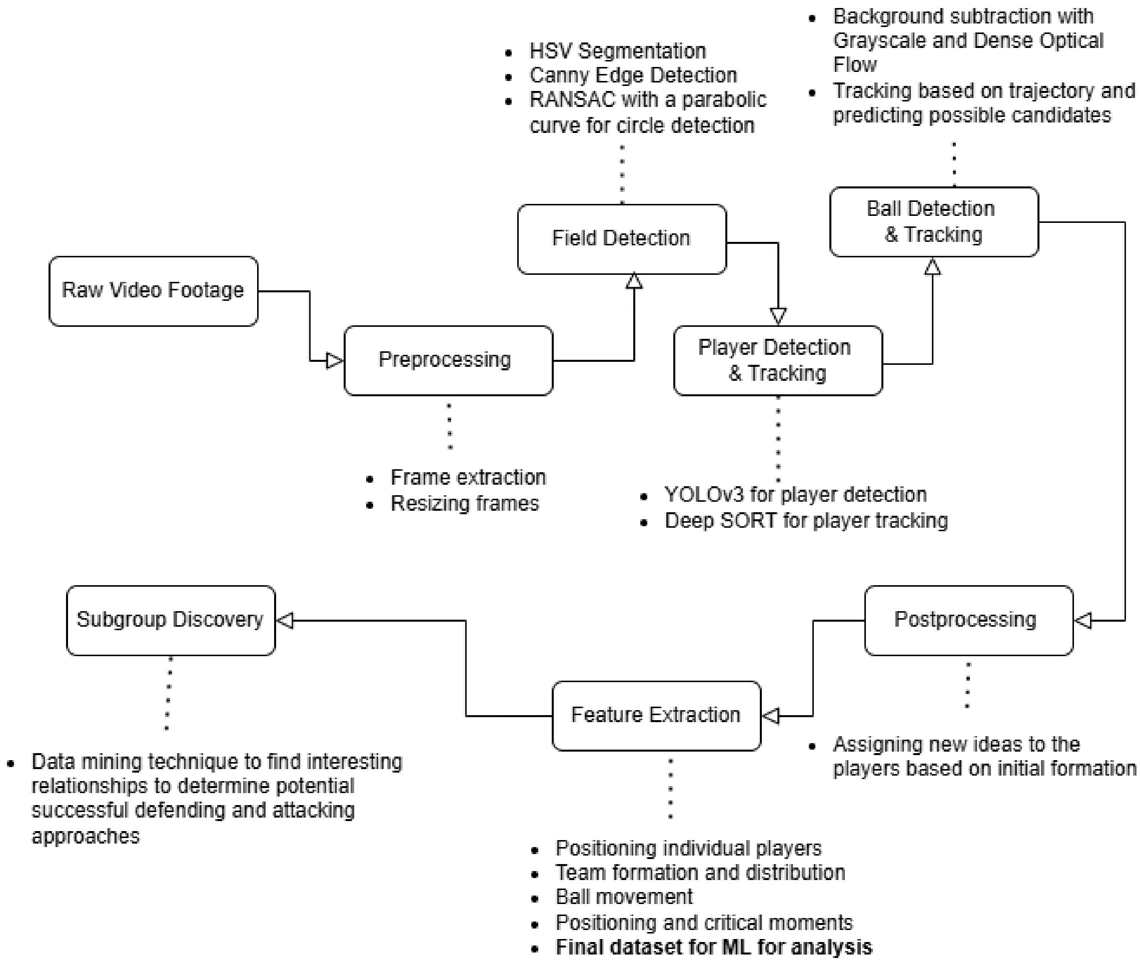


Figure 3: Overview of the full computer vision pipeline from video input to subgroup discovery used to analyze penalty corners

4.1 Video Preprocessing and Visual Detection

This subsection outlines the computer vision techniques used to extract raw positional and visual data from the video recordings of penalty corners. Elements such as field lines, ball, and players are identified and tracked frame-by-frame by using a combination of color space filtering, edge detection, object tracking, and geometric transformations. These preprocessing steps form the foundation of the entire analysis pipeline, enabling accurate detection of key events and formations during the execution of a penalty corner.

4.1.1 Field element detection

Circle detection

To analyze the spatial structure of the penalty corner setup, a custom script was developed to detect the top of the circle using a parabolic fitting method. This process begins with a color-based segmentation technique, where white field markings are extracted using thresholding in the HSV (Hue, Saturation, Value) color space. Unlike the RGB (Red, Green, Blue) color model, which directly represents color using light intensities as discussed in the next section, HSV separates the chromatic content (hue), the saturation (color intensity), and the brightness (value) of each pixel [SQP02].

This separation makes HSV particularly effective for detecting white lines on the pitch, as it allows the selection of pixels with low saturation and high brightness values, which are characteristic of white regions [KYYW22]. Moreover, the HSV model is more robust to variations in lighting and shadow, ensuring that slightly discolored or shaded white lines are still correctly identified as part of the field markings. This isolates the white lines commonly found on the hockey pitch from the background. Additionally, a background mask is applied to filter out non-relevant areas of the field, such as the top border and the lower 30% of the frame.

Once irrelevant regions are removed, the remaining foreground is further processed through Gaussian blurring, edge detection and morphological operations to reduce noise and enhance continuous edges. Edge detection is performed using the Canny algorithm which was introduced by Canny in 1983 [Can86]. This algorithm finds boundaries or edges by detecting areas where the intensity of an image changes sharply. These are often the outlines of objects, as portrayed in Figure 4. After detecting edges, the result may contain broken or disconnected lines especially due to noise or imperfect lighting. Morphological closing is therefore applied to "seal" small gaps and link edge fragments together, helping form continuous contours.

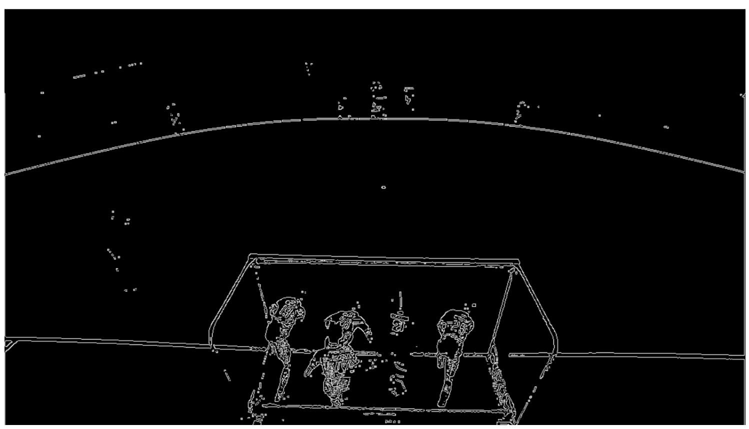


Figure 4: Edge detection performed using the Canny algorithm to find continuous edges and boundaries

From the resulting edge image, contours are extracted and converted into a set of (x, y) points representing line segments in the scene together with the applied top and bottom masks as depicted in Figure 5a. These points are sorted and passed into a RANSAC (Random Sample Consensus) regression model to fit a second-degree polynomial (quadratic curve), shown in Figure 5b. RANSAC is selected for its robustness to outliers and partial occlusions, which are common in real-world sports footage. Unlike conventional methods that use the full dataset and then remove outliers, RANSAC begins by selecting the minimum number of points needed to estimate the model parameters and then incrementally expands this set by including only consistent data points [FB81]. This sampling-based approach makes it particularly well-suited for noisy environments [Der10]. The resulting model estimates the parameters a, b, and c of the quadratic function $y = ax^2 + bx + c$, which defines the top arc of the penalty circle.

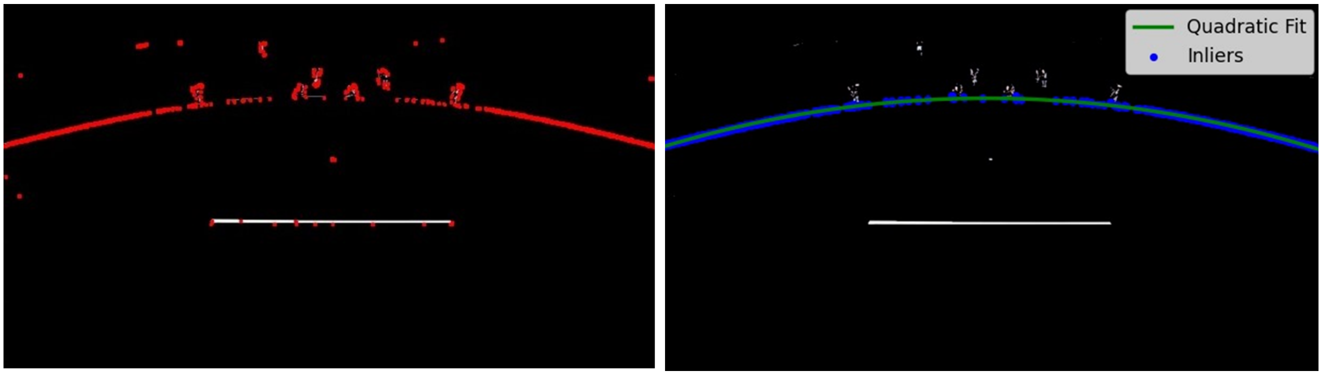


Figure 5: Visual pipeline for detecting the top of the circle. Figure a shows the extracted contour points (in red) after HSV-based thresholding and masking. Figure b illustrates the RANSAC regression fit (green curve) applied with identified inlier points (blue) to approximate the circle arc.

To improve robustness, the script fits parabolas over several frames (e.g., frames 3 to 5), and the resulting coefficients are averaged to obtain a stable estimate of the circle's curvature as shown in Figure 6. In later frames, this average curve is drawn onto the video, providing a clear and consistent reference for the top of the circle. This reference is essential for defining critical tactical zones and for calculating player positioning and approach angles relative to the top of the circle.

Figure 6: Visual representation of the circle detection process, showing the result after edge detection, contour extraction, and quadratic curve fitting using RANSAC.

Baseline Detection

Following the circle detection approach, the baseline of the hockey field is identified using a similar combination of color-based segmentation, edge detection, and post-processing techniques. As before, the white field markings are extracted in the HSV color space, and a background mask is applied to filter out irrelevant regions of the frame, particularly the top corners and upper border.

Once masking is complete, the image is converted to grayscale and processed using the Canny edge detector. The resulting edge map is passed through the probabilistic Hough Line Transform to extract straight line segments which idea was originally introduced by Hough in 1962 [Hou62]. To isolate the baseline specifically, detected lines are filtered based on orientation, favoring horizontal vectors, and their $y-$position within the frame. Only those with an angle close to zero degrees, a minimum length of 60% of the frame width, and located within the bottom 25% of the image are retained.

Compared to the circle, the baseline is more frequently fragmented due to visual obstructions such as defenders standing near or over the line, as well as the presence of the goal. To overcome this, a line merging algorithm is applied that combines short, nearly parallel, and spatially close segments into longer, continuous lines. This is based on both angular similarity and spatial proximity.

The result is a robust and interpretable baseline representation, which serves as a spatial anchor for calculating relative player positions and distances. Figure 7 shows an example of the fitted baseline. In this image, multiple red lines are detected due to visual gaps or player movement along the baseline. From these segments, the algorithm calculates an average slope and extracts representative $x-, y-$ coordinates to reconstruct a stable geometric baseline. This estimated baseline provides a reliable foundation for analyzing player movements and overall team strategies.

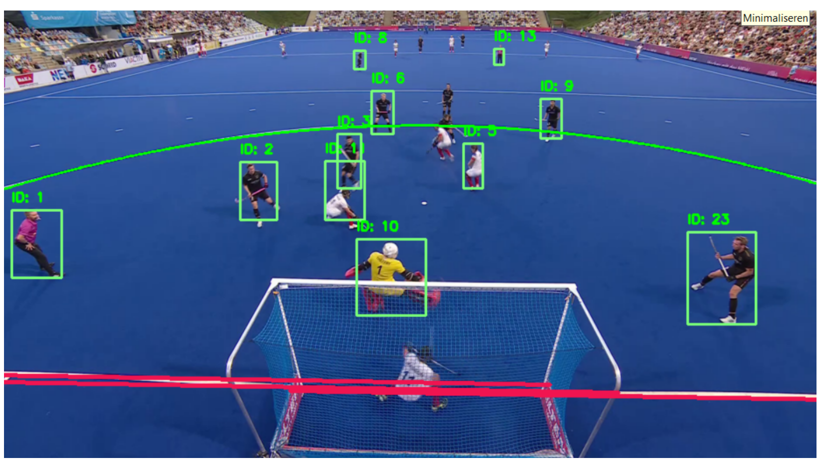


Figure 7: Example of detected baseline (red) and top of the circle (green) overlaid on the original frame. The baseline is reconstructed by averaging the orientation and location of detected line segments.

4.1.2 Ball Detection Pipeline

To accurately detect the ball during a penalty corner, it is essential to first reduce visual noise and exclude irrelevant parts of the frame. The first step focuses on removing fixed, non-relevant regions of the video frame which contains the crowd. This is done by applying a custom mask over two triangular areas in the top corners and a rectangular region at the top of the frame. This mask is applied before any analysis, ensuring that distractions and false positives from static areas are eliminated early in the process. This mask is applied for across the ball, player and team data, and is implemented before any analysis. This ensures that distractions and false positives from static areas with people and movements are eliminated early in the process.

After the static regions are masked, a mask is applied that focuses on detecting motion between frames. This study starts with applying a color conversion code that converts RGB images (Red-Green-Blue frames) into a grayscale image with Grayscale conversion. This saves a lot of computational effect and enhances motion accuracy as a RGB image is represented by:

$$I_{\text{RGB}} = (F_R, F_G, F_B) \tag{2}$$

Where $F_R(x, y)$ is the intensity of the pixel (x,y) in the red channel, $F_G(x, y)$ in the green channel, and $F_B(x, y)$ in the blue channel, while a grayscale image only has only one dimension and intensity $F(x, y)$ [KV10]. The Grayscale approach takes the luminance of each pixel using a weighted formula that reflects how humans perceive brightness [Cad09]. On top of that, a histogram equalization is implemented to enhance contrast and make small movement stand out more [ea87].

After these masks are implemented, a binary threshold is employed to isolate areas with significant motion in between frames. This is further refined by computing optical flow to detect fast-moving pixels, which are likely to belong to the ball. The optical flow is based on Dense Optical Flow which calculates motion vectors for every pixel to characterize the change of structures in the image plane [ea00]. By combining these steps, the script filters static objects and slow-moving elements out, such as slow walking players or gradual lighting changes.

The next step is to filter the detected motion blobs to isolate only those that are likely to represent the ball. Blobs are connected components, pixels, that are considerably denser than the surrounding background [DMZ11]. Since many moving blobs can appear in the frame, further modification is necessary. This is done using a combination of a visual characteristic, based on size, and motion trajectory. The script starts with filtering out potential candidates by suggesting that the ball blob must fall within a predefined area range of pixels. On top of that, for the initial detection, a condition is applied requiring the ball to be located in the lower right corner of the frame.

After the initial visual filtering, motion patterns are used to further narrow down the potential ball candidates and reduce false positives. This step is based on the assumption that, once played, the ball moves faster and more consistently than any other object in the scene. To support this, the script calculates the magnitude of optical flow vectors between frames to estimate the object's speed. Using this speed, it predicts the expected distance the ball would travel between frames. Combined with the direction (angle) of the motion vector, the script determines the ball's next location and constructs its following trajectory. This filtering step, therefore, discards all the candidates that do not follow the predicted path, either by deviating in angle or speed significantly. This parallel trajectory filtering ensures that only consistent ball movements are tracked across frames, improving the reliability of the detection. Figure 8 shows the detected ball in red and the purple circles show the predicted position in the next frame.

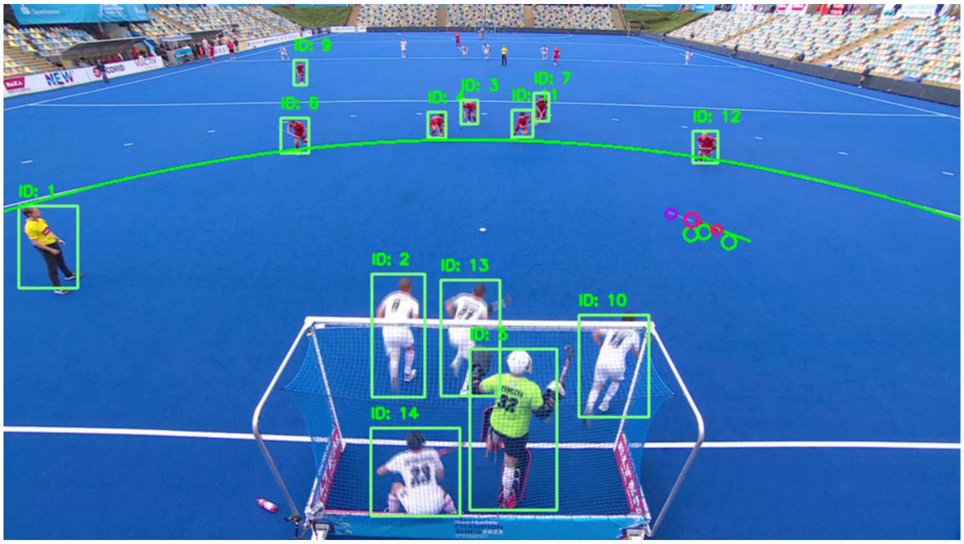


Figure 8: Frame showing successful ball tracking and prediction of the next position. The red trajectory markers represent the detected ball in the current frame, while the purple circle shows the predicted position for in the next frame. The green circles show possible candidates of the ball that do not align with the trajectory of the ball. These circles are therefore seen as noise.

An important part of the pipeline is identifying the exact frame where the ball reaches the top of the circle. This moment is critical for segmenting phases of play and engineering features related to both offensive and defensive strategy. When the ball is temporarily lost due to occlusion or blur, its trajectory is predicted using its prior motion vectors. This is visualized in Figure 9, where the red circle marks the estimated position at the top of the circle based on previous frames. This predicted moment is used to determine key tactical phases such as the initiation of the drag flick, defender pressure timing, and team compactness.

4.1.3 Player detection and tracking

To analyze player movements during penalty corners, this method first detects individual players using a deep learning object detection model YOLOv3 (You Only Look Once, version 3). This algorithm is applied to each video frame to identify and localize players on the field. YOLOv3 is chosen for its balance between detection speed and accuracy, making it well-suited for real-time analysis in sports videos [JEL+22].

Once players are detected, their positions are tracked across consecutive frames using Deep SORT (Simple Online and Realtime Tracking with a Deep Association Metric). Deep SORT builds on the original SORT algorithm by incorporating appearance based feature embeddings and Kalman filtering [MSAGS06]. This addition helps maintain consistent identity tracking across multiple frames, even in frames in which players briefly overlap or get occluded.

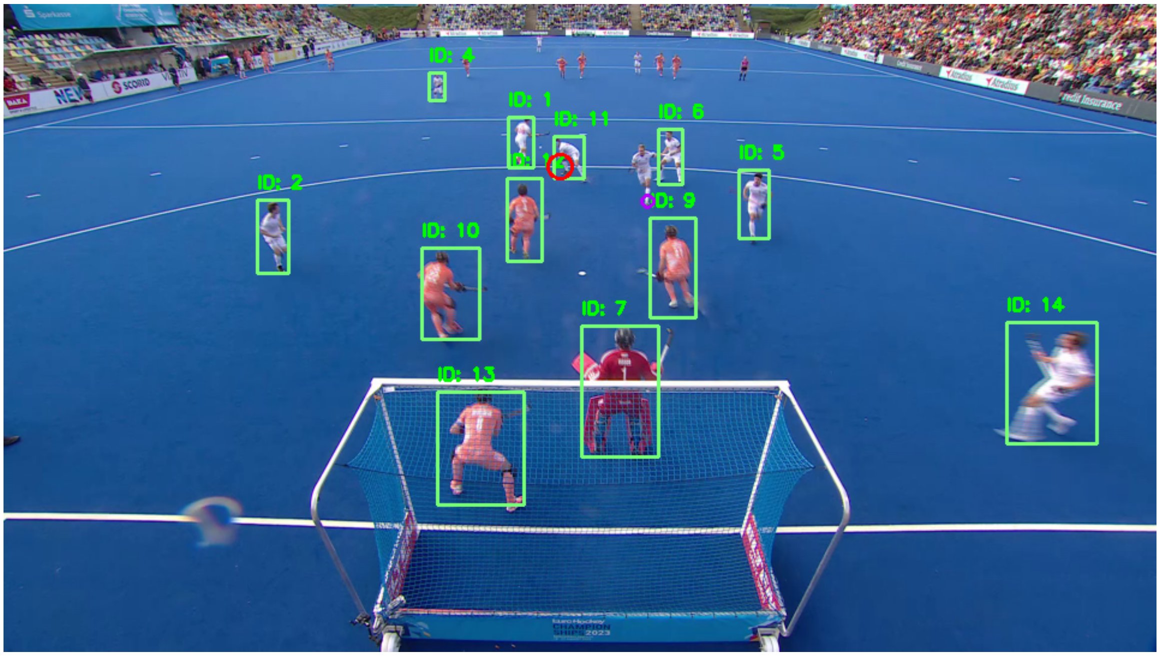


Figure 9: This frame shows the predicted position of the ball in red at the moment it reaches the top of the circle, after it was temporarily lost due to occlusions or unrecognizable shape.

This process results in a complete set of player trajectories, each linked to a unique player identifier. After detection, the player IDs are re-assigned based on their horizontal (left-to-right) position for both the attacking and defending team separately. This post-processing step ensures that player identities remain consistent across different videos, allowing for reliable comparisons of individual roles and movements. As this is a post-processing step, the IDs seen in Figure 10 are not sorted yet, this will be done afterwards. These trajectories form the basis for further analysis of individual and team behavior, such as positioning, movement patterns, and roles during a penalty corner.

Once players are detected and tracked, the next step is to determine which team each player belongs to. This is done using a two-phase method that combines initial positioning and dominant color clustering.

In the first phase, players are assigned to a team based on their initial position before the ball enters the field, which occurs in the lower right corner of the video. Since both teams start from well-defined positions, players can be grouped according to their relative location on the field. The defending team consists of up to five players positioned in the goal area at the bottom center of the frame. All other players are classified as attackers, as they are positioned closer to the shooting circle, ready to execute the penalty corner. The referee is also identified based on their unique positioning and excluded from both team labels.

In the second phase, the script analyzes the dominant color within each player's bounding box, reflecting their shirt color. Using K-Means clustering, players are grouped into two color-based clusters, which are used to either confirm or update the initial team assignment. This approach ensures that players are accurately classified, even if their initial position was unclear or if they entered the frame in a later stage during the corner. The same approach is also used to identify the goalkeeper, which is necessary for analyzing defensive formations and anticipations. In this case, the model only focuses on the previously identified defenders and applies K-Means clustering to detect the color outlier within that group. Since goalkeepers are obligated to wear a distinct jersey color, they can be reliably distinguished from their teammates using this method. The results of this clustering process and team seperation is depicted in Figure 10.

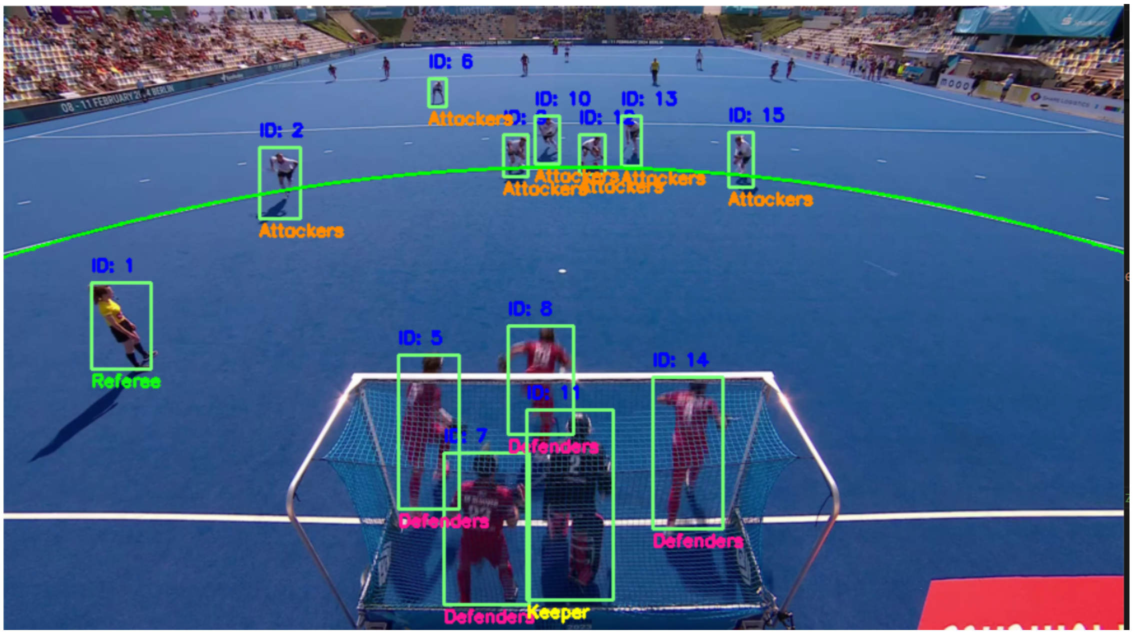


Figure 10: This frame shows illustrates the team assignment. Players are detected and tracked using YOLO and Deep SORT, after which they are classified as attackers or defenders based on their initial positioning and dominant jersey color. The goalkeeper and referee are identified separately to distinguish their separate roles.

4.2 Tactical Feature Extraction

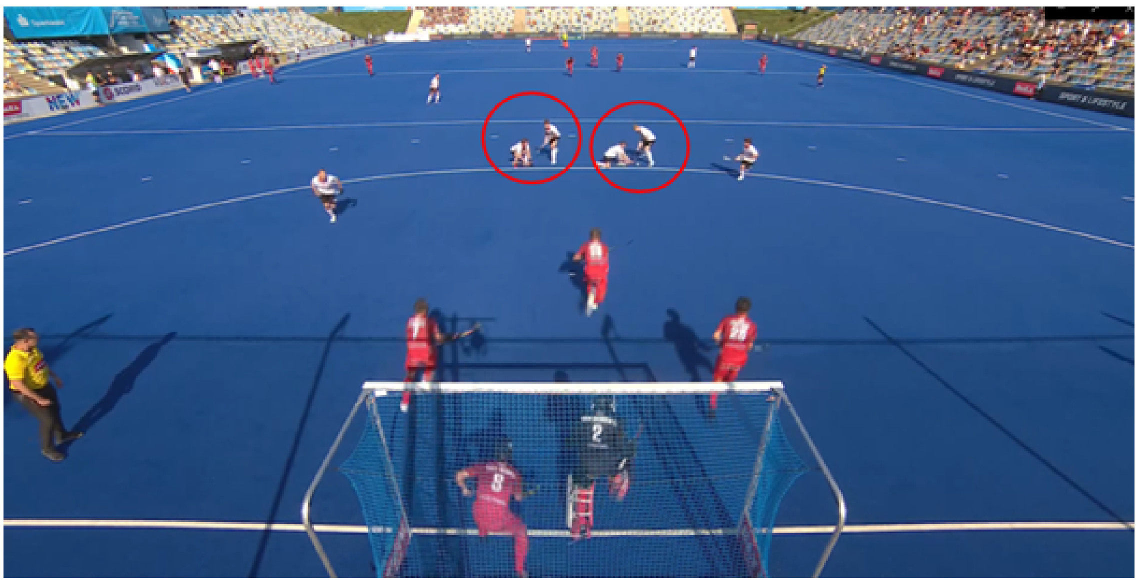
With the key visual elements extracted, such as player positions, the ball trajectory, and field boundaries, the next step is to convert this positional data into meaningful tactical features. These derived variables allow us to evaluate not just where players and the ball are located, but how they move and interact in the context of structured team strategies.

This subsection outlines how individual player metrics, such as positioning and movement variability, as well as team-level indicators, including player distribution and formation structure, are extracted. These features capture the strategic intention and coordination of both the attacking and defending teams, and serve as essential input for the later analysis of penalty corner effectiveness. To distinguish between team roles and their tactical opportunities and aim, this subsection is divided into offensive and defensive strategic features. A complete overview of all generated features is provided in Section 7 (Appendix).

4.2.1 Offensive Features

Ball Features

To analyze the offensive approach of a penalty corner, it is essential to accurately track the ball and extract relevant features. The first set of features concerns the speed of the ball, which is derived from its movement over consecutive frames. In addition, the x and y coordinates of the moment when the ball reaches the top of the circle are recorded, as this is the critical point where attackers typically receive the ball and initiate a shot or pass. The results are shown in Figure 9 where the centroid of the red circle is saved as the received top location. Finally, the model identifies which pair of attacking players, positioned at the top of the circle, receives the ball to initiate the potential shot or pass. The two sets of two attackers are positioned at the top to make it unclear for the defensive team which player will receive the ball as explained by [Ass22]. This is demonstrated in Picture 11. These features together provide a reliable basis for understanding how the play starts.



Picture 11: Showing the 2 sets of attackers of which one couple will receive the ball and initiate a shot

Player Features

To characterize the behavior of attacking players during a penalty corner, we extract spatial features for each attacker. Players are numbered from 1 to 8, ordered from left to right based on their average horizontal (x-axis) position. For each player, their mean x and y coordinates are calculated, as well as the standard deviation in both directions, to capture positional trends and movement variability. In addition, we focus on three critical moments in the attacking sequence that typically represent the following situations: 15 frames after ball reception (initiation of the shot), 25 frames after (shot execution), and 35 frames after (ball release of the stick, ball going towards the goal). The x and y coordinates of all attackers are recorded at each of these significant moments to understand their relative positions and movements as the play develops. These features serve to assess both team structure and potential passing options and rebound plays.

Team features

Beyond individual player metrics, we also extract features that describe the collective positioning and movement of the attacking team. At each of the three critical moments, 15, 25, and 35 frames after ball is received at the top, we compute a bounding box around all attackers to capture the area in which the attackers are distributed. The distribution has been further defined by taking the standard deviation of the movement of the attackers. We also calculate the change in distribution compared to the initial configuration, providing insights into how the team structure adapts as the play progresses. In addition, we compute the mean Euclidean distance between each attacker and the team's center (mean position), as well as the difference in these distances between the initial moment and the later frames. These features reflect the team's compactness or expansion during the penalty corner execution and help identify structured attacking formation and positioning versus improvised or disorganized attacks.

4.2.2 Defensive Features

Player Features
Similar to the offensive side, we extract spatial features for the defending players during the penalty corner. For each player we compute the mean and standard deviation of their x and y coordinates. Additionally, the x and y positions of defenders are recorded at the same three critical moments, 15, 25, and 35 frames after ball reception. This parallel structure enables direct comparison between the attacking and defending sides in terms of movement dynamics and formation discipline.

Team Features
To evaluate the defensive organization, we apply the same set of team-level features used for the attackers: bounding box, spreadingness, mean distances to the team center, and changes in these metrics across the three critical moments (15, 25, and 35 frames after ball reception). These features capture how the defensive unit expands, shifts, or holds its structure in response to the developing play.

In addition, we introduce a key feature unique to the defensive team, which is the initial formation before the ball enters the frame. Based on professional field hockey tactics, we identify one of four common defensive starting formations. This initial positioning is critical, as it lays the groundwork for how defenders move and organize themselves in the following frames. It strongly influences the team's structure and coordination at the three key moments of the penalty corner. These three moments, along with the defensive approach taken during each and the initial formation beforehand, are central to how coaches assess and develop effective defensive strategies.

Then, after determining the initial formation, at each critical moment, we further assess whether line stops are present. These are defenders positioned to the left, right, or both sides of the goalkeeper to block potential shots aimed at the corners of the goal. Lastly, we evaluate defender pressure by measuring how far defenders ran out from the goal area to pressure the shooter or intercept passes. Again, the number of players, the distance and distribution is a key element in the tactics of successfully defending a corner. All these features offer insight into the team's tactical response and adaptability under pressure of the scoring opportunity of the opponent.

4.3 Subgroup Discovery Analysis

To analyze the relationship between tactical player behavior and penalty corner success, this study applies Subgroup Discovery (SD) using the SubDisc tool [KGvDP21]. Unlike traditional machine learning models that focus on predictive accuracy, SD is designed to uncover interpretable subgroups—patterns within the data that are statistically associated with a desired target outcome.

Each discovered subgroup is defined by a logical rule (e.g., `mean_position_y_attacker_4` ≤ 113.4) and is evaluated based on how well it separates successful from unsuccessful penalty corners. Depending on the team role, the target class changes: for defending teams, success is defined as `target = 0` (no goal scored), whereas for attacking teams, success corresponds to `target = 1` (goal scored). This role-dependent framing allows the SD process to identify strategic patterns that are effective in both offensive and defensive contexts.

To evaluate the quality of each discovered subgroup, the Cortana Quality measure is used. Cortana Quality is based on the Weighted Relative Accuracy (WRAcc) metric, a widely accepted measure in subgroup discovery introduced by Lavrač et al.(1999) [lav99]. WRAcc captures the trade-off between subgroup relevance and coverage and is defined as:

$$WRAcc(R) = p(\text{Cond}) \cdot (p(\text{Class} \mid \text{Cond}) - p(\text{Class}))$$

Here, $p(\text{Cond})$ represents the relative frequency, or coverage, of the subgroup condition within the entire dataset, indicating how often the condition is met. $p(\text{Class} \mid \text{Cond})$ denotes the proportion of the target class (e.g., goal scored or not) within the subgroup, reflecting the local class distribution. Meanwhile, $p(\text{Class})$ represents the overall class probability in the entire dataset. The term $p(\text{Class} \mid \text{Cond}) - p(\text{Class})$ thus measures the contrast between the subgroup and the global distribution. WRAcc, and by extension Cortana Quality, favors subgroups that not only occur frequently (high coverage) but also display a meaningful deviation from the dataset average which (high contrast). This ensures that discovered patterns are both statistically significant and representative enough to generalize well in practical settings such as coaching strategy or match preparation.

Cortana Quality builds on the WRAcc measure by scaling its values to fall within the interval $[-1, 1]$, making it easier to compare subgroups across different datasets or experiments. A higher Cortana Quality score means the subgroup is both relatively large (it applies to many corners) and highly specific (the outcome in that group is very different from the average).

In addition, WRAcc has a useful geometric interpretation: it is proportional to the perpendicular distance from the diagonal line in ROC space (Lavrač et al., 2004) [LKFT04]. This means that subgroups with high WRAcc tend to have better true positive rates and fewer false positives, making them more effective at distinguishing successful from unsuccessful corner strategies. For this reason, Cortana Quality is particularly well suited for discovering tactical rules that coaches can trust and apply, such as formations or movements that consistently correlate with scoring or preventing goals. It offers both interpretability and performance, aligning perfectly with the needs of sports analytics, where clarity and real-world relevance are just as important as statistical accuracy.

In addition to defining the quality function (Cortana Quality), the subgroup discovery process also relies on a set of configurable parameters that guide the search through the space of possible subgroups. One key parameter is the minimum coverage threshold, which determines the smallest proportion of examples a subgroup must cover to be considered valid. In this study, we set a minimum coverage of 10% to ensure that discovered subgroups were both statistically robust and practically relevant. A maximum coverage of 90% was applied to avoid trivially large subgroups that mirror the overall dataset distribution.

The beam search strategy was applied with a search width of 100, which determines how many of the top-performing subgroup candidates are retained and expanded at each level of the search tree to find the most promising candidates [MVC20]. SubDisc supports various search strategies to guide this process. Depth-first explores subgroups along a single path to its full depth before backtracking, which can yield deep but narrow coverage [Kor85]. Best-first evaluates all current candidates and always expands the globally highest-scoring subgroup next, potentially missing diverse or early promising alternatives. ROC-beam prioritizes candidates that maximize the area under the ROC curve early in the search. In contrast, beam search limits the number of candidates per level while still exploring broadly across different directions in the search space.

Beam search is selected because it strikes a good balance between efficiency and diversity. With a search width of 100, the algorithm can retain and consider many promising subgroups at each iteration, enabling a wide exploration of tactical patterns. This is especially valuable in our case, where the dataset contains a high number of features derived from movement and positioning. Beam search helps avoid being overly biased by the first few rules while still keeping the process computationally manageable. It also increases the likelihood of finding multiple, distinct tactical insights rather than just one optimal rule set.

For numeric attributes, the numeric strategy was set to best, which dynamically determines the most informative way to split continuous variables at each depth. This setting evaluates all possible split points for a numeric variable within a candidate subgroup and selects only the single best-performing split to expand. This approach avoids arbitrary binning while maintaining computational tractability. Other strategies supported by SubDisc include numeric bins and best-bins, which generate bins at equal-width intervals within the current subgroup's value range . The difference in these two approaches lay whether all splits (bins) or only the best one (best bins) is retained. These strategies help ensure that meaningful subgroup patterns are discovered even in high-dimensional, continuous-feature spaces.

By setting these parameters and applying this technique to features, the subgroup discovery framework enables the identification of statistically significant patterns that reflect successful offensive and defensive strategies. These subgroups are then visualized through ROC curves and form the basis of tactical insights presented in the next section.

5 Results & Experimental Evaluation

The results of this study are presented in this section. As described in Section 3, a wide range of features were extracted from video data to capture player behavior, team strategy, and ball movement during penalty corners. Afterwards, these features were analyzed using subgroup discovery techniques to identify patterns associated with successful and unsuccessful outcomes. In this chapter, we present the most impactful single and combined feature conditions discovered for each team. These results are then compared to known field hockey strategies to assess their tactical validity and to provide evidence-based recommendations for both defending and attacking penalty corner setups.

5.1 Computer Vision Results

Circle Detection

This detection method was manually tested on multiple frames and consistently produced a stable and accurate representation of the circle. While the masks limit the region of interest, they do not bias the model, but instead improve precision by reducing irrelevant noise. This was a necessity given the fast-moving, visually complex scenes in penalty corner sequences. Because of this, no precise detection accuracy percentage is provided, but visual inspection shows successful fits in approximately 96% of the videos. Since this detection serves as a foundation for all subsequent tactical feature extraction, it plays a critical role in determining the reliability of the results. If the circle is not accurately detected, it can lead to misaligned reference points, which may distort the extracted features and ultimately affect the outcome of the analysis.

Baseline Detection

To evaluate the accuracy of the baseline detection, the videos were also manually reviewed. While the algorithm is able to detect the correct horizontal structure in all cases except one, due to rain, very minor inconsistencies can occur when defenders or the goalkeeper obstruct the line. These occlusions may lead to very small deviations in the y coordinates of baseline estimate and its slope, which in turn could impact downstream features, such as player distance to the goal. However, since this study used the average of all detected baselines, as shown in Figure 7, minor deviations are accounted for. As a result, they do not affect the extracted features, making the baseline detection method robust.

Ball Detection and Tracking

The accuracy of this detection step is crucial because it anchors time-based tactical features. If the moment of reception at the top of the circle is wrongly detected, it can shift all subsequent measurements (e.g., player distances, angles, or movement velocities), leading to unreliable feature values. Although no exact percentage of accuracy is computed here, manual inspection indicates that the pipeline correctly identifies or predicts the ball position in around 90% of all cases. The remaining 10% of cases, in which the ball is lost or incorrectly detected due to noise, can significantly distort the measurement of engineered features. As a result, improving the precision of ball detection in these instances would substantially enhance the overall reliability of the analysis.

Player Detection and Tracking

The identification of players using YOLO was generally very successful. Even players positioned unusually, such as crouched defenders or those partially occluded, were reliably detected in most frames. The Deep SORT tracking algorithm also performed well overall, maintaining consistent player IDs across frames. However, in some cases, particularly during heavy occlusion or when a player briefly left and re-entered the frame, the tracker incorrectly assigned a new ID to the same individual. This type of identity switch can affect the reliability of engineered features, such as the player's mean position, movement variability, or their role in tactical formations at critical moments. While exact accuracy cannot be quantified, we estimate that approximately 10–15% of player tracks were affected by such mismatches.

The dominant color-based approach also performed well. The goalkeeper was accurately classified in the majority of cases, with only 9 misidentifications out of the 97 analyzed penalty corners. Moreover, assigning players to a team based on uniform color, without prior knowledge of the initial formation, successfully identified several team members correctly. The use of a threshold-based method, where players are labeled as "unknown team" when classification confidence is low, contributed to the robustness of the approach by avoiding incorrect assignments. Overall, this method enables a reasonably accurate team labeling process, which is essential for analyzing team strategies, spacing, and individual roles during penalty corner execution.

5.2 Most Successful Defending Feature Outcomes

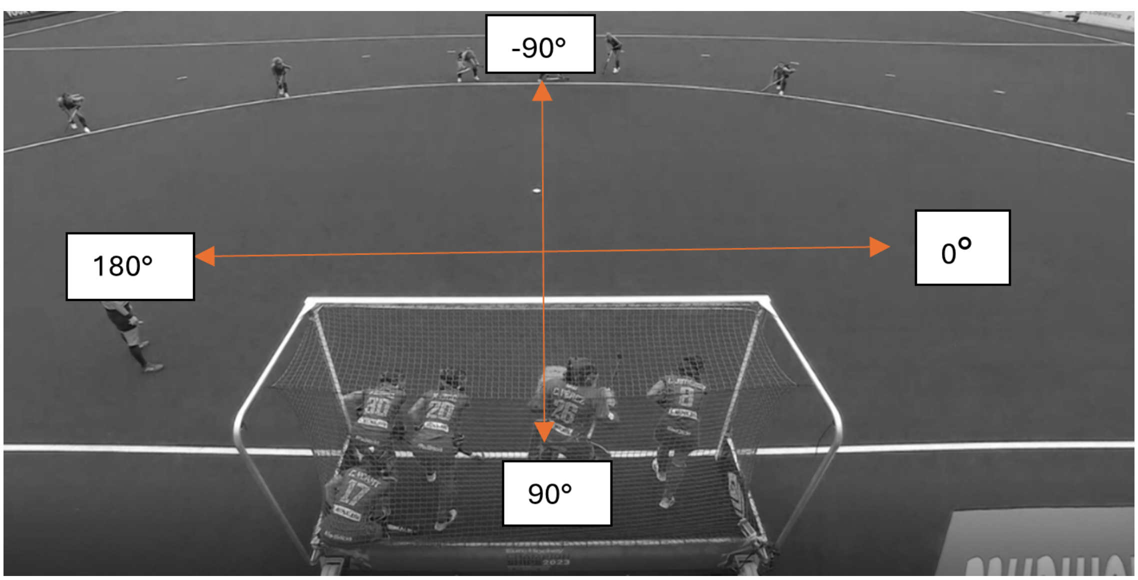
We start by focusing on the defending team. In a defensive context, a penalty corner is considered successful when the opposing team does not score a goal, so target = 0. The following section presents the top 5 single and combined feature subgroups that are strongly associated with successful defensive outcomes. In this section, only features that reflect actions performed by the defenders themselves are used. The aim of this study is to identify the most successful defensive strategies, thus focusing only on these defensive features rather than reactive responses to the offense.

5.2.1 Single feature results

The results in Table 1 present the top five single-feature subgroups most strongly associated with a non-scoring penalty corner. The first feature, `T1_angle_prev_highest`, reflects the movement angle of the flyer, the highest out-running defender, at critical moment T1. This is a very common strategy as it will pressure the attackers to play the ball fast [HC nd]. Moment T1 occurs shortly after the ball is received at the top of the circle. This angle is calculated using the defender's initial formation when the ball enters the frame (the start of the corner) and current position. All possible angle values are visualized from the red dot in Picture 12. An angle of $-81°$, for instance, corresponds to movement approximately 1 pixel upward and 0.15 pixels to the right (see Picture 12). A threshold of $\geq -81°$ implies that optimal flyer movement should not follow a straight vertical line (which would be $-90°$), but instead follow a diagonal trajectory to the right. This aligns with common defensive practice, where the flyer starts on the left or center-left side of the goal and runs slightly diagonally right toward the ball receiver to pressure the shooter. This has also been evaluated by Sportplan Hockey [Spo24]. Such positioning allows the flyer's forehand to align with the shooter's forehand, increasing the likelihood of intercepting a drag flick or blocking the shot line. This implies a flyer starting position on the left of the goal; however, the data on initial formations is not robust enough to make a definite statement about this.

The second, third, and fourth features highlight the spatial and movement behavior of the defender identified as `defender_2`, the second defender from the right. While this player can either serve as a flyer or a line stopper depending on the team's formation, Picture 13 illustrates how `defender_2` can be located to the left of the goalkeeper. This is one of the two roles that defender 2 can perform during the corner; the other role is serving as the flyer. The condition `mean_position_y_defender_2` ≤ 375 indicates that this player should not remain deep near the baseline ($y \approx 400$), which is typical for a line stopper. Similarly, `mean_movement_35_defender_2` ≥ 98 at T3 (35 frames after reception) suggests significant movement, unfitting with the relatively static line stopper role. Combined with the condition `mean_std_dev_x_defender_2` ≥ 13.3, which indicates greater horizontal variation, the data supports the recommendation that `defender_2` should act as a dynamic flyer, rather than a static line stopper. The data, therefore, suggests that `defender_2` should actively run out to challenge the ball, rather than remain fixed near the goal line.

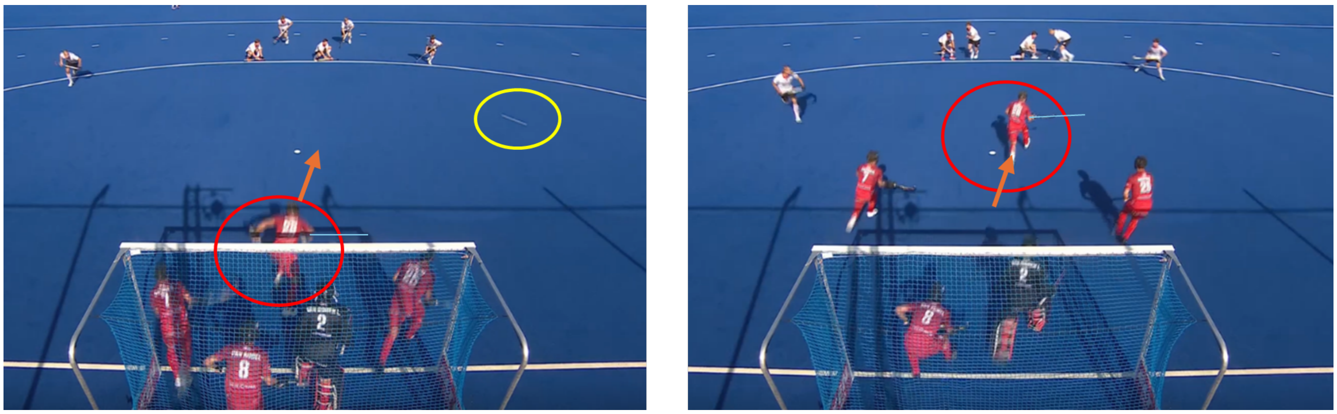
The fifth condition focuses on `T2_angle_prev_middle`, referring to the movement angle of the middle-most outrunning defender at T2, the moment the shot is initiated. A threshold of $\geq -95.3°$ again implies an out-running vertical movement, possibly with a slight curve. This aligns with typical tactics where the middle defender is also running out to help the flyer with pressuring the ball from the sides as demonstrated by nfhca (national field hockey coaches association) [Ass22]. In this way, the defenders reduce direct shooting opportunities by supporting the left and right side of the flyer. Therefore, the middle outrunner is either the left most player, running slightly rightwards to the top, or represents the most right defender and is expected to run toward the top of the circle in a straight line or slightly leftwards. The most right and left defenders cover the sides of the flyer by standing in a line from the shooter toward the goal, beyond the reach of the goalkeeper, as depicted in Picture 14. Based on these features, we cannot determine whether the data prefers the right or left player to serve as the middle outrunner. We can only conclude that the middle runner should still move upwards at critical moment T3 and not shift too far to the right.



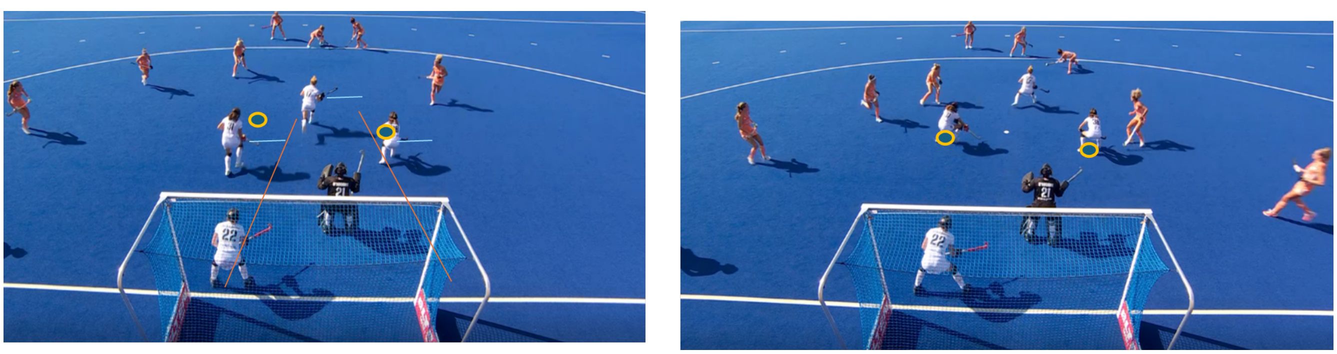
Picture 12: Illustrating all possible movement angles that players can take, calculated from the point of intersection.

Coverage	Quality	Target Share	Positives	Condition
45	0.3710	0.8889	40	`T1_angle_prev_highest` ≥ -81.38
45	0.3185	0.8666	39	`mean_position_y_defender_2` ≤ 375.47
36	0.2968	0.8888	32	`mean_movement_35_defender_2` ≥ 98.06
58	0.2914	0.8276	48	`mean_std_dev_x_defender_2`≥ 13.32
42	0.2762	0.8571	36	`T2_angle_prev_middle` ≥ -95.25

Table 1: Top five single-feature subgroups associated with successful defensive corners (Target = 0)



Picture 13: Showing the ball in yellow and the flyer in red and his movement when the ball has just arrived at the top of the corner. Picture a shows the initial formation, followed by a flyer run out in picture b and the angle between these movements represent `T1_angle_prev_highest`. In this case, the angle will be slightly larger than $-81.38°$



Picture 14: The yellow circles in picture a represent the positioning of their feet in a following frame b. It illustrates a situation in which the possible angle the middle out runners make satisfy the condition of this subgroup discovery as being the most right or most left player

To interpret the values in Table 1, it is important to understand what each metric represents. The coverage column shows the number of corners that match a specific single-feature condition. For instance, the top condition (`T1_angle_prev_highest` $\geq$ `-81.38`) is present in 45 out of the 97 corners. Of those 45 corners, 40 were successfully defended, which corresponds to a target share of

88.9%. This is considerably higher than the overall success rate (baseline) of 73.2%, suggesting that this condition is associated with a higher-than-average chance of defensive success.

The Cortana Quality score is a normalized metric ranging from -1 to 1. A score greater than 0 indicates that the subgroup contains a higher proportion of the target class (e.g., successfully defended corners) than the overall dataset, whereas a score below 0 suggests that the subgroup performs worse than the baseline and thus worse than random selection. This quality score accounts for both the size of the subgroup (coverage) and its consistency (target share), while also comparing it to the overall class distribution. Higher CQ scores reflect patterns that are not only more concentrated in positive outcomes but also more distinctive relative to the dataset as a whole. Therefore, a high CQ value such as 0.3710 suggests that the identified subgroup is both statistically meaningful and tactically relevant.

Although the earlier results highlight the top 5 single subgroups based on Cortana Quality, the combinations shown in Figure 15 (ROC curve) represent those that most effectively discriminate between positive and negative outcomes across the full dataset. These are selected to maximize the area under the ROC curve (AUC = 0.738) for the target value of 0, which reflects a successful defensive penalty corner. As a result, the ROC-based rule set includes a different combination of features, such as `T2_percentage_uitloop_middle` and `Mean Distance Defender Difference T1`, that score lower in individual quality but work well together. This difference arises because subgroup quality is locally optimized, whereas the ROC curve captures global classification performance. `T2_percentage_uitloop_middle` therefore asks the middle defender to not run further than 67.8% of the circle, probably to ensure rebound control and to support the flyer. `Mean Distance Defender Difference T1` refers to the total average difference of the defending team in comparison to their initial formation. Therefore, this suggests that the defending team should play dynamically and actively run outwards, away from the goal, with multiple defenders.

5.2.2 Multiple feature results

In addition to single-feature subgroups, several combined conditions emerged as strong indicators of effective defensive penalty corner execution, as shown in Table 2. These combinations reflect even more nuanced tactical decisions. The first combination includes the conditions `mean_position_y_defender_2` ≤ 375.47 and `mean_position_x_defender_3` ≥ 414.39. The lower y-position of `defender_2` again suggests that this player should actively run out, avoiding deep positioning near the baseline. In contrast, `defender_3`, often positioned slightly wider, should not drift too far left but instead remain more central or right, with an x-position above 414.

If `defender_3` acts as a line stopper, this helps ensure that they protect the goal area effectively, avoiding positioning too far right where they would be covering shots that would otherwise miss the goal. If `defender_3` instead acts as a flyer, a higher x-axis position ensures that the left side of the goal is protected, while still applying pressure to the shooter near the center of the pitch. On a standard 1100-pixel wide frame, moving below $x = 414$ would place the player closer to the left side of the goal and further from the center, potentially leaving a critical area more exposed to a direct shot.

The second subgroup is defined by two conditions, the horizontal variability of `defender_1` must

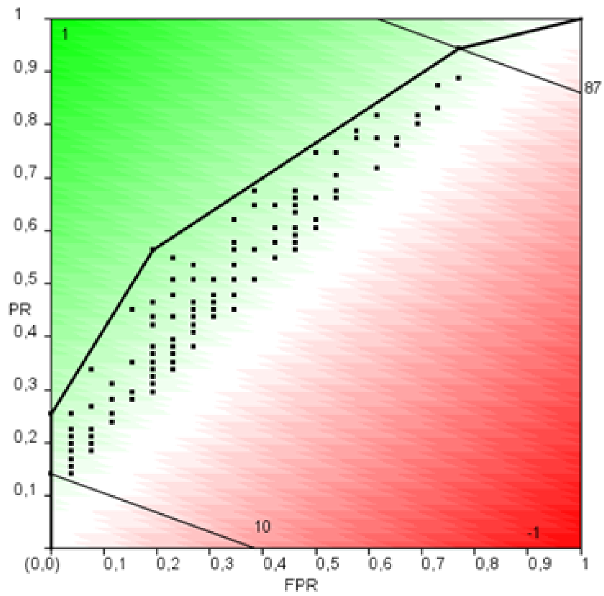


Figure 15: Showing the ROC with area under the curve of 0.738 for Target = 0. This figure shows the curve of the most successful defensive approach with the following conditions: `T2_percentage_uitloop_middle` $\leq$ 0.679, `T1_angle_prev_highest` ≥ -81.384, `Mean Distance Defender Difference T1` ≥ 58.891

be at least 10.38, and the `Bounding_Box_Area_Defenders_T3` of all defenders at frame T3 must not exceed 51336. Here, `defender_1`, who is often assigned to the lower middle out-running role [Ass22], should demonstrate a degree of horizontal variability in their movement, as illustrated in Picture 14 with the yellow circle on the left. This implies that the defender should not run in a straight line, but rather attempt to minimize the shooting area by forming a triangular shape with the flyer and the other out-runner. This setup is highlighted by the orange lines showing the flyer's reach and the corresponding zones where the out-runners must provide coverage. The blue lines depict their defensive reach with their sticks. A compact bounding box at T3 (the frame at which the shot is executed) indicates that defenders should avoid spreading too widely. Instead, they should form a tight unit, effectively reducing shooting lanes and blocking tip-in or rebound opportunities. This positioning is especially effective for intercepting direct shots and neutralizing unpredictable ball deflections or redirections by the attacking team.

The third combination revisits the condition `T1_angle_prev_highest` ≥ -81.38 for the flyer and combines it with `mean_std_dev_x_defender_3` ≤ 30.64. As discussed earlier, a flyer angle greater than $-81°$ indicates a diagonal rather than straight vertical movement, improving pressure on the shooter by aligning the flyer's forehand with the anticipated shot trajectory, as depicted with the blue line in Picture 14.

The inclusion of `defender_3`'s horizontal variability further refines this subgroup. When `defender_3` serves as a line stopper beside the goalkeeper, they should make only minimal lateral adjustments to fine-tune their positioning relative to the shot angle and the goalkeeper's coverage. If `defender_3` instead acts as a flyer, their trajectory should remain direct and efficient, following the shortest path to the top of the circle and avoiding unnecessary lateral deviations that could delay interception.

Both interpretations emphasize the importance of role clarity and disciplined movement for effective penalty corner defense.

The combined-feature subgroups presented in Table 2 are particularly noteworthy due to their very high target shares, ranging from 90.9% to 97.2%, meaning that nearly all corners within these subgroups were successfully defended. Although the coverage of each rule lies between 34 and 44 corners, the high Cortana Quality scores, all above 0.40, demonstrate strong discriminative power. These scores indicate that the identified patterns are not only significantly better than the baseline success rate of 73.2%, but also statistically robust and tactically meaningful. The results reinforce the importance of both individual defender positioning and collective team compactness at critical moments during the penalty corner. Coaches should therefore use these combinations of insights to train players on specific movement coordination that consistently lead to successful defensive outcomes.

Coverage	**Quality**	**Target Share**	**Positives**	**Conditions**
36	0.4544	0.9722	35	`mean_position_y_defender_2` $\leq$ 375.47 $\wedge$ `mean_position_x_defender_3` $\geq$ 414.39
34	0.4263	0.9705	33	`mean_std_dev_x_defender_1` $\geq$ 10.37 $\wedge$ `Bounding_Box_Area_Defenders_T3` $\leq$ 51336.0
44	0.4095	0.9090	40	`T1_angle_prev_highest` $\geq$ -81.38 $\wedge$ `mean_std_dev_x_defender_3` $\leq$ 30.64

Table 2: Top three combined-feature subgroups associated with successful defensive corners (Target = 0)

Again, this ROC-curve represents some different features than discussed before, as seen in Figure 16. `T1_speed_highest` and `T1_angle_top_highest` are a mainly significant combination, as both features have not been explained before. The first feature, `T1_speed_highest`, indicates that the flyer should run with a speed faster than 107 pixels at the moment the ball is received at the top. As the baseline's mean y is around 425 and the top circle y is 135, the flyer should have run with a speed of approximately 5.3 m/s. The feature `T1_angle_top_highest` explains the angle the flyer is making with the recipient location of the ball and the anticipated shot direction. Since this angle is now measured from the top, the degree scale is different: 90° is a straight line toward the top, while 0° points directly to the right of the top. An angle larger than 48° suggests that the flyer should probably run slightly rightward while keeping the angle smaller than a diagonal line (45°).

For the best subgroups with a depth of three, the area under the curve increased slightly to 0.775 for Target = 0. However, this modest improvement, combined with the absence of notably new or impactful features, led us to exclude these results.

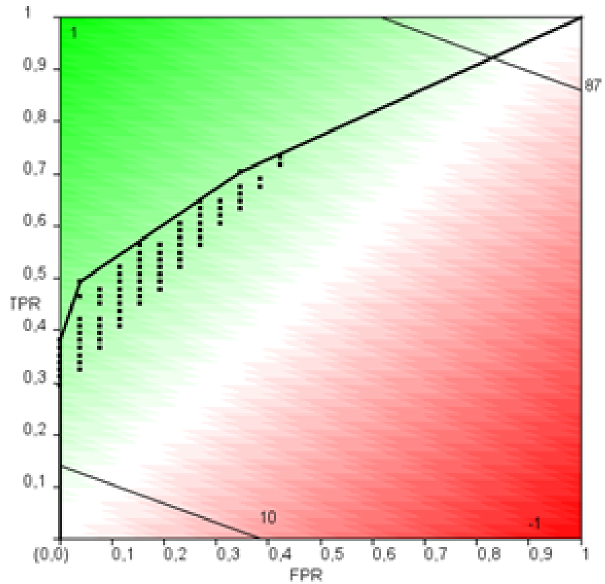


Figure 16: Showing the ROC with area under the curve of 0.758 for Target = 0. This figure shows the curve of the most successful defensive approach with the following conditions: `Mean_movement_35_defender_2` $\geq$ 98.061 $\wedge$ `Mean_std_dev_y_defender_2` $\leq$ 103.664, `Mean_position_y_defender_2` $\leq$ 375.474 $\wedge$ `Mean_position_x_defender_3` $\geq$ 414.393, `T1_speed_highest` $\geq$ 107.927 $\wedge$ `T1_angle_top_highest` $\geq$ 48.593.

5.3 Worse Defending Feature Outcomes

5.3.1 Single feature results

The following subgroups represent the 3 least successful defensive outcomes, where the target equals 1, indicating that the opponent scored. These results are based exclusively on defensive features. The first subgroup from Table 3, `mean_position_y_defender_2` $\geq$ 376.12, was already discussed in the previous section and once again highlights the importance of not allowing defender 2 to remain too close to the baseline (around 380). This supports the notion that defender 2 should act as a flyer, actively running out to apply pressure. The second and third subgroups involve the standard deviation of the x-position for defenders 2 and 1, respectively. Both have low variability thresholds, suggesting that limited horizontal movement is associated with unsuccessful defensive outcomes. This aligns with expectations in high-level hockey, where defenders must remain adaptable and mobile, adjusting their positioning laterally to respond to ball movement, attacker placement, and potential rebounds. Therefore, a higher horizontal variability reflects more responsive and effective defensive behavior.

The subgroups in Table 3 highlight defensive configurations associated with a higher success rate than the dataset's baseline of 26.8%. Each of these conditions results in a target share below 44%, indicating that the majority of corners meeting these criteria led to a goal. The Cortana Quality scores range from 0.2876 to 0.3466, which, while a little lower than scores for subgroups from previous subsection, still reflect meaningful deviations from the baseline.

The ROC curve with an AUC of 0.728, shown in Figure 17, also emphasizes the importance of

Coverage	Quality	Target Share	Positives	Condition
50	0.3466	0.4000	20	`mean_position_y_defender_2` $\geq$ 376.11
37	0.3196	0.4324	16	`mean_std_dev_x_defender_2` $\leq$ 12.95
43	0.2876	0.3953	17	`mean_std_dev_x_defender_1` $\leq$ 10.26

Table 3: Worst-performing single-feature subgroups associated with unsuccessful defensive outcomes (Target = 1)

individual defender positioning and horizontal variability. Notably, the feature `Mean Distance Defender Difference T1` once again suggests that a greater level of dispersion among defenders contributes to more successful defensive outcomes.

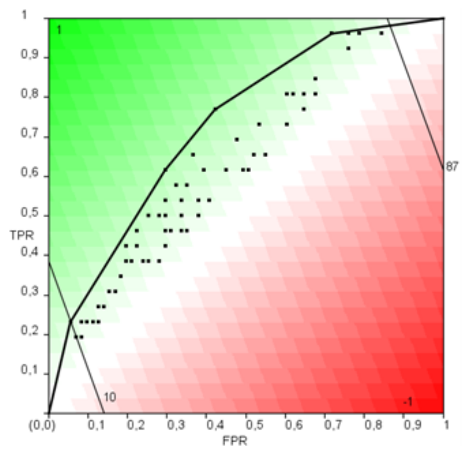
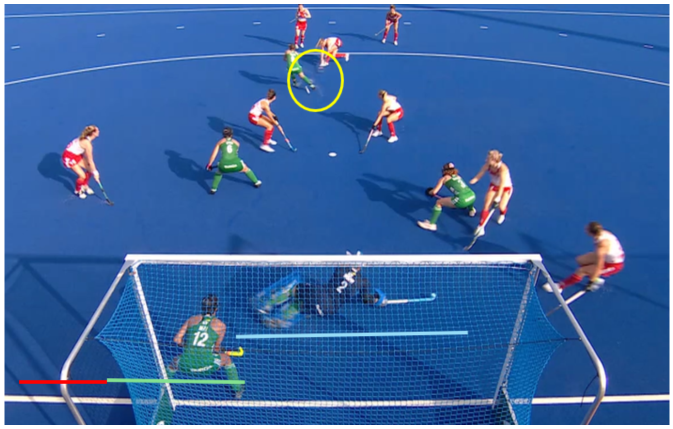


Figure 17: Showing the ROC with area under the curve of 0.728 for Target = 1. This figure shows the curve of the worse defensive approach with the following conditions: `Mean_Distance_Defender_Difference_T1` $\leq$ 57.590, `Mean_std_dev_x_defender_2` $\leq$ 12.954, `Mean_position_y_defender_2` $\geq$ 376.115, `Mean_std_dev_y_defender_3` $\leq$ 95.494

5.3.2 Multiple feature results

The combined subgroups at depth 2 do not show much new information and similar features as in subgroup at depth 1. Subgroups at depth 3, however, reinforce the value of having fast defenders who can run out effectively toward the top of the circle, creating defensive spread and disrupting attacking plays. As the corresponding table did not yield many additional significant patterns, we focus instead on the ROC curve with a clean trajectory and an AUC of 0.865, as shown in Figure 19. This curve represents the best-performing rule set among all the compared depth-3 subgroups.

Notably, this is the first instance where the line stopper's position is included as a predictive factor for defensive success. Specifically, the left line stopper's x-position at T3 is expressed as a percentage relative to the horizontal axis, with the ball reception lo The rule indicates that the line stopper should not be positioned more than 23the left of the ball reception point. A position further to the left could place the defender outside the goal frame, unable to block any shot, as illustrated by the red line in Figure 18. Furthermore, if positioned too far left, the line stopper may be unable to cover the vulnerable zone just outside the reach of the goalkeeper's left side, his reach is indicated by the light blue line in Picture 3. This vulnerability is particularly critical because, unlike the right side, where the goalkeeper's stick can offer protection, the left side relies heavily on proper line stopper coverage.



Picture 18: Illustrating the correct positioning of the left line stopper with a green line adjusted to current ball location. The light blue line demonstrates the reach of the keeper and therefore the required support of the line stopper in the left corner.

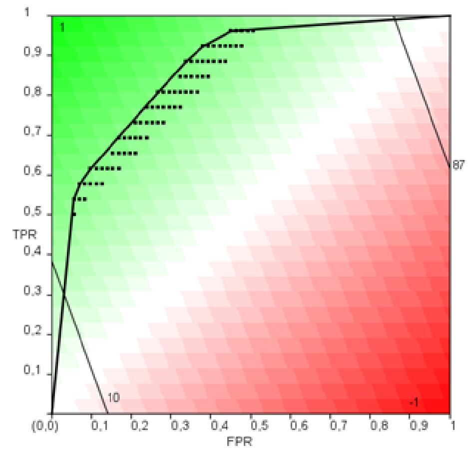


Figure 19: ROC curve with an area under the curve of 0.863 for Target = 1. This figure shows the curve of the worst defensive approach, based on the following conditions:
`Mean_position_y_defender_2` $\leq$ 376.115 $\wedge$ `Curr_Spreadiness_Defender_T1` $\leq$ 40.214 $\wedge$ `Mean_position_x_defender_3` $\geq$ 435.745
`Mean_std_dev_x_defender_1` $\leq$ 10.261 $\wedge$ `Mean_position_y_defender_3` $\geq$ 358.548 $\wedge$ `Mean_std_dev_x_defender_3` $\geq$ 2.687
`Defender_Lijnstop_T3_Left` $\leq$ -0.229 $\wedge$ `Mean_std_dev_y_defender_3` $\leq$ 88.011 $\wedge$ `Mean_std_dev_x_defender_3` $\geq$ 2.687
`Mean_std_dev_x_defender_3` $\geq$ 2.687 $\wedge$ `Mean_position_y_defender_2` $\leq$ 374.378 $\wedge$ `Spreadiness_Defender_Difference_T1` $\leq$ 1.447

5.4 Most Successful Attacking Feature Outcomes

Following the analysis of defensive strategies, this section focuses on the attacking side of penalty corners, where the objective is to maximize the likelihood of scoring (Target = 1). Out of 97 analyzed corners, 26 resulted in a goal, corresponding to a baseline success rate of 26.8%. As in the defensive evaluation, player positions, movements, and tactical elements were extracted from video data and analyzed using subgroup discovery methods. This chapter presents the most influential single and combined attacking feature conditions associated with scoring outcomes. These results are interpreted in the context of known offensive strategies in field hockey, aiming to identify which tactical patterns contribute most to effective penalty corner execution.

5.4.1 Single feature outcomes

The analysis of the top single-feature subgroups is illustrated in Table 4. The successful attacking penalty corners reveal several key positioning and movement patterns. First, the feature `mean_position_y_attacker_4` $\leq$ 113.42 suggests that `attacker_4`, often the first pair stopper who receives the ball often and therefore, he should remain positioned high near the top of the circle to receive the ball as the mean top Y is $Y = 135.38$. This aligns with the view of nfhca (national field hockey coaches association) [Ass22]. This illustrates that the player should stay outside the circle. This positioning likely supports scenarios such as rebound control or counters, where the player must be prepared to react quickly to a deflected ball.

Second, the feature `mean_movement_15_attacker_3` ≤ 56.01 indicates that `attacker_3`, typically the shooter of the second couple *koppel*, should remain relatively stable in the early moments of the play for rebound and for faking or actually receiving the ball. Additional support comes from the complementary condition `mean_movement_15_attacker_3` ≥ 5.83, which shows that the attacker should not be entirely static but make only minor, controlled adjustments to maintain formation and adjust it position for an accurate reception.

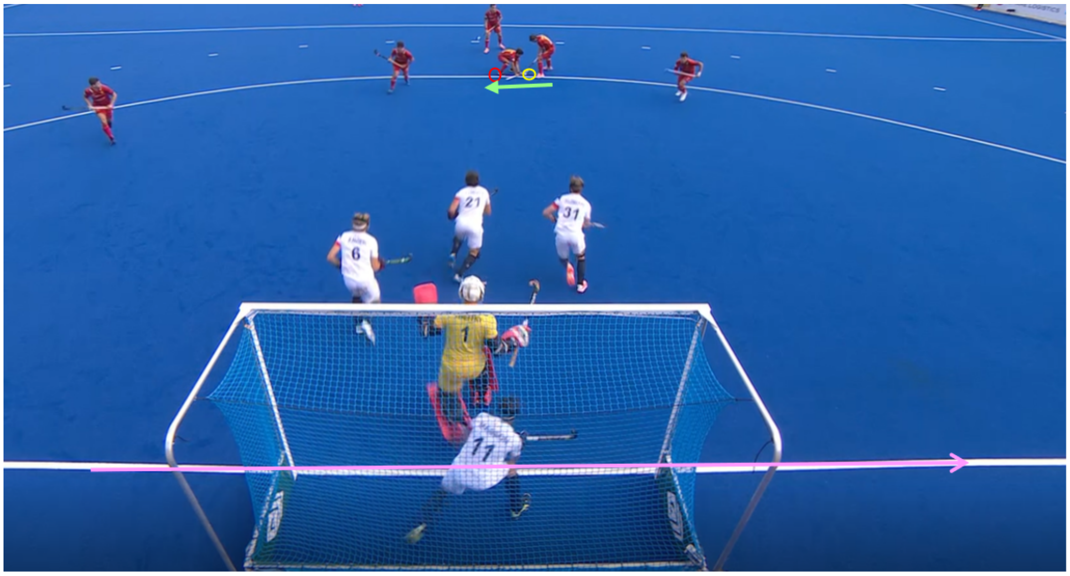
The feature `Angle_to_Top_of_Corner` ≥ 176.38 emphasizes the importance of the approach angle between the location of the reception of the ball and the top of the circle. This calculation is the angle between baseline vector (pink) which is always from left to right and the vector between the coordinates of reception region and the top circle (green). The angle between these two vectors captures how directly the ball approaches the top of the circle relative to the baseline. Such a high angle implies a location nearly horizontal leftwards from the top as depicted in Picture 20. This supports that the top left of the corner and often the first couple has the most scoring opportunities as shown in [Edu20] and that a clean and accurate pass from an inserter located on the baseline is required.

Finally, the condition `Corner_X` ≥ 503.0 confirms that the ball should be passed near the central region of the circle (mean $X = 505$) and at least more rightwards than x-coordinate 503. This location enables optimal shooting opportunities with shooting chances towards both left and right regions.
When focusing on the performance, Table 4 presents the five most predictive single-feature subgroups associated with successful attacking penalty corners (Target = 1). While the Cortana Quality scores range from 0.2546 to 0.3185, and the target shares lie between 33.8% and 38.5%, these values still represent a notable improvement over the baseline scoring rate of 26.8%. Importantly, the coverage values are relatively high, each subgroup applies to over 50 corners, highlighting their practical relevance while having a lower target share (consistency). Collectively, these findings reinforce the importance of spatial positioning and precision in ball delivery to maximize offensive effectiveness.

Coverage	**Quality**	**Target Share**	**Positives**	**Condition**
52	0.3185	0.3846	20	`mean_position_y_attacker_4` ≤ 113.42
65	0.2930	0.3538	23	`mean_movement_15_attacker_3` ≤ 56.01
58	0.2865	0.3620	21	`Angle_to_Top_of_Corner` ≥ 176.38
71	0.2611	0.3380	24	`mean_movement_15_attacker_3` ≥ 5.83
64	0.2546	0.343	22	`Corner_X` ≥ 503.0

Table 4: Top five single-feature subgroups associated with successful offensive outcomes (Target = 1)

Figure 21 presents the ROC curve with an AUC of 0.717 for Target = 1, indicating the most successful offensive subgroup based on combined features. The condition `Ball_speed` ≥ 460.79 highlights the importance of a fast injection, as a quicker pass provides attackers with more opportunity to shoot while limiting defenders' reaction time. Defenders therefore, have less time



Picture 20: The pink line illustrates the baseline vector, always from left to right. The green arrow shows the direction vector of the top of the circle (red dot) and the location of the ball reception (yellow). This illustrates that the positioning of the ball is almost in a straight line opposite of the baseline.

to run out and move to adjust their positioning for optimal defending. The remaining conditions involving `attacker_3` and `attacker_4`—already discussed in earlier sections—reinforce the need for strategic positioning and controlled movement of the first couple to optimize scoring chances during the penalty corner.

5.4.2 Multiple feature results

The most successful depth-2 subgroups, shown in Table 5, again emphasize the importance of controlled movement by attackers 3 and 4, the first and often main couple, while also introducing broader team-level spatial dynamics. The condition `Y_Mean_Attacker_T1` ≤ 140.71 suggests that the average vertical position of all attackers at the moment of reception should remain close to the top of the circle (with the average Y top circle $Y = 135$ as reference). This implies that while some attackers can begin running towards the goal, a portion of the attacking team should remain lower around the circle. This supports strategies that promote positional ambiguity, making it harder for defenders to anticipate who will receive the ball, and also improves opportunities for rebound control.

Additionally, the condition `Bounding_Box_Area_Attackers_T1` ≤ 71858.0 indicates that at the point of ball reception, attackers should not be overly dispersed. A more compact formation supports faster passing options and coordinated execution, while maintaining enough spacing to remain unpredictable and responsive. The subgroups in Table 5 exhibit relatively high Cortana Quality

scores, all exceeding 0.42, which indicates strong subgroup performance in terms of both purity and contrast with the overall dataset. While the individual target shares range from approximately

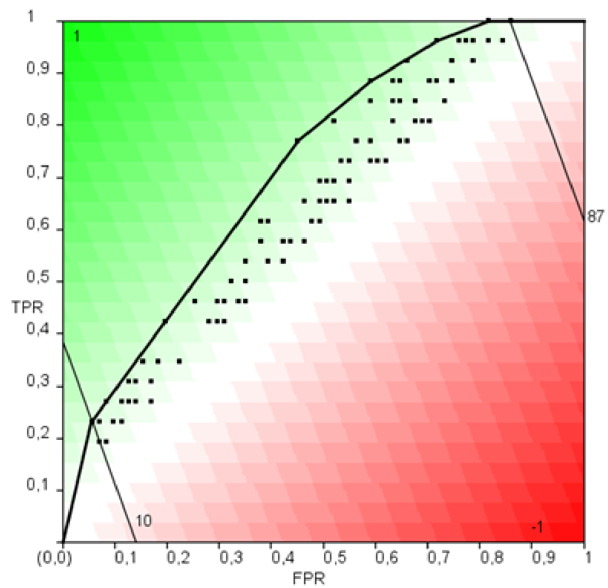


Figure 21: Showing the ROC with area under the curve of 0.717 for Target = 1. This figure shows the curve of the most successful offensive approach with the following conditions: `Ball_speed` $\geq$ 460.790 `Mean_position_y_attacker_4` $\leq$ 113.421, `Mean_movement_15_attacker_3` $\leq$ 56.008 `Mean_std_dev_y_attacker_3` $\geq$ 1.915

40.7% to 51.5%, which may appear moderate in absolute terms, they are notably higher than the baseline scoring rate of 26.8%. This suggests that these combinations of features substantially increase the likelihood of scoring compared to average penalty corners, as also demonstrated by the Cortana Quality score.

Among all subgroup depths analyzed, the depth-3 subgroups stands out with the highest area under the curve (AUC = 0.859) for attacking success. Two features stand out with the attackers as `Bounding_Box_Area_Attackers_T1` $\leq$ 71850 indicates that attackers should not be overly dispersed at the moment of ball reception, supporting compactness for coordinated action around the goal. The condition `Spreadiness_Attacker_Difference_T1` $\geq$ 0.798 on the other hand, introduces a new insight that attacking spreadiness should remain at least 79% of the original formation, ensuring sufficient spatial coverage for reacting to rebound balls which can be played towards any direction.

Coverage	Quality	Target Share	Positives	Condition
51	0.4377	0.4313	22	`mean_movement_15_attacker_3` $\geq 5.83 \land$ `Y_Mean_Attacker_T1` ≤ 140.71
51	0.4377	0.4313	22	`mean_movement_15_attacker_3` $\leq 56.01 \land$ `mean_position_y_attacker_4` ≤ 123.04
59	0.4301	0.4067	24	`Bounding_Box_Area_Attackers_T1` $\leq 71858.0 \land$ `mean_movement_15_attacker_3` ≥ 2.83
33	0.4284	0.5151	17	`mean_position_y_attacker_4` $\leq 113.42 \land$ `Y_Mean_Attacker_T1` ≥ 123.5
41	0.4209	0.4634	19	`mean_position_y_attacker_4` $\leq 113.42 \land$ `mean_movement_15_attacker_3` ≤ 55.47

Table 5: Top five double-feature subgroups associated with successful offensive outcomes (Target = 1)

5.5 Worse Attacking Feature Outcomes

5.5.1 Single feature results

The analysis of the worst-performing offensive subgroups from Table 6, where Target = 0, so no goal is scored, with baseline 73.2%, reveals patterns related to suboptimal individual attacker movements. The condition `mean_position_y_attacker_4` ≥ 113.44 suggests that the stopper of the first couple remains positioned too close to the goal. To improve scoring opportunities, this attacker should stay on the top of the circle. This player is namely a potential stopper and thus increasing the chances to stop the ball for a shot and wait behind the circle for a potential rebound as shown in Picture 20.

Additionally, the condition `mean_movement_35_attacker_3` ≥ 12.04 shows that the shooter of the second couple exhibits excessive movement at critical moment 3 (when the shot is taken). A more stable position may be necessary to maintain accuracy and control during the set play. Finally, the condition `mean_movement_35_attacker_6` ≤ 7.07 highlights insufficient movement by `attacker_6`, typically the one in a straight line from the initiator. This player is expected to tip the ball or actively participate for a rebound. This player does not receive the ball so he should run towards the goal. The relatively low movement observed may reflect a lack of participation, potentially reducing the chances of scoring.

Table 6 demonstrate the performance of these features. While the Cortana Quality scores are modest, ranging from 0.2611 to 0.3044, they still indicate a meaningful deviation from the baseline success rate of 73.2%, particularly given the high target shares between 85.4% and 92.3%. These high target shares in this context reflect a large proportion of unsuccessful outcomes within each subgroup. The relatively lower coverage of some subgroups, especially the third (26 corners), indicates that these patterns are specific but impactful. The ROC curve for the worst-performing single-feature attacking subgroups, shown in Figure 22, yields an area under the curve of 0.708. This ROC-curve highlights limited or excessive movement by key attackers at critical moments, which may reduce scoring opportunities.

Coverage	Quality	Target Share	Positives	Condition
44	0.3044	0.8636	38	mean_position_y_attacker_4 $\geq$ 113.43
41	0.2621	0.8536	35	mean_movement_35_attacker_3 $\geq$ 12.04
26	0.2611	0.9230	24	mean_movement_35_attacker_6 $\leq$ 7.07

Table 6: Worst-performing single-feature subgroups associated with unsuccessful offensive outcomes (Target = 0)

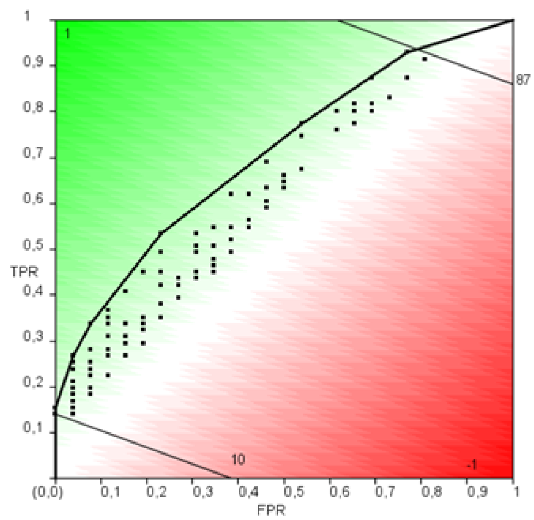


Figure 22: ROC curve with an area under the curve of 0.708 for Target = 0. This figure shows the curve of the worst offensive approach, based on the following conditions:
Mean_position_x_attacker_7 $\leq$ 512.078, Mean_std_dev_y_attacker_3 $\leq$ 1.826,
Mean_movement_35_attacker_6 $\leq$ 7.071, Mean_position_y_attacker_4 $\geq$ 113.435

5.5.2 Multiple feature results

The double-feature subgroup is shown in Table 7. These features mainly focus on the y positioning of the players and the distance moved. The conditions `Mean_Distance_Attacker_Difference_T1` ≥ 9.72 and `Mean_Distance_Attacker_Difference_T2` ≥ 16.50 suggest that the attacking team must deviate a lot from their their initial position, if you want to minimize scoring opportunities. Excessive early deviation from the starting structure must therefore be mitigated.

Additionally, the condition `mean_std_dev_y_att_1` ≤ 12.87 indicates that `attacker_1`, the most left-positioned player who typically runs toward the goal for tip-ins or rebounds, exhibits too little vertical variation. A greater range of movement in the y-direction increases the chance of connecting with a deflection or rebound near the goal, and thus this limited y-movement likely reduces scoring potential.

Coverage	Quality	Target Share	Positives	Condition
40	0.4057	0.925	37	`mean_position_y_attacker_4` $\geq 113.45 \wedge$ `Mean_Distance_Attacker_Difference_T1` ≥ 9.71
40	0.4057	0.925	37	`mean_position_y_attacker_4` $\geq 113.43 \wedge$ `Mean_Distance_Attacker_Difference_T2` ≥ 16.49
51	0.4030	0.8823	45	`mean_std_dev_y_att_1` $\leq 12.86 \wedge$ `mean_position_y_attacker_6` ≥ 109.34

Table 7: Worst-performing double-feature subgroups associated with unsuccessful offensive outcomes (Target = 0)

The ROC curve of depth 2 exhibits the same individual movement and positioning features already discussed in this Chapter. However, by combining these into a double-feature subgroup, the overall classification performance improves considerably, with the area under the curve increasing from 0.708 to 0.767. This combination provides a more reliable signal for identifying ineffective offensive strategies.

5.6 Perspective on data collection

This subsection provides a reflection on the data collection and feature extraction processes used throughout the study. It discusses the performance and challenges encountered during the detection of field elements, player movements, and ball tracking. Each subsection highlights specific limitations, proposed improvements, and overall evaluations of the methods used, offering insight into the reliability and robustness of the extracted data that formed the basis for the subsequent analyses.

5.6.1 Field detection

The detection of the field elements, including the top of the circle and the baseline, was successful. The parabolic fitting approach combined with RANSAC proved to be effective in identifying the top arc of the circle with high accuracy. This method produced a stable and well-formed curve, even in the presence of partial occlusions, lightning changes and noise. Similarly, the baseline detection performed reliably by focusing on horizontal vectors in the lower portion of the frame. The algorithm was able to handle common visual obstructions, such as defenders standing near or over the baseline, by merging shorter, nearly parallel line segments into longer continuous lines. Overall, the field detection process provided robust geometric references that supported the spatial analysis of player positioning and ball movement throughout the penalty corner sequence.

5.6.2 Player movements

The detection of players was generally successful using the YOLO algorithm, especially in frames where players were clearly visible. However, the static top-down camera angle, filmed directly from above the goal, presented certain challenges in identifying players and their team formations. In several cases, especially during the initial moments of the penalty corner, the defensive formation was difficult to distinguish due to the limited visibility of body features and the compactness in which they start. From this perspective, players were often reduced to small objects where only the backs of their shirts were visible, making it difficult to infer their posture or orientation.

Another challenge encountered during the collection of individual player data was the tracker occasionally losing sight of a player. In several video sequences, the tracker would temporarily lose a player and later reassign their identity incorrectly as a new individual. Although such errors occurred infrequently, they affected the overall consistency of the tracking process. A potential solution to this issue could involve a more advanced tracking approach that predicts a lost player's next likely position. This approach would be similar to ball tracking method, as it will compare the possible location with the location of newly detected players. If the predicted and actual positions are within a defined proximity, the new detection can be reassigned the identity of the previously lost player.

Furthermore, assigning players to the correct team based solely on jersey color proved occasionally unreliable, particularly under poor lighting conditions or when players were occluded. Despite these limitations, the combination of YOLO for detection and a tracking algorithm for maintaining temporal consistency performed effectively, providing an overall solid basis for analyzing player movement and team formations.

5.6.3 Ball movement

The detection of the ball, particularly from the inserter to the top of the arc, was generally successful. However, tracking the ball beyond the reception point proved significantly more challenging due to several limitations. The video quality was relatively low, creating a less sharp and clear representation of the ball. Also, a hockey ball itself is small and fast-moving, making it difficult to detect with traditional methods and gets easily mixed with noise of other irrelevant moving pixels and objects. Compounding this issue is the fact that the ball's appearance varies considerably between frames as it will sometimes appear as a small dot, while in motion, it may blur into an elongated shape. This inconsistent visual representation, especially when blended with the field lines or shadowed areas, alters the ball's color and shape. These challenges all make it difficult to isolate the ball using basic color, shape and size thresholding techniques. As a result, tracking the ball during the shot phase and rebound sequences was not reliably possible. Addressing this limitation would likely require a more advanced approach, such as a deep learning-based tracking model trained specifically to recognize the ball under a wide range of conditions, shapes, and motion blurs.

5.7 Perspective on Subgroup Discovery

The use of subgroup discovery proved highly effective in identifying the most successful defensive and offensive approaches during penalty corners. The subgroup discovery tool, developed externally, performed reliably and provided meaningful distinctions between different playing strategies. It allowed for the clear identification of key features associated with successful and unsuccessful outcomes, enabling the formulation of concrete tactical recommendations. Additionally, the tool made it straightforward to explore more complex relationships through depth-2 and depth-3 subgroups, without significantly increasing the workload. The automatic generation of ROC curves and visualization of results further enhanced the analysis, helping to validate findings and support clear communication of the results. Overall, the subgroup discovery process played a critical role in answering the research question and structuring the practical advice for optimizing penalty corner strategies.

6 Conclusion and further research

In this chapter, final summarized recommendations for both offensive and defensive penalty corner strategies are presented. Based on the results of the subgroup discovery analysis, the most significant findings are summarized into practical tactical advice. Furthermore, this chapter briefly reflects on the extent to which the research question has been answered through the application of visual computing and machine learning techniques. Finally, suggestions for future research directions are discussed, identifying areas where further advancements can be made to enhance player tracking, ball detection, and tactical analysis.

6.1 Final Recommendations Defensive Play

For the defending team, success was most strongly associated with dynamic and aggressive defender movements. On individual level, the focus was mainly on flyer, defender 2 and 3. In particular, the flyer (highest outrunner) should run diagonally upwards with a movement angle larger than $-81°$, applying early pressure on the shooter while aligning forehand-to-forehand to intercept shots. Defender 2, positioned near the left of the goalkeeper, should not remain static close to the baseline (around $y \approx 410$), but must actively run outwards, ensuring a position further away from the baseline. Furthermore, the movement of this defender at critical moment 3 should be substantial, with enough variability in positioning, combined with sufficient horizontal movement variability. Lastly, player 3 should avoid staying close to the baseline and showing a lack of movement. This player should especially have a large variation in its y-coordination.

The tactic for most defenders contains horizontal flexibility, as indicated by the importance of a greater standard deviation in their horizontal position. Additionally, defensive units should have moved significantly by critical moment 1, while avoiding excessive dispersion at the moment of the shot. This is important as maintaining a compact formation around the goal blocks direct shooting lanes and prepares for rebounds. Nevertheless, the dispersion of the defenders should always be somewhat larger compared to their initial formation. These findings highlight the importance of fast, proactive defenders, role clarity between flyers and line stoppers, and maintaining compactness during critical phases of the penalty corner.

6.2 Final Recommendations Offensive Play

On the offensive side, the most successful penalty corners emphasized precise player positioning and fast ball movement. Focusing first on individual contributions, the first couple's stopper (attacker 4) should remain high at the top of the circle ($y \leq 113$) to control for rebounds and counters. Meanwhile, the shooter of Couple 2 (attacker 3) should maintain limited but controlled movement in the early phases to stabilize the setup. A fast ball pass by the inserter is critical, creating the ball to reach a movement of over 460 pixels per second which is similar to approximately 21 m/s. This helps minimizing defender reaction time‘and minimizing time to run out towards the top of the circle to pressure the shooter. Additionally, the ball should ideally be injected toward the central-right side of the circle ($x \geq 503$), optimizing opportunities to shoot both left and right. The highest scoring opportunities were found when the first couple executed the shot, rather than the second couple. Therefore, the first shooter must show substantial movement at critical moment 3 (when executing the shot), while the corresponding stopper should remain positioned close to the top of the circle to receive and control the ball effectively.

Considering the team positioning, compactness at reception is required. This has been reflected by a bounding box around attackers smaller than 71850 pixels^2 to enhance scoring chances. Movement towards the goal was essential, with the mean y-coordinate of attackers reaching at least 123-124 at critical moment 1, allowing for greater presence near the goal for rebounds and tip-ins. Additionally, the mean distance between attackers relative to their initial formation should not become excessively large, ensuring sufficient spacing to react effectively to rebounds. These strategies collectively indicate that offensive success hinges on controlled structures, strategic positioning, fast ball injection, and maintaining options for rebound opportunities.

6.3 Conclusion

This study was done to answer the research question: "What would be the most successful approach for a penalty corner, for both the attacking and defending team?" Through the application of visual computing, feature extraction, and subgroup discovery techniques, the analysis successfully identified the key player and team behaviors most strongly associated with successful penalty corner outcomes. By focusing on video footage captured from a top-down static camera angle, detailed data on player positioning, movement patterns, and ball trajectories were obtained. The results have provided clear tactical insights for both offensive and defensive strategies, fulfilling the primary objective of this research.

Furthermore, the study addressed the proposed sub-questions by first identifying which individual player movements contributed most significantly to successful outcomes. Secondly, it distinguished between the most and least effective team behaviors for both attacking and defending situations. These findings were compared against the limited existing literature on hockey strategies, providing additional validation of their tactical relevance. In doing so, this research not only answered the core research question but also generated actionable recommendations that can directly support coaching practices. Ultimately, the integration of machine learning and visual computing proved to be a powerful approach for enhancing tactical understanding in field hockey penalty corners. Especially, considering the largely unexplored research area in this sport, this research represents one of the first studies to systematically apply such methods in this context.

6.4 Further Research

While this research provided valuable insights into the tactical optimization of penalty corners, there are several opportunities for further research. One of the primary limitations of the current study was the relatively small dataset. Future work should incorporate a larger and more diverse set of penalty corners, including footage from different tournaments, more teams, and playing conditions. A broader dataset would improve the generalization of the findings and allow for deeper analysis of variations in tactical approaches.

In terms of data collection, future studies could benefit from improving the ball tracking capabilities. Although the initial injection phase was captured accurately, the ball was not followed after the shot was initiated due to its small size, fast movement, and variation in appearance across frames.

Developing a deep learning-based ball detection model, specifically trained to recognize the ball under various motion and lighting conditions, would allow researchers to track the ball throughout the entire penalty corner sequence. This would enable analysis of the most successful approaches of a shot. Not only would this be beneficial to track the shot trajectory, but it would also help to analyze rebound situations, tip-ins, and defensive clearances after the initial shot.

Additionally, more advanced tracking algorithms for players could be developed to address occasional identity switches or tracking losses, especially under heavy occlusion near the goal. Incorporating predictive tracking models that anticipate player movement trajectories could further enhance tracking reliability.

Finally, future studies could explore more sophisticated subgroup discovery techniques or combine them with other machine learning methods such as clustering or sequence modeling. This would allow for the analysis of temporal patterns in player movements across longer sequences rather than static snapshots, providing even deeper tactical insights into dynamic penalty corner phases.

7 Appendix

Category	Feature Names
Ball Information	`Ball speed`, `Distance to Top of Corner`, `Angle to Top of Corner`, `Corner X`, `Corner Y`
Couple Information	`First Koppel`, `First Koppel 1`, `First Koppel 2`
Players & Teams	`Players`, `Defender Start Formation`
Defender Line Stoppers	`Defender Lijnstop T1 Left/Right`, `T2 Left/Right`, `T3 Left/Right`
Defender Positions (mean)	`X mean Defender T1/T2/T3`, `Y mean Defender T1/T2/T3`
Attacker Count	`Number of Attackers T1/T2/T3`
Attacker Positions (mean)	`X Mean Attacker T1/T2/T3`, `Y Mean Attacker T1/T2/T3`
Attacker Spreadiness	`Curr Spreadiness Attacker T1/T2/T3`, `Spreadiness Attacker Difference T1/T2/T3`
Attacker Distance Metrics	`Mean Distance Attacker Difference T1/T2/T3`
Attacker Bounding Box Area	`Bounding Box Area Attackers T1/T2/T3`
Defender Spreadiness	`Curr Spreadiness Defender T1/T2/T3`, `Spreadiness Defender Difference T1/T2/T3`
Defender Distance Metrics	`Mean Distance Defender Difference T1/T2/T3`
Defender Bounding Box Area	`Bounding Box Area Defenders T1/T2/T3`
Defending Flyer and running out Formation Peak Stats for highest/middle/lowest out runner at T1/T2/T3	`% uitloop`, `% x from the reception of the ball`, `speed`, `angle_prev`, `angle_top`

Table 8: Overview of all the generated features for subgroup discovery

Category	**Feature Names**
Individual Defender Mean Position (X/Y) for defender 1-5	`mean_position_x/y_defender` (1–5)
Individual Defender Std Dev (X/Y) for defender 1-5	`mean_std_dev_x/y_defender` (1–5)
Defender Movement for defender 1-5	`Individual mean_movement_15/25/35_defender` (1–5)
New Defender Metrics (X/Y) for new defender 1-2	`mean_position_x/y_new_def`, `mean_std_dev_x/y_new_def`, `mean_movement_15/25/35_new_def`
Individual Attacker Mean Position (X/Y) for attacker 1-8	`mean_position_x/y_attacker` (1-8)
Individual Attacker Std Dev (X/Y) for attacker 1-8	`mean_std_dev_x/y_attacker` (1-8)
Individual Attacker Movement for attacker 1-8	`mean_movement_15/25/35_attacker` (1-8)
New Attacker Metrics for attacker 1-2	`mean_position_x/y_new_att_1/2`, `mean_std_dev_x/y_new_att_1/2`, `mean_movement_15/25/35_new_att_1/2`
Unknown Players 1-3	`mean_position_x/y_unknown_1/3`, `mean_std_dev_x/y_unknown_1/3`, `mean_movement_15/25/35_unknown_1/3`
Label / Goal	`Target`

Table 9: The continuation of all generated features for subgroup discovery

References

[AAAJ21] Imran Ahmed, Misbah Ahmad, Awais Ahmad, and Gwanggil Jeon. Top view multiple people tracking by detection using deep sort and yolov3 with transfer learning: within 5g infrastructure. *International Journal of Machine Learning and Cybernetics*, 12, 11 2021. doi:10.1007/s13042-020-01220-5.

[Ass22] National Field Hockey Coaches Association. New coach toolkit: Field hockey penalty corners, 2022. URL: https://nfhca.org/resources/.

[BIKP19] Matija Buric, Marina Ivašić-Kos, and Miran Pobar. Player tracking in sports videos. pages 334–340, 2019. doi:10.1109/CloudCom.2019.00058.

[BJE+09] Y. Benezeth, P. M. Jodoin, B. Emile, H. Laurent, and C. Rosenberger. Review and evaluation of commonly-implemented background subtraction algorithms. *MoBT8.10*, 2009. URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4760998.

[Cad09] M. Cadík. Perceptual evaluation of color-to-grayscale image conversions. *Computer Graphics Forum*, 27(7):1745–1754, 2009. doi:10.1111/j.1467-8659.2008.01319.x.

[Can86] John F. Canny. A computational approach to edge detection. Technical Report 6, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1986.

[CM13] Bodhisattwa Chakraborty and Sukadev Meher. A real-time trajectory-based ball detection-and-tracking framework for basketball video. *Journal of Optics*, 42:156–170, 01 2013. doi:10.1007/s12596-012-0108-7.

[Der10] Konstantinos G. Derpanis. Overview of the ransac algorithm. Technical report, Department of Computer Science and Engineering, York University, 2010. URL: http://www.cs.yorku.ca/~kosta/CompVis_Notes/ransac.pdf.

[DGHW00] T. Darrell, G. Gordon, M. Harville, and J. Woodfill. Integrated person tracking using stereo, color, and pattern detection. *International Journal of Computer Vision*, 37:175–185, 2000.

[dH24] The Fédération Internationale de Hockey. Interactive hockey data hub, 2024. URL: https://www.fih.hockey/datahub.

[DMZ11] D. A. D'Ippolito, J. R. Myra, and S. J. Zweben. Convective transport by intermittent blob-filaments: Comparison of theory and experiment. *Physics of Plasmas*, 18(6), 2011.

[DT05] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 886–893, 2005. doi:10.1109/CVPR.2005.177.

[ea87] Stephen M. Pizer et al. Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, 7(2), 1987.

[ea00] Luis Alvarez et al. Reliable estimation of dense optical flow fields with large displacement. *International Journal of Computer Vision*, 39:41–56, 2000.

[Edu20] Field Hockey Education. Field hockey penalty corner defense — positions responsibilities, 2020. URL: https://www.youtube.com/watch?v=vLK_uwQvB4A.

[FB81] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. doi:10.1145/358669.358692.

[(FI23] International Hockey Federation (FIH). Penalty corner rules trials, 2023. URL: https://www.fih.hockey/.

[Gar09] J. Garganta. Trends of tactical performance analysis in team sports: bridging the gap between research, training and competition. *Centre of Research, Education, Innovation and Intervention in Sport (CIFI2D) Faculty of Sports University of Porto Portugal*, 9:81–89, 2009.

[GLO16] Yanming Guo, Yu Liu, and Ard Oerlemans. Deep learning for visual understanding: A review. *Neurocomputing*, 197:27–48, 2016.

[Har04] M. Harville. Stereo person tracking with adaptive plan-view templates of height and occupancy statistics. *Image and Vision Computing*, 22(2):127–142, 2004.

[HC nd] HC Nuth. Spelhervattingen - de verdedigende strafcorner, n.d. URL: https://www.hcnuth.nl/files/library/vakblad%20hockey/spelhervattingen%20-%20de%20verdedigende%20strafcorner.pdf.

[Hea19] E. Herold and et al. Machine learning in men's professional football: Current applications and future directions for improving attacking play. *International Journal of Sports Science & Coaching*, 14(6), 2019. doi:10.1177/17479541198793.

[Hig16] David Higham. Finding the optimal background subtraction algorithm for euro-hockey 2015 video. *Procedia Engineering*, 147:637–642, 2016. URL: https://www.sciencedirect.com/science/article/pii/S1877705816307081, doi:10.1016/j.proeng.2016.06.225.

[Hou62] Paul V. C. Hough. Method and means for recognizing complex patterns, 1962. URL: https://patentimages.storage.googleapis.com/9f/9f/f3/87610ddec32390/US3069654.pdf.

[JEL+22] Peiyuan Jiang, Daji Ergu, Fangyao Liu, Ying Cai, and Bo Ma. A review of yolo algorithm developments. *Procedia Computer Science*, 199:1066–1073, 2022.

[KGvDP21] Arno Knobbe, Jeremie Gobeil, Richard van Dijk, and Willem Jan Palenstijn. Subdisc: A subgroup discovery toolkit. https://github.com/SubDisc/SubDisc, 2021.

[Kor85] Richard E. Korf. Depth-first iterative-deepening: An optimal admissible tree search. *Artificial Intelligence*, 27(1):97–109, 1985. doi:10.1016/0004-3702(85)90084-0.

[KV10] Tarun Kumar and Karun Verma. A theory based on conversion of rgb image to gray image. In *International Conference on Computer Graphics, Vision and Information Security*, 2010.

[KYYW22] I. Kurniastuti, E. N. I. Yuliati, F. Yudianto, and T. D. Wulan. Determination of hue saturation value (hsv) color feature in kidney histology image. *Journal of Physics: Conference Series*, 2157:012020, 2022. doi:10.1088/1742-6596/2157/1/012020.

[lav99] Inductive logic programming, 9th international workshop, ilp-99, bled, slovenia, june 24–27, 1999, proceedings. In Nada Lavrač, Saso Dzeroski, and Peter Flach, editors, *Lecture Notes in Artificial Intelligence*, volume 1634 of *Lecture Notes in Computer Science*. Springer, 1999.

[LKFT04] Nada Lavrač, Branko Kavšek, Peter Flach, and Ljupčo Todorovski. Subgroup discovery with cn2-sd. *Journal of Machine Learning Research*, 5:153–188, 2004.

[McK00] Stephen McKenna. Tracking groups of people. *Computer Vision and Image Understanding*, 80(1):42–56, 2000.

[MDPH23] Dimitrios Meimetis, Ioannis Daramouskas, Isidoros Perikos, and Ioannis Hatzilygeroudi. Real-time multiple object tracking using deep learning methods. *Neural Computing and Applications*, 35:89–118, 2023. doi:10.1007/s00521-021-06391-y.

[MSAGS06] Rafael Muñoz-Salinas, Eugenio Aguirre, and Miguel García-Silvente. People detection and tracking using stereo vision and color. *Image and Vision Computing*, 25(6):995–1007, 2006.

[MVC20] Clara Meister, Tim Vieira, and Ryan Cotterell. Best-first beam search. *Transactions of the Association for Computational Linguistics*, 8:795–809, 2020. doi:10.1162/tacl_a_00346.

[OW13] Wanli Ouyang and Xiaogang Wang. Joint deep learning for pedestrian detection. *Computer Vision and Pattern Recognition (CVPR)*, 2013.

[PSAR21] Narinder Singh Punn, Sanjay Kumar Sonbhadra, Sonali Agarwal, and Gaurav Rai. Monitoring covid-19 social distancing with person detection and tracking via fine-tuned yolo v3 and deepsort techniques. *Sustainable Cities and Society*, 2021. doi:10.48550/arXiv.2005.01385.

[SHJ15] Ajmal Shahbaz, Joko Hariyono, and Kang-Hyun Jo. Evaluation of background subtraction algorithms for video surveillance. pages 1–4, 2015. doi:10.1109/FCV.2015.7103699.

[Spo24] Sportplan. Short corner defence (3:1), 2024. URL: https://www.sportplan.net/drills/Hockey/Set-Pieces/Short-Corner-Defence-3-1-h103scd2.jsp.

[SQP02] Shamik Sural, Gang Qian, and Sakti Pramanik. Segmentation and histogram generation using the hsv color space for image retrieval. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, volume 2, pages 589–592, 2002.

[SV13] Andrews Sobral and Antoine Vacavant. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. *Computer Vision and Image Understanding*, 122:4–21, 2013.

[TKBM99] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintenance. In *International Conference on Computer Vision (ICCV)*, pages 255–261, 1999. URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=791228.

[YLXT07] Xinguo Yu, Hon Leong, Changsheng Xu, and Qi Tian. Trajectory-based ball detection and tracking in broadcast soccer video. *Multimedia, IEEE Transactions on*, 8:1164 – 1178, 2007. doi:10.1109/TMM.2006.884621.

[Zha19] Zhong-Qiu Zhao. Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*, 30(11):3212–3232, 2019. doi:10.1109/TNNLS.2018.2876865.

[ZXT10] Zhengtao Zhang, De Xu, and Min Tan. Visual measurement and prediction of ball trajectory for table tennis robot. *IEEE Transactions on Systems, Man, and Cybernetics*, 59(12):3195–3205, 2010.