AALBORG UNIVERSITET
STUDENTERRAPPORT

School of Medicine and Health (SMH

Frederik Bajers Vej 7D, 9220 Aalborg Øst

DK - 9220 Aalborg Øst
Tlf. +45 99 40 37 48

Title:	**Developing an Algorithm for Analysing Passes in a Football Match using spatio-temporal data**
Semester:	**10th semester**
Project period:	**1st Feb – 6th June 2019**
ECTS:	**30**
Supervisor:	**Anderson de Souza Castelo Oliveira**
Project group:	**19gr10208**

Jacob Gandrup

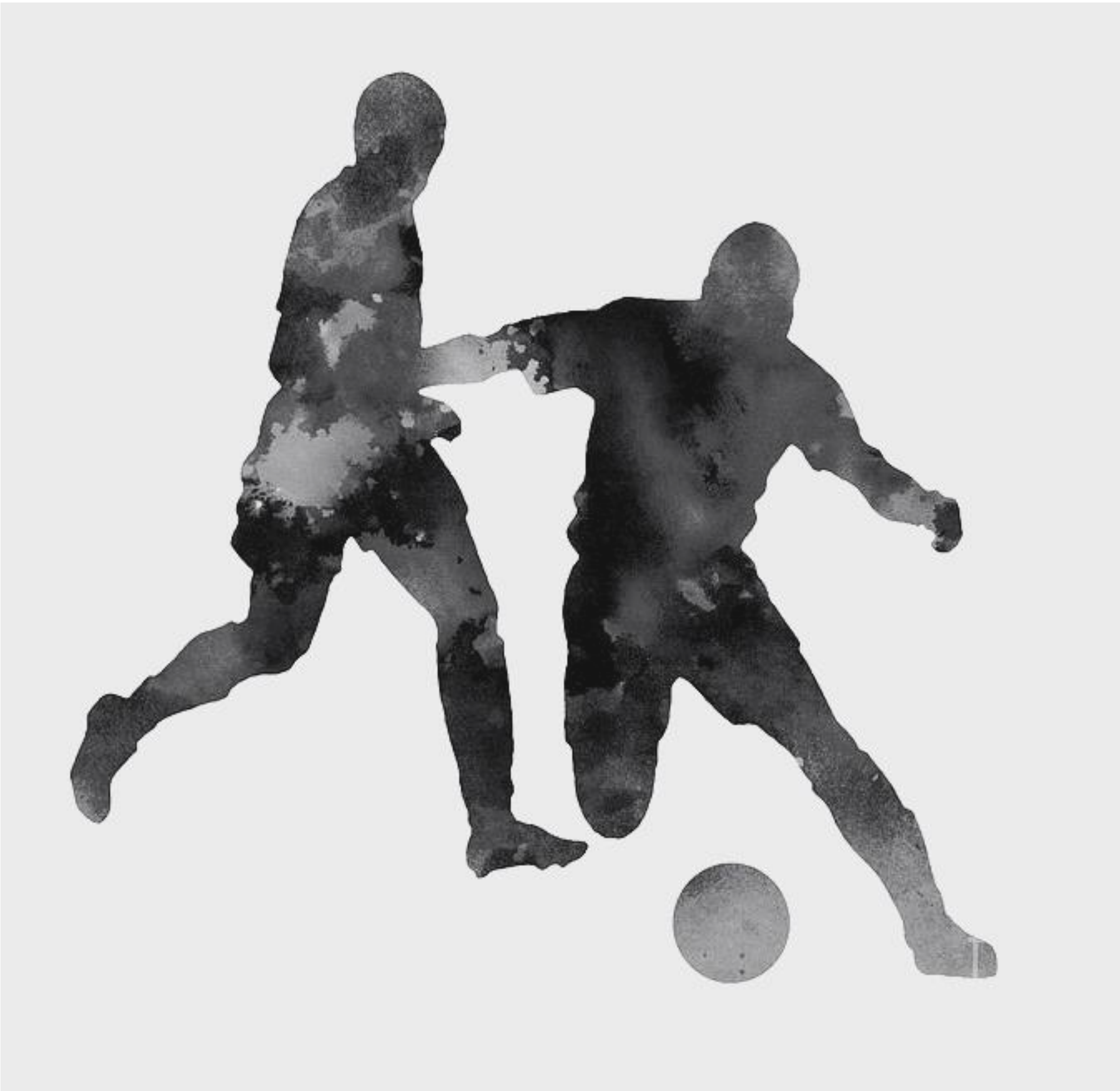
Morten Blegvad Jørgensen

Analysing performance in football is an important task for coaches and by the substantial development of technologies which have become a consistent component in sports, a huge amount of data is collected every day. This data can potentially provide information of the tactical or technical performance, for coaches to evaluate. However automated methods need to be developed in order to extract this information.
The purpose of this thesis was to develop and evaluate an automated algorithm to detect and classify passes in a football match using spatio-temporal data. The proposed **method** consisted of multiple steps including the application of suitable filters, pass detecting procedures and classifying pass based on length and direction which were detectable by the algorithm. The data consisted of one half of a football match played in the Danish Superliga and was collected by an optical tracking system.
The results showed a substantial accuracy of the detected passes by the algorithm compared to the ground truth created by human observers (F-score = 0.80). **In conclusion** the method showed to be a useable tool for the detection of passes, however certain limitations regarding constraints in the algorithm and the data needs to be considered in future research.

Characters: [62320]
Pages: [45]
Appendix: [8]

Ved at underskrive dette dokument bekræfter hvert enkelt gruppemedlem, at alle har deltaget lige i projektarbejdet og at alle således hæfter kollektivt for rapportens indhold.

Developing an algorithm for analysing passes in a football match using spatio-temporal data



Master's thesis

By Morten B. Jørgensen & Jacob Gandrup

Acknowledgement

The authors of this thesis would like to express our gratitude to Football Club Midtjylland, Denmark, for providing the data used in present thesis. Furthermore, we would like to thank our supervisor Anderson de Souza Castelo Oliveira for excellent guidance throughout this thesis.

Table of content

Terminology

The following terminology includes a definition of different terms used in this thesis which are related to the fundamental actions of a football match. The definitions of these terms are based on the study of Liu et al.[1].

Pass	An intentional played ball from one player to another player.
Clearance	An attempt to kick the ball away from an area close a team's own goal field.
Dribble	An attempt to move the ball past an opponent player. If it is successful, the player will maintain possession of the ball.
Interception	When an opponent intercepts a pass and thus has possession of the ball.
Through Ball	Another type of pass which is played through the last line of defense to a team member.
Shot	An attempt to score a goal made by any part of the body, apart from the arms, either on or off the goal. The outcomes of a shot could be a goal, shot on target, shot off target or a blocked shot.
Wall pass	A sequence of two passes between two players. One player passes the ball to a team member who returns the ball to the first player by a new pass with one touch.

Literature search

Following describes the literature search procedure related to the field of interest of this thesis regarding the development of an algorithm to detect and classify passes in football by spatio-temporal data of the ball and players.

The research included in this thesis has its origin in different fields, such as machine learning, sports science and statistics. The research procedure was mainly conducted on Primo (Aalborg University Library) as it grants access both to various databases, i.e. PubMed, Scopus and Web of Science and to different types of literature, i.e. textbooks, journals, scientific articles, manuscripts and web pages. Additional searching tools were applied, such as Google Scholar and the chain search procedure was conducted on relevant research to utilize and include new relevant references. Furthermore, keywords found in new literature were applied for future research.

To specify the relevant research, the following key words were used:

‘Tracking data’, ‘spatio-temporal data’, ‘event detection’, ‘football passes’, ‘association football’, ‘tactical football approaches’, ‘machine learning’ etc.

1 Introduction

Association football, commonly known as football (soccer), is the most popular sport in the world based on media coverage, economy and its more than 250 million players [2]. A lot of money is earned and spent due to this popularity, and the total revenue of the top 20 richest clubs in the world was recorded to €8.3bn in 2019 [3].

Being successful is a crucial criterion for the revenue of top football clubs as a total of €2.04bn will be distributed to the teams competing in the biggest European club tournament, Union of European Football Association (UEFA) 2018/2019 Champions League and the 2018 UEFA Super Cup [4]. To achieve the objective of being successful, football clubs and teams are constantly investigating and developing new approaches of how to analyze football and hereby improve performance of both the teams and the individual players [5].

The incorporation of technologies to improve sports, including football has in recent years become a new and consistent component in the field of sports research. Especially the development of tracking systems has experienced a substantial increase in football due to innovations in image processing and object detection [6]. This has led to a higher amount and more precise tracking data of both the ball and players which makes it possible to achieve a better understanding of the underlying actions in football [7][8]. Tracking data is often combined with event data which is a representation of the important occurring events in a match, such as freekicks, goals and ball possession etc. [1], [6]. This event data is manually noted by people who spend around ten hours per game [9]. However, the current processing and analysis methods in the research of football are limited and thus encounter difficulties in handling and quantifying this data due to the complex and dynamic nature of the game of football [10], [11]. This often results in the data mainly being used to observe physical loading and performance, such as the total distance covered, and high intensity runs. Valuable information to analyze the tactical approaches of a team and the complex individual player requirements, i.e. the passing ability and the tactical level, is therefore not being extracted [8], [10], [12].

Previously, football analysis was done manually and mostly consisted of simple event data such as number of shots, tackles, passes, crossings and goals, describing *what* happened, and not *where, why* and *by whom* [10]. However, this procedure does not cover the complex aspects of the game, is more time consuming and becomes impractical as the volume of the data increases [6], [13]. The development of automated methods to exploit the potential of improved spatial-temporal data and to discard the need for manual notation could therefore be beneficial. The obvious advantage will be a reduction of workload, but it will also improve the knowledge of player evaluation, decision-making for coaches and talent scouts to make informed choices, managing training and improve transfer decisions. Furthermore, it could support the development of tactical strategies and contribute to an improved viewing experience of televised matches [6], [10].

When looking more specifically at the evaluation of player performance, one parameter which is often of huge interest, is the player's ability to perform certain types of passes combined with the quality and impact of these passes [14]. Furthermore, passes are the most frequently occurring event in football [15], which states the importance of this parameter. If the previously mentioned methods of analyzing spatio-temporal data in relation to passes were to be developed, coaches and scouts would be able to track the development of the player's passing level, spot new talents and achieve valuable information about the passing strengths and weaknesses of opponent teams [5].

Previous studies have conducted methods to evaluate the quality of passes based on spatio-temporal data [6]–[8], [16], [17]. These studies based the quality of a pass on different assumptions of the most effective passes, such as the probability of creating goal scoring opportunities [16], [17] or how much a pass influenced the organization of the opponent team based on passing characteristics as velocity, length and direction [8]. Chawla et al. evaluated the quality of a pass by an ordinal measure of "Good, Ok or Bad" by multiple variables, such as player position, velocity and the dominant area of a player [18]. It is possible that a tool to objectively assessment passes is more valuable for coaches, as the practiced type of passes, such as long and direct, short and controlled and the direction of a pass, which defines the playing style, is highly based on individual preferences of the coaches [14], [19]. However, no previous studies to the author's knowledge have provided the characteristics in relation to specific types of passes. Furthermore, before being able to evaluate the performance of passes based on tracking data of the players and the ball, a method to detect and identify the type of pass needs to be developed. Irnai et al.[9], developed a model to detect different evens, including passes. However, they did not assess the detected passes in relation to performance.

The aim of this thesis is thereby to develop and evaluate an algorithm to automatically detect passes based on trajectory data of the ball and players and event data from a football match and furthermore to classify different types of passes, based on the length and direction, which are detectable by the algorithm.

2 Theory

This chapter contains the basic theory and background knowledge which this thesis is based on.

2.1 TRACAP Optical tracking and spatial-temporal data

The data used as the input for the algorithm in the present thesis were collected by the TRACAP Optical Tracking system [20]. This is a non-intrusive sports analysis tool that uses a 3D localization technology for the tracking of objects by two compact TRACAB Super-HD camera units. The company, ChyronHego, states that it "*provides the most accurate, independently-verified, consistent, and reliable sports performance data*". The cameras use stereo technology to capture the whole playing area from various angles with a sampling frequency of 25 Hz and with a maximum delay of three frames in real-time. The software analyzes each image by disclosed tracking algorithms to extract spatial-temporal data, which can be divided into spatial trajectories and event logs. Additionally, it can provide performance metrics, such a velocity, acceleration, team formations and running distance [20].

2.1.1 Spatial trajectories

A spatial trajectory is defined by the path that a moving object (players and ball) follows through space (playing field) as a function of time. Thus, denoted as $\{(x_1, y_1, t_1), (x_2, y_2, t_2), \ldots, (x_N, y_N, t_N)\}$ where x_i, y_i, equals the coordinates of the objects in motion at time t_i and N as the total number of elements in the series [21]. The spatial trajectories extracted from the TRACAP system consist of X and Y positions of players and referees and X, Y and Z positions of the ball in centimeter from origo at given time stamps (Figure 1). The Z positions are defined as the height of the ball from the ground [20].

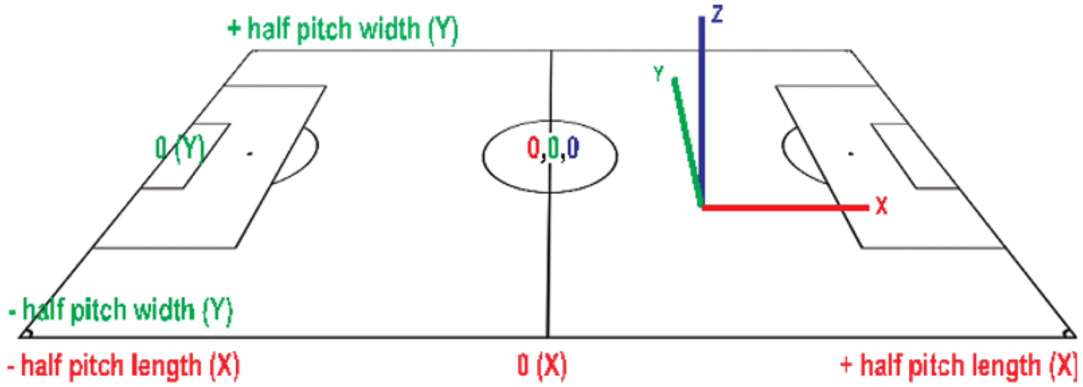


Figure 1: The coordinate system of the pitch [20].

However, the ball and player trajectories captured by tracking systems in football also provide challenges. Since the players move around for rather a long amount of time within in the playing field, which is a spatially constrained area, the player trajectories can be dense. Additionally, the players frequently change their movement direction and velocity to adapt to the progress of the game, which can make the trajectories cluttered when visually inspecting them, as illustrated in figure 2 [5].

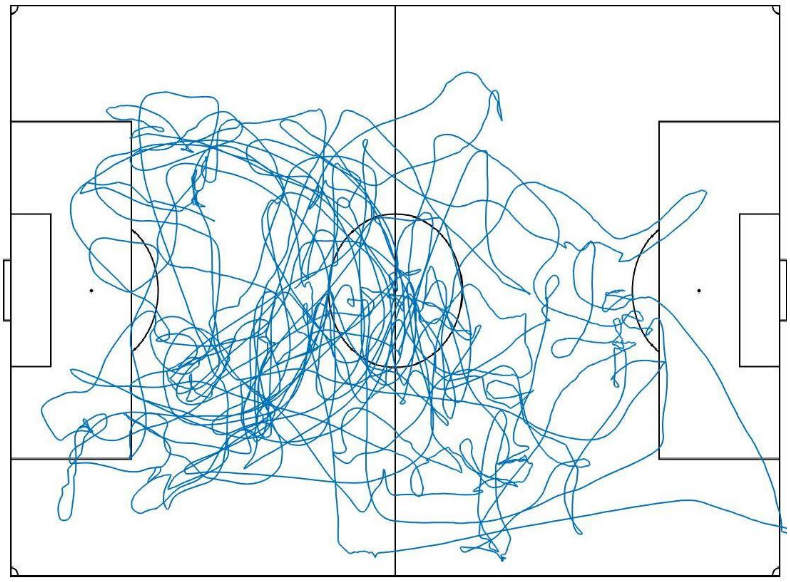


Figure 2: Example of a filtered movement trajectory of a player after one half in a football match.

2.1.2 Event data

The events captured by TRACAP include event flags of the team in possession of the ball and when the ball is either in open-play(alive) or not (dead) [20]. The event log is characterized by three different parameters: the time stamp of the event, the type of the event and the players that are involved. It is hereby possible to segment a match into intervals by the different types of events. An example could be to divide a match into open-play stages based on the stops in a match such as freekicks, fouls, etc. The main feature that makes event data different from spatio-temporal data is the qualitatively aspect, since event data are not dense and are only collected at a certain event. However, it includes more semantic information [5].

2.2 Ball possession and passes

2.2.1 Ball possession

For a team to win a football match, the team is required to score more goals than the opponent team. This means that the ability to score goals is the most crucial determinant of a successful football team [22], [23]. Scoring a goal is often a result of a goal scoring opportunity being created.

Various situational actions in a football match can lead to the creation of a goal scoring opportunity. An example of such a scenario is a standard situation (corner kick, free kick and throw in) where the ball is aimed at a team member positioned inside the opponent's goal field. Another example, from an open-play scenario is an interception or a tackle where the opponent team is caught in a disorganized situation. This creates the opportunity for a counterattack which involves moving the ball towards the opponent's goal as fast as possible. Other scenarios occur when one team has possession of the ball and then produces a goal scoring opportunity by a through ball or a sequence of short passes between multiple players to disrupt the defensive organization of the opponent [24], [25].

The last scenario concerns ball possession, which is a highly investigated performance parameter in the analysis of football performance [26]. Previous studies have found that successful teams have both higher ball possession and the ability to maintain control of the ball for longer periods of time during a game. However, maintaining a higher ball possession is not a direct criterion for success, as teams tend to have less ball possession in games which they have won compared to lost games. This phenomenon can be explained by tactical dispositions which depend on the team being in the lead or behind [25].

Ball possession is furthermore a parameter determined by a process of interactions dependent on multiple contextual factors, such as tactical strategies, the venue, the quality of the teams and the current score [27]. Ball possession is defined by the beginning of the moment a player can perform an action with the ball, typically after receiving a pass from another player. It ends when that player is not able to perform an action with the ball any longer, usually after the player either passes the ball to another player, gets tackled by an opponent or a game interruption [13].

2.2.2 Passes

The ability to perform successful passes is an essential factor in order to maintain ball possession or to create scoring opportunities, hence making it an essential technical skill in football [5], [14]. Cintia et al.[15], found that passes are the most frequent occurring event in football matches. More than 300,000 passes were thus performed in a total of 444 games from the 2014 World Cup and the Italian Serie A 2013/2014 season (figure 3). The ability to perform passes of a certain quality are therefore a significant parameter when the overall performance players are evaluated [5], [8].

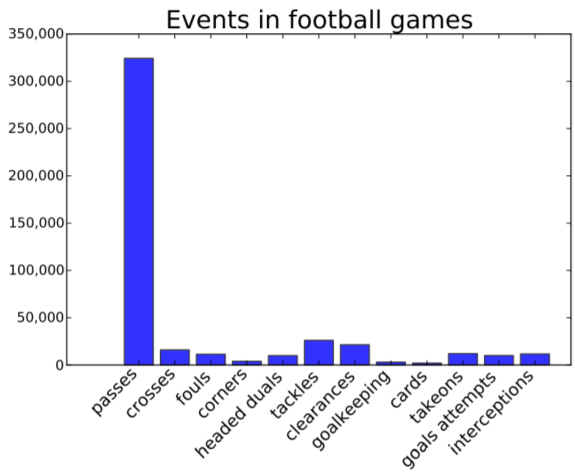


Figure 3: The total number of events per category which occurred in a total of 444 games (64 games player in the 2014 World cup and 380 games player in the Italian Serie A) 2013/2014 [15].

According to literature, there are certain parameters of a pass that are recognized as predictors for success, i.e. the number of passes, rate of completed passes, the number of passes in a forward direction and the length of a pass [14], [19], [28]. The tactical approach of different teams can furthermore be determined by analyzing the parameters of the different types of passes, as some teams try to create opportunities by long direct passes, whereas other teams aim for the usage of short passes, hence a more possession based style of play [19].

Teams that practice the short possession tactic, utilize a more reliable procedure of passing the ball between the players on the team compared to teams that practice long passes which requires a higher skill level and are often less precise. Prioritizing a short passing tactic includes other benefits as the team often maintains ball possession for a longer period. This is important both in relation to creating scoring opportunities and in relation to denying ball possession of the opposing team. However, this style of play places high demands on the players' stamina, speed, creativity and technical passing level. [14], [19] In opposition, teams practicing a more direct play style aim to create goal scoring opportunities by as few passes as possible, often associated as the counter attacking play style [19]. The parameters which defines a direct play style is long passes defined by the direction and the length [14]. The long forward pass is generally of higher risk of missing the intended receiver due to a certain requirement of velocity, precision and vision of the opponent [29]. Jones et al.[25], investigated which type of play style that is the most successful. They found that successful teams in English Premier League 2001-2002 season both had a greater share of ball possession and longer periods of ball possession. However, as mentioned in the section 2.2.2, having ball possession is not a direct criterion of success. The importance of including the tactical approach in relation to the context of the game, was shown by Tenga et al.[30]. They found that a direct counter-attacking style of play is the most effective concept against teams with a disorganized defense, but not when the opponent was well organized [30].

Various studies have investigated different approaches of quantifying the quality and impact of passes based on tracking data. The majority of these studies only use assessment methods based on the probability of creating goal scoring opportunities [11], [26]. Hence, a pass is merely evaluated as a “good pass”, if it creates a goal scoring chance or increases the possibility for a shot on goal. Because of this, forward passes are the only passes being classified as good passes despite sideways or backwards passes potentially provide more value.

A possession-based style of play is, as mentioned previously, advantageous against organized teams, and this play style often includes passes in sideway or backward direction pass. Even though these passes do not necessarily result in a scoring chance, both types of passes can disrupt the defensive organization of the opposition and thereby create space for scoring opportunities or aim to relieve pressure of the opponent [8]. Furthermore, Fernandez & Bornn, [31], stated that space creation and the disruption of the opponent’s organization are some of the most crucial approaches in modern football [31]. In addition, Goes et al.[8], found that a pass with a length of 19 to 30 *m* and a velocity of 10.7 *m/s* had the most impact on the opponent’s organization. Horton et al.[5], also considered the distance, angle and velocity of a pass as fundamental parameters to characterize a pass.

3 Method

This chapter describes the proposed method related to the aim in the current thesis of developing an algorithm to automatically detect passes based on trajectory data of the ball and players and event data from a football match and furthermore to classify different types of passes which are detectable by the algorithm. The main components in this method include a classification of different passed, a step by step description of the procedure in the algorithm and a notation protocol to create a ground truth.

3.1 Data description

The data used in this thesis consisted of tracking data of the ball and players, event data and a televised video recording of a football match played in the Danish Superliga 2018/2019 season at Vejle Stadium between Vejle Boldklub and Football Club Midtjylland. The tracking data were collected by the TRACAP optical tracking system [20] which records the X and Y positions of the players and the X, Y and Z positions of the ball at 25 Hz, as previously described in section 2.1.

The data from the match were divided into a test set and an experimental set, which consisted of the first and second half, respectively. The test set was used to develop the algorithm, and the experimental set was used to evaluate the algorithm. The only data included from the event data was the event which that states when the ball is either alive or dead. The possession flags were ignored as a part of the algorithm involved detecting ball possession and furthermore because the algorithm was aimed to be as independent from other input than tracking data of ball and player as possible. Furthermore, this decision was made to make it possible to implement the algorithm on tracking data collected at other settings, such as a training session, and thereby with no need for manual notation of events. An overview of the structure of the method, which will be described in following sections, is presented in figure 4.

The TRACAP system [20] provided additional input variables: Match ID, Team type, Pitch dimensions, Jersey number (Player ID) and Periods. Table 1 presents a description of the most important input variables included in the experimental data set. The data was imported, preprocessed, analyzed and visualized using MATLAB (The MathWorks®, Inc., Natick, MA, USA).

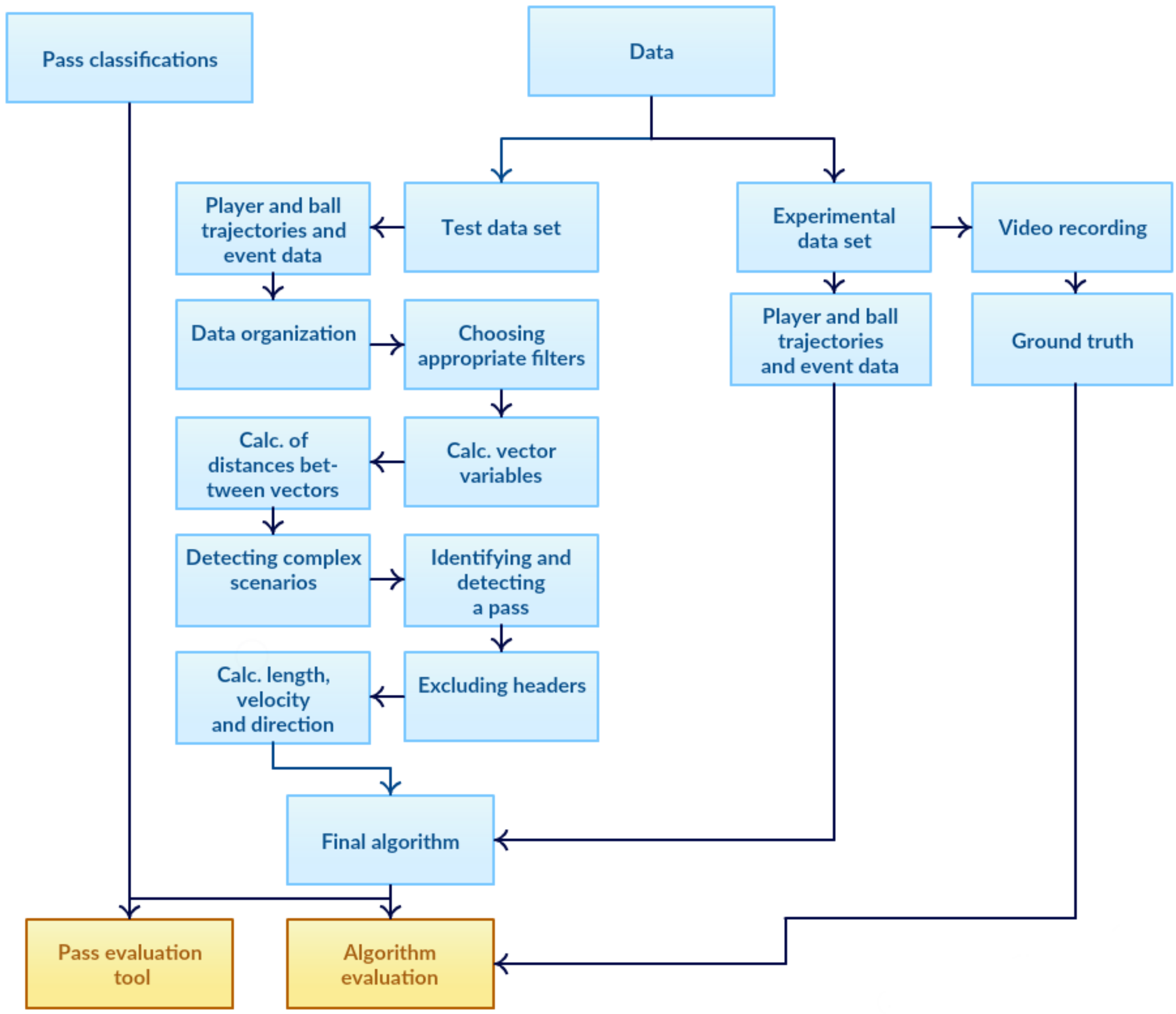


Figure 4: Overview of the structure of the proposed method. Calculation is denoted by calc.

Table 1: The most important input variables included in the experimental data set provided by the TRACAP software.

Variable	Description
Match ID	Unique ID for the match
Team type	Home or away team and referee
Pitch dimensions	The width and length of the pitch
Player ID	The players jersey number
Player position	X and Y positions
Ball position	X, Y and Z position
Event flags	Alive flag: Alive or dead
Periods	Start- and end-frame from the first and second half

3.2 Classifying different types of passes

This section includes the process of classifying the different types of passes which were identifiable by the algorithm. The purpose of these classifications were to create an objective measure of each pass which included characteristics from the different tactical approaches in football, where especially the length and direction of a pass was found to be important determinants for the playing style of a team [14], [19], [27].

3.2.1 Length of a pass

The passing length was defined as either short, medium or long based on research of passing exercises in football [32] and the study by Rösch et al.[33], who evaluated the performance of 588 football players by a variety of different tests, including a short (11m) and a long passing (31m) test. A short pass was therefore defined by a length of maximum 11m and a long pass by a minimum of 31m. No test of passes between the distance of 12 to 31m was conducted by Rösch et al.[33], and a pass with this length was therefore defined as a medium pass.

3.2.2 Direction of a pass

The passing direction was defined as either a backward, sideway and forward pass based on the study by Goes et al.[8], who calculated the displacement of the ball and the angle of the pass in by the X and Y positions of the passer and the receiver in relation to the longitudinal and horizontal axis. A backward pass was defined by the ball moving towards a team's own goal by an angle of -45° to -45° on the longitudinal axis A sideway pass was defined by the ball moving towards the side of the sidelines of the pitch by an angle between -45° to 45° in relation the horizontal axis. A forward pass was defined by the ball moving towards the opponent's goal with an angle of 45° to 45° on the longitudinal axis [8].

Based on a combination of the length and direction parameters of a pass, a total of nine different types of pass were classified:

1. **The backwards short pass**: defined by a maximum distance of 11 m played in the backward direction of -45° to -45°.

2. **The backwards medium pass**: Defined by a distance between 12 to 30 m played in the backward direction of -45° to -45°.

3. **The backwards long pass:** Defined by a distance between 31 to 90m played in the backward direction of -45° to -45°.

4. **The sideways short pass:** Defined by a maximum distance of 11 m played in the sideway direction of -45° to 45°.

5. **The sideways medium pass:** Defined by a distance between 12 to 30 m played in the backward direction of -45° to 45°.

6. **The sideways long pass**: Defined by a distance between 31 to 90m played in the sideways direction of -45° to 45°.

7. **The forwards short pass**: Defined by a maximum distance of 11 m played in the backward direction of 45° to 45°.

8. **The forwards medium pass**: Defined by a distance between 12 to 30 m played in the backward direction of 45° to 45°.

9. **The forwards long pass**: Defined by a distance between 31 to 90m played in the backward direction of 45° to 45°.

3.3 Visualization tool

In the development and testing process of the algorithm, it was necessary to visualize the positions of the players and the ball. An animation of the game was therefore created by the plotting function in MATLAB (The MathWorks®, Inc., Natick, MA, USA), which is illustrated in figure 5 and presented in Appendix 5. This allowed for visualization of the different positions of the ball and players, passes, synchronization, notation of the data and the video. Furthermore, it was used to test the different approaches in the algorithm which will be elaborated in the following sections.

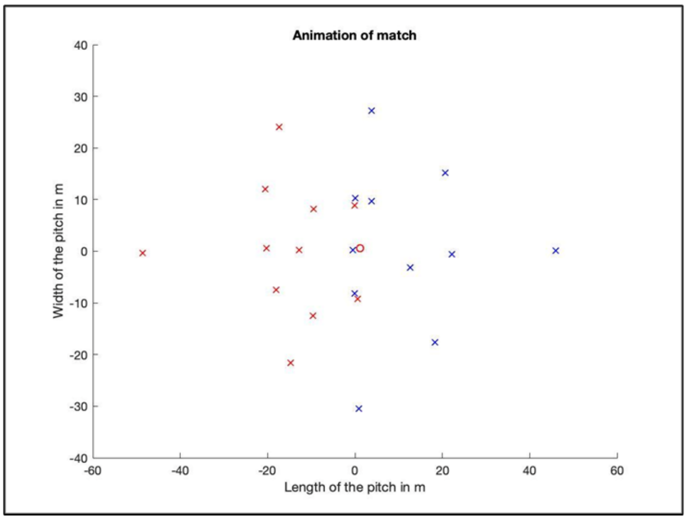


Figure 5: Visualization tool. The blue and red crosses denote the positions of the players on the home team and away team, respectively. The red circle denotes the position of the ball. The x- and y axis represent the distance from the centerline to the back- and sideline in meters, respectively.

3.4 Algorithm procedure

This section includes a description of the multiple steps in the algorithm, which can be characterized as a preprocessing procedure of the trajectory and event data by filtering, segmentation, indexing and computation of variables to detect and label each pass based on the classifications presented in section 3.2.

3.4.1 Initial steps

3.4.1.1 Data organization

The initial step in the algorithm involves a procedure of organizing the following of the input data: Match ID, Team type, Player ID and periods. The match is set to $M = M_1, \ldots, M_i$, the periods in the match to H = {*First half, Second half*} and the players who played in the match to $P = \{p_1, p_2, \ldots, p_i\}$. Each player is additionally denoted as playing for either the home team or away team and to the playing position in the team formation. The pitch dimensions are furthermore included to define the playing field. The imported data was found to be unstructured in defined manner in relation to the movement trajectories of the players and a MATLAB script was created to re-organize the data. This script is presented in Appendix 2.

3.4.1.2 Filtering

For further processing it was required to smooth the raw trajectories of the ball and players. It was especially crucial to find a filter that fits the high accelerations and velocities which the ball in football often is exposed to. The raw ball trajectories provided by TRACAP [20] furthermore showed to contain errors, such as unrealistic high velocities and accelerations, which were observed during the development of the algorithm (Figure 6 & 7). Two different filters were tested in this thesis; a modified Butterworth filter (4^{th} order, low-pass, 3 Hz cut-off) in relation to the filter used by Moura et al.[34], and a Kalman filter which was stated to be an appropriate filter for ball trajectories by Zheng & Zhou, [21]. The Kalman was presumed advantageous as it models both the measurement noise which can be affected by the location of the ball from the camera and the dynamics of a trajectory to provide an estimate of the ball trajectory [21](A script of the application of Kalman filter is presented in Appendix 3). By a visual assessment of the two filters, it was found that the Kalman filter delayed the peak-values from the accelerations and the velocities as opposed to the Butterworth filter. Since the Butterworth filter showed the best fit, it was thereby selected (Figure 6 & 7).

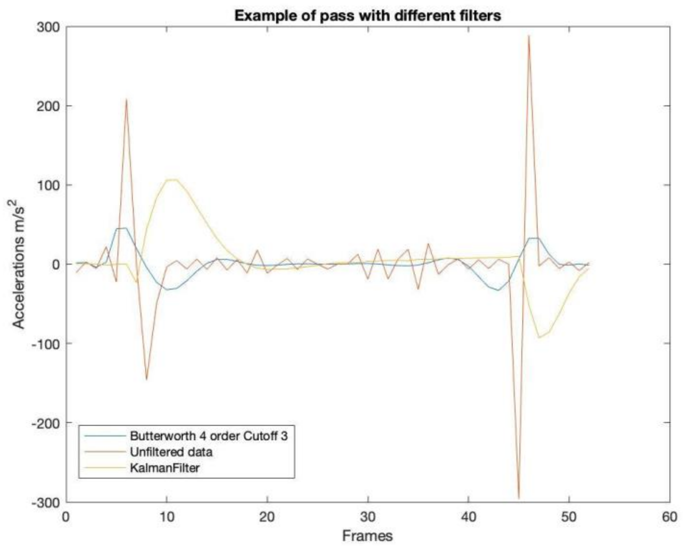


Figure 6: The acceleration of the ball after a pass computed from the raw trajectory of ball is shown on the y-axis (orange curve). The frame count is furthermore shown on x-axis . The yellow and blue curve furthermore shows the acceleration of the ball after the application the Kalman- and Butterworth filter (4th order, low-pass, 3 Hz cut-off).

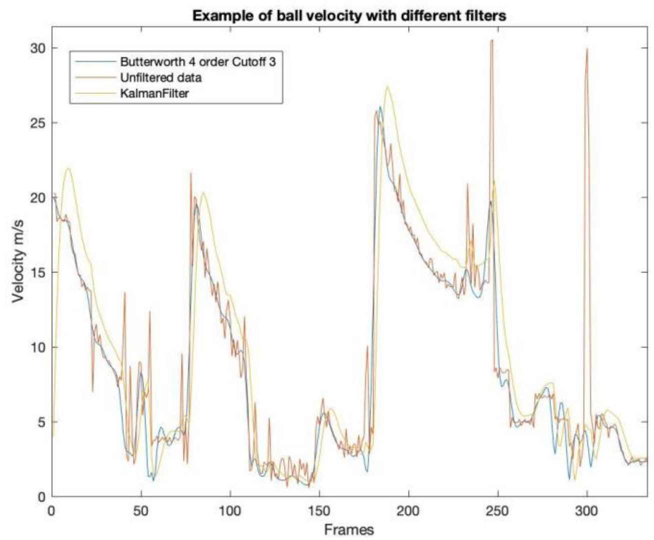


Figure 7: The velocity of a pass computed from the raw trajectory of ball is shown on the y-axis (orange curve). The frame count is furthermore shown on x-axis. The yellow and blue curve furthermore denotes the velocity of the pass after the application the Kalman- and Butterworth filter (4th order, low-pass, 3 Hz cut-off).

3.4.1.3 Calculating vector variables

The filtered trajectories of the ball and players was used to compute the following variables at each time stamp *t*: Displacements, *D*, velocities, *v,* and accelerations, *a,* by equation (1, 2 and 3), respectively.

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}, \;\; v = \frac{\Delta d}{\Delta t}\,, \;\; a = \frac{\Delta v}{t} \qquad (1, 2, 3)$$

where x_1 and y_1 is denoting the initial X and Y position of a given trajectory and x_2 and y_2 is denoting the next X and Y position at the given time stamp.

3.4.1.4 Calculating distances between vectors

The next step includes calculating the distance from the players, p_i, to the ball, b_i, in the X and Y positions for every timestamp t, the distance corresponds to the Euclidean distance, d, which are the straight-line distance between two points, and is calculated by equation (4):

$$d\ (p\ , b) = \sqrt{(p_1 - b_1)^2 + (p_2 - b_2)^2 + . . . + (p_n - b_n)^2} \qquad (4)$$

3.4.2 Identification of a pass

The process of identifying the occurrence of pass initially involves an identification of when the individual ball possession (IBP) starts and ends for a player [13]. (A script representing the steps in this section is shown in appendix 4)

3.4.2.1 Identifying the start of individual ball possession

The start of IBP was defined as when a player is physically able to interact with the ball, characterized as the *physical area*. To detect if the position of the ball is within the physical area, a threshold, T_p, of 1m radius was applied [Equation (5)], which was determined by a trial and error procedure conducted on the test data set.

$$physical\ area\ =\ d\ (p, b)\ <\ 1\ m\ (T_p) \qquad (5)$$

The Euclidean distance between the ball and player was calculated without taking the z-coordinate of the ball into account. However, if the trajectory of the ball passes above a player, it could result in ball possession being incorrectly assigned [13]. To solve this issue a threshold, T_b, for the Z-coordinate of the ball, was set to 2.5m [Equation (6)], to ignore when the trajectory of the crosses the T_b of a played with a height <2.5m. This threshold was determined by a trial and error procedure conducted on the test data set.

$$Exclude\ ball\ possession\ =\ ball\ \ z - coordinate\ >\ T_b \qquad (6)$$

3.4.2.2 Identifying the end of ball possession

When the ball is alive, it is assumed that only one player has IBP and can perform a pass. As the procedure of detecting the start of IBP has been defined, the next step involves defining how to identify the end of IBP. Four scenarios for the end of IBP for a player are is possible:

1) A player loses IBP – detected by the ball moving out of T_p.

2) The ball is dead (an interruption of the game) – detected by event data.

3.4.3 Detection of a pass

A pass in this thesis was defined as a pass performed by the lower extremities. The algorithm was furthermore set to only recognize passes when the ball is in an open-play scenario which is detected by the usage of the alive and dead flags in the event data. A script representing the steps in this section is shown in appendix 6 and 7.

The first step of identifying a pass, is to detect the beginning of a pass which is defined by the end of IBP for a player (T_p < *1m)* and IBP is not assigned to any other player, this means that the ball is not within the T_p threshold of any player.

The next step is then to detect if the pass is either success or unsuccessful. A successful pass is detected if a player on the same team then acquires IBP, as shown in Figure 8. An unsuccessful pass is detected if a player on the opponent team acquires IBP or the ball is the detected as dead.

Indexing a successful pass

Time	Home	Away	Ball
...			
20.1	3	0	0
20.2	0	0	1
20.3	0	0	1
20.4	0	0	1
20.5	4	0	0
...			

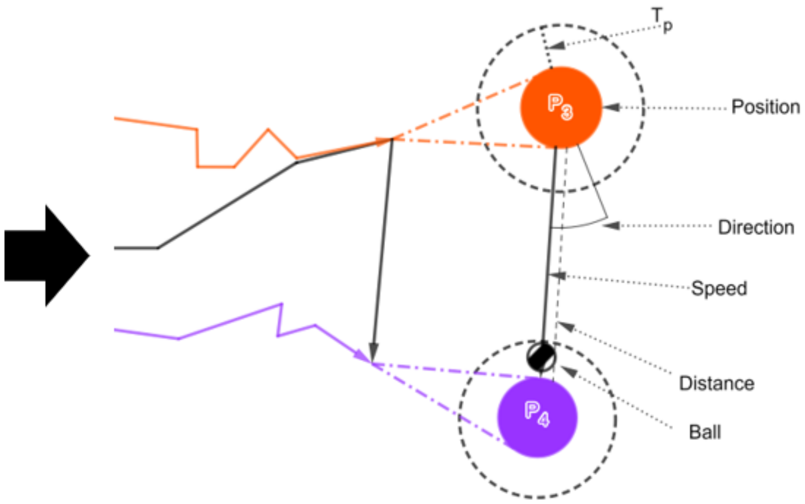


Figure 8: **Left:** Indexing of a successful pass: In the Home and Away column, the value denotes the player ID of the player with IBP at a given time stamp, a value of 0 denotes no assigned IBP. The Ball column denotes if any player on either of the teams has IBP, 0 denotes no IBP and 1 denotes IBP. **Right:** Three trajectories (black = ball, orange = player 3 (P3) from the home team, purple = player 4 (P4) from the home team) which illustrates a successful pass by the ball moving from the T_p threshold of P3 to the T_p threshold of P4.

3.4.3.1 Identifying complex passing scenarios

Certain problems occurred in the detection procedure of the passes due to complex situations. These situations and the solutions are described by the following examples:

1) When a player loses IBP and then the same player regains IBP after a certain amount of time, i.e. when a player performs a successful long dribble. The was solved by detecting if any other player acquires IBP in the meantime. This will not be detected as a pass.

2) When the ball trajectory crosses the T_p threshold of another player without the player interacting with the ball, thus resulting in an incorrect identified pass as illustrated in Figure 9. This problem was solved by applying an acceleration threshold, T_a, when the ball is moving within the T_p threshold of a player. This was based on the assumption of that a change in the direction of the ball only can occur due to an interaction by a player, which either increases or decreases the acceleration of the ball. The T_a threshold was set to ±10 m/s^2 by investigating the peak accelerations during IBP (Figure 10).

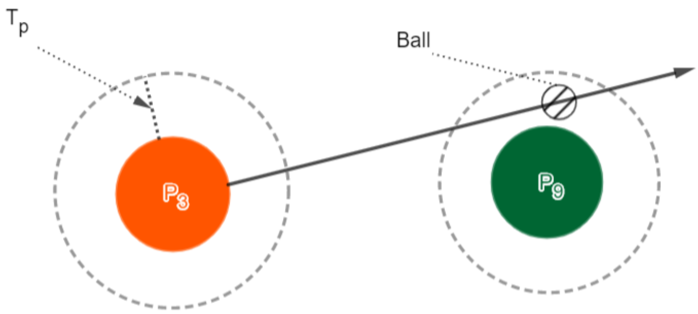


Figure 9: An illustration of an identified problem when the trajectory of the ball crosses the *Tp* threshold of player 9 (P_9) after a pass played from player 3 (P_3). The circular dotted line around each player denotes the *Tp* threshold.

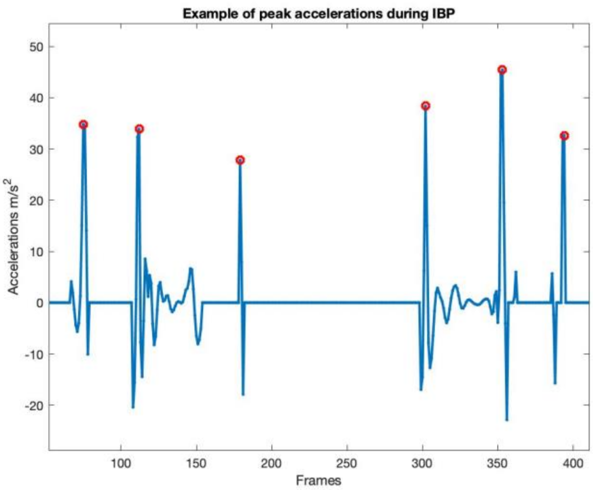


Figure 10: Identification of the acceleration of each peak when the ball is positioned with the T_p *threshold* denoted by the red circles

3) When the ball is positioned within the T_p threshold of two players and both thereby are identified as having IBP. It is therefore difficult to identify which player that is performing a pass (Figure 11). Problem 2 is obviated by assigning the pass to the player closest to the ball.

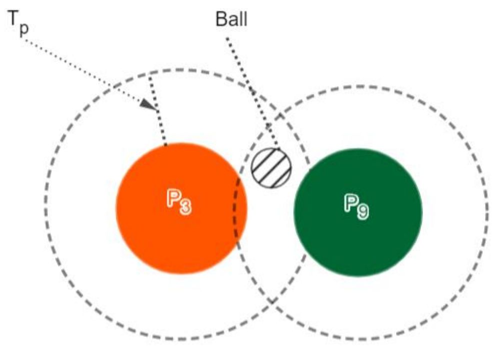


Figure 11: An illustration of an identified problem by the ball positioned within the *Tp* threshold of two players, P_3 and P_9. The circular dotted line around each player denotes the *Tp* threshold.

3.4.4 Calculating the length, direction and velocity of a pass

The start and end of a pass has thus been defined. It is now possible to calculate the length, direction and velocity of each pass.

3.4.4.1 Passing length

The length of a pass played from the X and Y position of the passing player (A) to the X and Y position of the receiver (B) was calculated by the usage of formula (1) (Figure 10). Long passes can potentially exceed the distance of 31m, therefore a boundary is set to 90m to exclude all passes above that value, furthermore passes with a duration <10 s are also excluded as this could indicate corrupted data [8].

3.4.4.2 Passing velocity

The velocity of a pass is calculated based on the duration and length of the pass by formula 2.

3.4.4.3 Passing direction

To detect if a pass is being played either towards the opponent's goal or a team's own goal, the playing direction for each team was accounted for. An input dialog box was applied to the algorithm to define the direction for each team. $x1 > x2 \mid x1 < x2$. Where the x-axis represents the longitudinal dimension of the field. The same rule was included for the y-axis to detect whether the ball was passed left or right. The direction was calculated by the displacement of the ball and the angle of the pass in relation to the X and Y coordinates of the passer and the receiver [8]. The angle is calculated by equation (7):

$$\alpha = tan^{-1}\frac{Xpass}{Ypass} \tag{7}$$

Where Xpass denotes the displacement of the ball on the x-axis and Ypass denotes the displacement of the ball on the y-axis (figure 10).

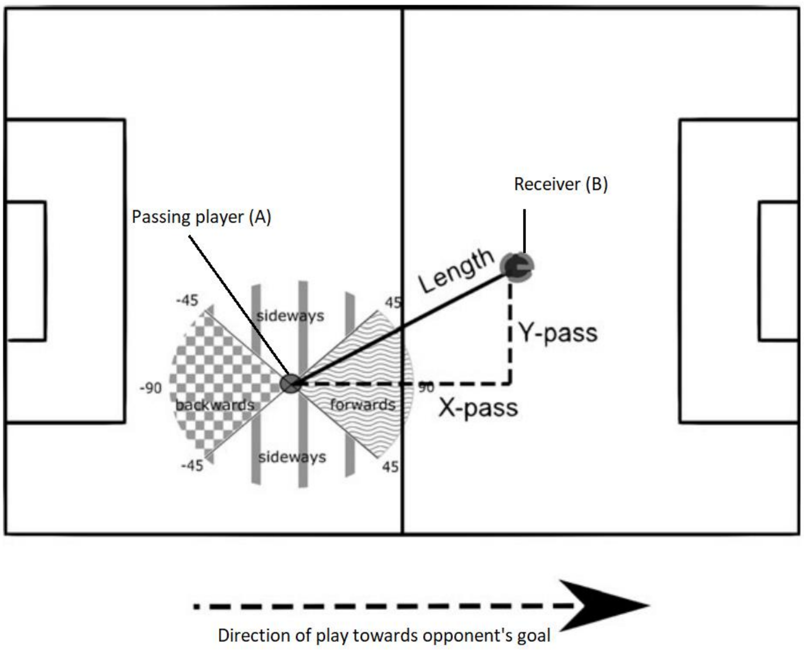


Figure 10 - Schematic illustration of a pass (position A) to the receiver (Position B). The direction of the pass is represented by the length and angle (°). X-pass and Y-pass denotes the displacement of ball on the X and Y axis, respectively (Goes et al.[8], modified).

3.4.5 Excluding headers as a pass

To exclude all actions which are not performed by the lower extremities, such as headers (Liu et at., 2015). When a player is performing a pass, the algorithm is set to check the Z-coordinate. A threshold, T_h, for the z-coordinate of the ball is set to *1.06 m* and all passes performed when the Z-positions of the ball exceeds this threshold is hereby excluded. This threshold was created as a basic estimate of the center of mass (COM) of the players. It was determined based on the mean height of the teams in the Danish Super 2018/2019 season (184.08 ± 1.76cm) [35] and the Clauser and Colleagues' Body Segment Parameters, *Total Body* proportions of the segments of 0.5881 [36]. When testing this threshold, it was found that the z-coordinate was unreliable when the ball is within the T_p threshold, therefore a time window of *-0.2s,* prior to the moment of the pass, is set to check if the height was above 1.06m.

3.4.6 Algorithm output

Based on the different on detection of each pass explained in the previous steps and the calculation of the length and direction following outputs were produced by the algorithm:

1. A detection of each occurred pass in relation to the given time in seconds
2. A labelled classification of each pass assigned to a player and team
3. The completion rate of the different kind of passes (Script is presented in appendix 8)
4. Characteristics of the labelled of passes
    - Length
    - Velocity
    - Acceleration

3.5 Notation protocol

This section describes the process of creating the ground truth by observation and manual notation of each pass by viewing a televised video recording of the match.

One author (internal observer) and one independent observer (external observer) created the notations of a pass by separately viewing the televised video recording of the given match, thus resulting in two sets of ground truth. The observers were additionally provided with a second-by-second playback control, all the time needed to perform the notation and a cross-scheme to note the occurrence of a pass at the given time in seconds.

3.5.1 Protocol procedure

With the purpose of creating a consistent notation process, the observers were provided with a written protocol which described the notation procedure (Appendix 1). The current thesis furthermore followed the procedure and principles explained by Lincoln and Guba, [37] of consistent observation:

> *Prolonged engagement:* To ensure a consistent notation process, various actions were performed to secure that the observers understood the culture and the phenomenon of football. The idea was to ensure that the observers were able to identify when a pass was performed. This included viewing video recordings of other matches.
>
> *Member checking:* The difference of the assigned notation by the observers were checked after a match was analyzed. The passes which were noted differently by the observers were selected. The video recordings of these passes were then viewed by the observers to discuss the rationale behind these notations. The intention of this was to secure consensus of the identification of a pass.

3.7 Evaluation methods

This section introduces the methods used to assess the inter-rater agreement of the two sets of ground truth and the accuracy of the algorithm.

3.7.1 Inter-rater agreement

The agreement between observers can be expected to be limited due to human errors [38] and to assess the agreement between the two sets of noted passes by the observers, the current study used Cohen's kappa, κ [39]. It accounts for the possibility of an agreement to occur by chance and is defined by the following equation (8):

$$\kappa = \frac{P_o - P_e}{1 - P_e} \tag{8}$$

where, P_o, is the segment of observed agreement between the observers, and P_e, is the segment of observations which are expected to agree by chance. The Cohen's kappa was interpreted based on the following guidelines as proposed by Landis & Koch, [40]:

- *K = 0* indicates the agreement equal to chance
- *K < 0.1 - 0.20* indicates a slight agreement
- *K < 0.21 - 0.40* indicates a fair agreement
- *K < 0.41 - 0.60* indicates a moderate agreement
- *K < 0.61 - 0.80* indicates a substantial agreement
- *K < 0.81 - 0.99* indicates a nearly perfect agreement
- *K* = 1 indicate a perfect agreement

Even though Cohen's kappa does not have statistical significance, it does however provide a measure for the similarity of the outcome of the observers [40].

3.7.2 Accuracy

To assess and test the accuracy of each detected pass in the algorithm, the metrics *F-score, Precision* and *Recall* (sensitivity) were used in relation to the ground truth set by the external observer. These metrics are a commonly applied evaluation tool in machine learning [41] and are computed by the three parameters; *true positives (TP), false positives (FP)*, *false negatives (FN),* which further are defined as:

TP - are evident if a pass detected by the algorithm also is noted in the ground truth.

FP - are evident if a pass detected by the algorithm not is noted in the ground truth

FN - are evident if a pass noted in the ground truth not is detected by the algorithm.

Precision is defined as the measure of quality of the algorithm by the number of correct positive results divided by the number of the predicted positive results [Equation (9)].

$$\frac{TP}{TP} + FP \tag{9}$$

Recall is defined as a measure of completeness by the number of correct positive results divided by the number of actual positive results [Equation (10)].

$$\frac{TP}{TP} + FN \tag{10}$$

F-score is defined as the weighted average of the recall and precision [Equation (11)].

$$2 * \frac{Recall * Precision}{Recall * Precision} + FN \tag{11}$$

The F-score is interpreted based on the following guidelines: *0 < F < 1* where *F = 1* indicates perfect accuracy [41].

4 Results

In this chapter, the results of the current thesis based on the experimental data set is presented. The results of inter-rater agreement are presented first, followed by the results of the accuracy of the algorithm and then succeeded by the results of the labelled passes and lastly the descriptive statistics of the characteristics of each labelled pass.

4.1 Results of inter-rater agreement

Table 2: Confusion matrix showing the inter-rater agreement between the internal and external observer by the true positive values, false negative values, false positive values and true negative values. Positive and negative, are denoted as Pos. and Neg., respectively

		INTERNAL OBSERVER		
		Pos.	Neg.	Total
EXTERNAL OBSERVER	Pos.	331	13	344
	Neg.	21	2579	2600
	Total	352	2592	2944

Table 2 shows the inter-rater agreement between the internal and external observer based on the noted and not noted passes each time stamp (n=2944). The total number of true positives = 331, false negatives = 21, false positives = 13 and true negatives = 2579. The total of number of disagreements between the observers was 34, which yielded a Cohen's kappa value of 0.94

4.2 Algorithm accuracy

Table 3: Confusion matrix showing the true positive values, false positive values and false negative values between the ground truth and the algorithm. Positive and negative, are denoted as Pos. and Neg., respectively

		GROUND TRUTH		
		Pos.	Neg	Total
ALGORITHM	Pos.	266	74	340
	Neg.	62	2644	2706
	Total	328	2718	

Table 3 shows the TP, FP and FN values between the ground truth and algorithm which the calculation of the precision, recall and F-score were based on. The total number of true positives = 266, false negatives = 62, false positives = 74. The total of number disagreements between the observer and the algorithm was 136.

Table 4: Accuracy of the algorithm by the metrics of precision, recall and F-score.

	PRECISION	RECALL	**F-SCORE**
Algorithm - External	0.81	0.78	**0.80**

Table 4 shows the accuracy of algorithm by metrics Precision, Recall and F-score, which was 0.81, 0.78 and 0.80, respectively.

4.3 Results of labelled passes

Table 5: The individual and total number of labelled passed for each player on the home team. Each player is denoted by the playing position in the team formation: Goalkeeper (GK), central back (CB), right wing (RW), left wing (LW), central midfielder (CM) and central forward (CF). Backwards direction, sideways direction, forwards direction is respectively denoted as BW, SW and FW. An example of the completion rate for a CM is presented outlined in the table.

	SP			MP			LP			TOTAL PASSES			
	BW	SW	FW	BW	SW	FW	BW	SW	FW	Total BW	Total SW	Total FW	Total Passes
GK	0	0	1	0	1	7	0	4	7	0	5	10	15
CB	0	3	1	1	1	2	0	0	2	1	4	4	9
CB	0	1	0	0	2	0	0	0	0	0	3	1	4
CB	1	2	2	2	3	3	0	0	3	3	5	6	14
RW	0	6	2	3	5	1	0	3	1	3	14	5	22
LW	0	5	2	3	4	2	1	0	2	4	9	9	22
CM	2	2	0	0	3	0	0	0	0	2	5	1	8
CM	1	5	1	2	6	2	0	2	2	3	13	4	20
SUC (%)	100	60	100	0	100	100	0	100	50	100	50	50	80
CM	1	7	2	0	2	2	1	1	2	2	10	6	18
CF	3	5	2	0	5	0	0	1	0	3	11	2	16
CF	1	1	2	0	3	0	0	0	0	1	4	3	8

Table 6: The individual and total number of each labelled passes for performed by each player on the away team. Each player is denoted by the playing position in the team formation: Goalkeeper (GK), central back (CB), right midfielder (RM), left midfielder (LM), central midfielder (CM) and central forward (CF). Backwards direction, sideways direction, forwards direction, short pass, medium pass, long pass is respectively denoted as BW, SW, FW, SP, MP and LP.

	SP			MP			LP			TOTAL PASSES			
	BW	SW	FW	BW	SW	FW	BW	SW	FW	Total BW	Total SW	Total FW	Total Passes
GK	0	0	2	0	7	0	0	2	1	0	9	3	12
CB	1	0	1	2	11	5	0	1	5	3	12	11	26
CB	2	0	3	0	21	1	0	3	2	2	24	6	32
CB	0	5	1	4	11	5	1	0	5	5	16	11	32
RM	2	0	2	0	0	3	1	0	0	3	0	5	8
LM	1	2	1	0	4	2	0	0	0	1	6	3	10
CM	1	1	2	2	6	1	0	1	0	3	8	3	14
CM	1	3	1	2	2	2	0	0	3	3	5	6	14
CF	2	6	0	3	3	0	0	2	0	5	11	0	16
CF	1	1	0	0	2	1	0	0	0	1	3	1	5
CF	1	3	3	2	0	2	0	1	0	3	4	5	12

Table 5 and 6 shows number of the individual and total number of labelled passes performed by each player on the home and away team respectively. A central midfielder is furthermore outlined in table 5 which shows the passing completion rate of the player related to each and the total number of labelled passes. The outlined player performed a total of 20 passes with a completion rate of 80%.

Table 7: Total number and completion rate of each labelled pass for the home team, respectively. Backwards direction, sideways direction, forwards direction, short pass, medium pass, long pass and pass completion rate are respectively denoted as BW, SW, FW, SP, MP, LP and SUC.

	SP			MP			LP			TOTAL PASSES			
	BW	SW	FW	BW	SW	FW	BW	SW	FW	Total BW	Total SW	Total FW	Total Passes
TOTAL	9	37	15	11	35	19	2	11	19	22	83	51	156
SUC. (%)	66	67	40	100	88	36	100	90	36	86	79	45	69

Table 8: Total number and completion rate of each labelled passes for the away team. Backwards direction, sideways direction, forwards direction, short pass, long pass, and pass completion rate are respectively denoted as BW, SW, FW, SP, MP, LP and SUC.

	SP			MP			LP			TOTAL PASSES			
POS.	BW	SW	FW	BW	SW	FW	BW	SW	FW	Total BW	Total SW	Total FW	Total Passes
TOTAL	12	21	16	15	67	22	2	10	16	29	98	54	181
SUC. (%)	50	52	56	100	86	36	100	70	38	83	70	49	67

Table 7 and 8 shows the total number of each labelled passes and the related completion rate for home and away team respectively. The home team performed a total of 22 backwards passes with a completion rate of 86%, 83 sideways passes with a completion rate of 79% and 54 forwards passes with a completion rate of 49%. The away team performed a total of 29 backwards passes with a completion rate of 83%, 98 medium passes with a completion rate of 70% and 54 long passes with a completion rate of 49%. The home and away team performed a total of 156 and 181 with a completion of 69% and 67%, respectively.

4.4 Descriptive statistics of the characteristics of each labelled pass

Table 9: Descriptive statistics (mean ± standard deviation) of the length, velocity and acceleration of each labelled pass for the home team.

		LENGTH		VELOCITY		ACCELERATION	
SHORT	Forwards	5.9	±3.1	17.5	±10.9	56.2	±54.1
	Backwards	5.2	±2.8	7.5	±3.2	48.0	±55.9
	Sideways	6.4	±3.0	11.1	±9.0	59.8	±57.8
MEDIUM	Forwards	16.0	±3.7	12.3	±5.5	72.0	±67.4
	Backwards	19.7	±4.7	11.3	±3.4	52.4	±47.0
	Sideways	19.5	±6.1	13.3	±4.4	53.0	±48.5
LONG	Forwards	49.1	±12.1	12.6	±2.6	84.9	±46.2
	Backwards	36.6	±2.2	13.0	±3.1	37.8	±13.6
	Sideways	40.6	±9.9	13.7	±3.4	69.3	±31.3

Table 10: Descriptive statistics (mean ± standard deviation) of the length, velocity and acceleration of each labelled pass for the away team.

		LENGTH		VELOCITY		ACCELERATION	
SHORT	Forwards	5.4	±2.7	8.5	±5.7	56.6	±80.2
	Backwards	7.3	±3.1	11.7	±5.3	53.0	±52.5
	Sideways	6.2	±3.3	14.2	±9.5	42.0	±38.8
MEDIUM	Forwards	20.4	±5.2	12.6	±4.5	46.2	±35.3
	Backwards	19.7	±5.9	10.6	±2.3	75.1	±68.8
	Sideways	20.1	±5.7	12.3	±4.1	54.8	±44.3
LONG	Forwards	52.8	±10.2	12.1	±2.6	64.6	±49.5
	Backwards	33.7	±2.2	10.2	±0.3	120.3	±103.1
	Sideways	36.2	±5.5	13.2	±4.4	74.3	±44.8

Table 9 and 10 shows the mean and standard deviation of the length, velocity and acceleration in the passing moment of each labelled pass for the home and away team, respectively.

5 Discussion

Ground truth

The assessment of the notations created by the two observers showed a Cohen's kappa value of *0.94,* based on inter-rater agreement shown in Table 2, which indicates a nearly perfect agreement and furthermore, that it is possible to objectify the occurrence of a pass by observation. Even though Cohen's kappa does not provide a measure for the statistical significance, as previously stated, it indicates that the ground truth created by the internal and external observer provided a nearly perfect measure to evaluate the algorithm.

One of the ground truth sets was created by an internal observer, and this included a risk of the observer being biased. Since the inter-rater agreement was based on notations by an external observer, it can be assumed that the observation created by the internal observer did not influence the agreement.

As previously stated in section 3.5, the notation protocol was conducted by viewing a televised video recording which included replays of certain events. During the replays, it is possible that the game continues, which is not visible for the observer, thus resulting in a missed notation of a pass. As the algorithm does not account for the replays, a difference between the detected passes by the algorithm and the noted passes by the observers is possible. This problem was solved by manually notating the start and end of each replay and hereafter exclude these time intervals in the experimental data set. In the future this problem could be avoided by a video recording without replays.

Algorithm

Table 4 shows a precision of 0.81, recall of 0.78 and F-score of 0.80 on average, which indicates a substantial accuracy in relation to the algorithms ability to detect passes. The substantial accuracy was calculated from the TP, FP and FN variables, which shows a number of 266, 62 and 74 in table 3. The total number of disagreements between the algorithm and the observer were found to be 136. There are multiple constraining aspects in relation to the methodological setup in the current thesis which could have influenced the experimental results and caused this number of disagreements. The following will include a discussion of these constraints, the possible effects, and if there are any potential and future solutions.

The data available were limited to one half of a match and the accuracy of the algorithm can because of this not be generalized to other matches. In the future, an evaluation of the algorithm should be conducted on a larger number of matches played on different stadia and by different teams. This could encourage the identification of complex interactions and thereby optimize the pass detection procedure. It could additionally reveal if the classified passes performed by each player or team varied because of factors like the venue, playing either home or away or by the level of the opponent team, as found by Lago et al.[27], and thereby will provide additional information for coaches.

As stated in section 3.4.1.2, it was observed that the raw trajectories of the ball provided by TRACAP [20] were not perfect, especially regarding the z-positions and the accelerations of the ball, which were also found by Irnai et al.[9], and Link et al.[13]. These errors can be considered as a limitation of the data as the displacement, acceleration and velocity variables were computed from this. By an investigation of a pass where the ball travels through the air where it is presumed to be solely affected by gravity and air resistance [21], high accelerations were still found. This could indicate either errors in the data and insufficient filtering. Future research should test the application of other dynamic filters. Furthermore, as the exclusion of head passes were based on this threshold related to the Z-coordinate, it is possible that some head passes were not excluded. Furthermore, as the time-window was set to 0.2 sec, prior to the execution of the pass a problem can occur when detecting a pass subsequently to a high pass. If the raw ball trajectories were without errors, it can be assumed that a higher accuracy could be obtained.

Furthermore, the problem of quantifying a dynamic invasion sport, such as football where complex interactions between the players occur on a regular basis, needs to be considered [10], [11]. As described in section 3.4.3.1, a part of the identified passing and possession situations proved to be complex and difficult to detect. Some of these situations, such as a successful dribble or when the trajectory of the ball crossed the *Tp threshold* of a player were solved, but others are still evident. Some of them, because they have similar characteristics to a pass and thus are difficult to separate from. Player actions, such as a clearance, a shot on goal, a save by a goalkeeper and a long dribble intercepted by an opponent player were identified as some of these actions which, in some situations, were found to fulfill the criteria of a pass. An overestimation of the number of passes due to incorrect identification is therefore a risk that needs to be considered in relation to the precision metric. Evidently, another situation which proved difficult to quantify is when a receiver of a pass does not impact the direction of the ball. The criterion which requires a change in the acceleration of the ball within the T_p threshold is thereby not fulfilled, thus potentially effecting the recall metric as some passes are not detected. A solution to an exclusion of these actions, like a shot on goal, could be an identification of the characteristics of the actions by utilization of event data or a manually notation of a significant number of shots on goal and then calculate the mean velocity to determine an exclusion threshold for shots. By an identification of these characteristics it would furthermore be possible detect shots and create an additional type of event in the algorithm.

An additional complicated problem is identified, when multiple players are positioned within the same *Tp threshold,* as they both can interact with the ball as described in section 3.4.3.1. In the current thesis, no solution to this problem was found and it is considered that that this problem confirms the limitation of analyzing football solely based on trajectory data and the importance of recognizing these complex actions in future research.

As no machine learning techniques were applied, as in the study by Irnai et al.[9], to train and optimize the criteria thresholds which were required to be fulfilled in order to detect a pass, the algorithm presented in this thesis can be considered as being somewhat rigid. The future steps for this algorithm are thereby to apply machine learning.

The current thesis aimed to detect passes solely based on ball and player trajectories. However, the algorithm was dependent on event data to detect the interruptions in the match. Future research should aim to detect these interruptions in order to produce a fully automated recognizing model, however certain events, i.e., “yellow card” is impossible to detect according to Gudmondsson & Wolle, [7], which states that additional information is needed to detect similar events.

When comparing the accuracy of the model presented in this thesis to other studies detecting similar events, such as Irnai et al.[9], who showed a f-score of 0.86 for passes and Khan et al.[42], . who showed an accuracy of 0.84 when detecting ball kicks and 0.92 in ball possession it can be stated that the algorithm presented is nearly as accurate. However, in this comparison it needs to be considered that Irnai et al.[9], used a machine learning model which, as previously described, potentially can optimize a recognition system, and that Khan et al.[42], detected events which not are as complex as a pass.

Pass classifications

When analyzing the results, the accuracy of the algorithm and its constraints discussed in the previous section needs to be accounted for. The general idea of the algorithm was to develop an automated algorithm, to avoid the requirement of manual notation and to detect and label each pass to a classified pass to provide the initial steps of quantifying a football pass based on spatio-temporal data. The classified passes are limited in the information achieved regarding how they can impact a game, however they do provide coaches with a simple tool to evaluate the individual and team performance in relation to passes. Even though the data included in this thesis was limited to one half of a match and a statistical analysis of the tendencies was therefore futile. Despite the limited data, following section will include several examples of how a coach could benefit from this data and how the results can benefit football research.

Table 5, 6, 7 and 8 shows the number of different labelled passes performed, as well as the completion rate, which previously was stated as important predictors for success [14], [19], [28]. In table 7 it can be observed that the home team performed a total of 11 medium backwards pass with a completion rate of 100%, a total of 35 medium sideways passes with a completion rate of 88% and a total of 19 long forward passes with a completion rate of 36%. This can give coaches knowledge about the difficulty level of certain types of passes and furthermore that the performance of a long forward pass involves a greater risk as they produced a lower completion rate which is in accordance with Steiner, [29]. Additionally, an individual completion rate, as shown in table 8, of different types of passes can yield information of the individual performance in a match, which is crucial for team practicing a possession-based style of play [25]. In table 6 it can be observed that the away team performed a total of 25 passes more than the home team whereas 15 of these is labelled as sideways. This could indicate a specific tactic of space creation and disrupting the opponent's organization, which was found to be an impactful approach in football [31]. This indication furthermore exemplifies how a coach could analyze the tactics of an opponent team based on the number, direction and length of a pass [19]. It can be observed in Table 5 and 6 that the only three players, positioned as central defenders for the away team, performed more than 25 passes, which from a coach's perspective could indicated that the opponents favors a tactic based on the passes made by the defenders.

Table 9 and 10 contains descriptive statistics of the mean and standard deviation of the length, velocity and acceleration in the initial moment of the labelled passes for the home and away team, respectively. The initial application of these results can provide future research aimed at classifying passes with estimates of the characteristics of each pass, as no previous studies have investigated this. The velocity and acceleration were included to provide an additional measure to quantify the characteristics of each pass, as both Goes et al.[8], and Power et al.[31], found the velocity of a pass to be a determinant of success.

It can be observed that the short, medium and long passes showed the highest mean values of 6.4 and 6.2, 19.7 and 20.4 and 49.1 and 52.8, for the home and away team respectively. It can be considered that these results were expected due to the definitions of the passes, which additionally is confirmed as the short passes, as an example, showed lower standard deviation values compared to the medium and long passes. It can however provide some indications of the mean length related to the direction of a pass. E.g. the long forward pass which showed a value of 49.1 and 52.8 for the home and away team, respectively. The classification of each pass regarding the included characteristics, such as length and velocity, of each pass could thus be optimized. Another procedure to optimize these characteristics could be to perform a cluster analysis on a larger data set. Future development of the algorithm will aim to implement this.

Steiner stated, that to perform a successful long forward pass it requires a certain amount of velocity. However, the values of the velocity regarding the long forward pass does not indicate this. The labelled passes furthermore show rather similar values for all labelled passes in exception of the short sideways for the away team and short forward for the home team which both show a standard deviation around of 9.5 and 10.9, respectively.
The standard deviation of the accelerations shows very high values. This indicates the problem regarding the applied filter and errors in the data discussed in the previous section.

6 Future perspectives

In the discussion, different suggestions are proposed of how to improve the algorithm and add additional parameters. This chapter includes how the algorithm can be implemented on tracking data collected in other domains and settings

The algorithm presented in the present thesis were only tested in a match setting, however the information gained from this assessment tool can also be transferred to training sessions to improve tactical and technical performance and development. It is evident in the literature that analysis of tracking data in football is mainly conducted on football matches. The evaluation of performance in training is especially related to the physical parameters to determine the training load for players, especially, in relation to injury prevention [43].
The importance of applying tools to evaluate technical performance is in accordance with the former technical director of Shanghai International Port Group Football Club and founder of Optima Football Consulting Mads Davidsen. He stated that training account for 90% of time spent on the pitch, whereas matches only accounts for 10% [44]. Even though the algorithm is dependent on event data to detect the interruptions in a match, it is still possible to implement on data collected in a training session if tracking systems are available in the training arenas. As the algorithm can detect if the ball moves out of the playing field, if the field dimensions is defined by the coaches.

Implementing the algorithm on tracking data can potentially provide endless possibilities to evaluate performance and recognize patterns in the game of football. By small modifications of the algorithm is it furthermore possible to apply in other invasion games such as handball and basketball. The method of classification and identification of passes can be furthermore be used as a valuable supplement to artificial intelligence (AI) and machine learning methods. By identification and classification of the different passes the different event can be used to train neutral networks to perform a similar identification by a much more automated procedure. The authors believe that the classification and identification methods can be a valuable supplement to AI.

7 Conclusion

In this thesis, an automated algorithm based on spatio-temporal data of the ball and players and event data from a real football match has been developed to detect passes and classify different passes. The algorithm showed to provide substantial accuracy based on the evaluation of the Precision, Recall and F-score metrics, which indicates that the algorithm can be implemented as a useable objective assessment tool for coaches despite limitations regarding the data and constraints of the algorithm. The current thesis furthermore stated the importance of identifying certain complex situations when quantifying passes and that future research should implement machine learning techniques and larger data sets to optimize the algorithm.

References

[1] H. Liu, W. Hopkins, M. A. Gómez, and J. S. Molinuevo, “Inter-operator reliability of live football match statistics from OPTA Sportsdata Inter-operator reliability of live football match statistics from OPTA Sportsdata,” *Int. J. Perform. Anal. Sport*, vol. 13, pp. 803–821, 2013.

[2] TotalSportek, “25 World’s Most Popular Sports (Ranked by 13 factors),” 2017. [Online]. Available: 25 World’s Most Popular Sports (Ranked by 13 factors). [Accessed: 01-May-2019].

[3] Deloitte, “Rising stars,” *Manchester, United Kingdom: Deloitte Sports Business Group.*, 2018. .

[4] UEFA.com, “How clubs’ 2018/19 UEFA Champions League revenue will be shared.” [Online]. Available: https://www.uefa.com/uefachampionsleague/news/newsid=2562033.html. [Accessed: 01-Jun-2019].

[5] M. J. Horton, *Algorithms for the Analysis of Spatio-Temporal Data From Team Sports*, no. January. 2018.

[6] S. Chawla, J. Gudmundsson, J. Estephan, and M. Horton, “Classification of Passes in Football Matches using Spatiotemporal Data,” *ACM Trans. Spat. Algorithms Syst*, vol. 3, no. 2, p. 30, 2017.

[7] J. Gudmundsson and T. Wolle, “Towards automated football analysis: Algorithms and data structures,” *Proc. 10th Australas. Conf. Math. Comput. Sport*, 2010.

[8] F. R. Goes, M. Kempe, L. A. Meerhoff, and K. A. P. M. Lemmink, “Not Every Pass Can Be an Assist: A Data-Driven Model to Measure Pass Effectiveness in Professional Soccer Matches,” *Big Data*, vol. 7, no. 1, 2018.

[9] T. Irnai, A. Uchiyama, T. Magome, and T. Higashino, “Play recognition using spatio-temporal relationship of football tracking data,” *2017 10th Int. Conf. Mob. Comput. Ubiquitous Network, ICMU 2017*, vol. 2018-Janua, no. October, pp. 1–2, 2018.

[10] P. Lucey, D. Oliver, P. Carr, J. Roth, and I. Matthews, “Assessing Team Strategy using Spatiotemporal Data,” in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2013.

[11] R. Rein, D. Raabe, and D. Memmert, “‘Which pass is better?’ Novel approaches to assess passing effectiveness in elite soccer,” *Hum. Mov. Sci.*, vol. 55, no. July, pp. 172–181, 2017.

[12] H. D. Johansen, S. A. Pettersen, P. Halvorsen, and D. Johansen, “Combining video and player telemetry for evidence-based decisions in soccer,” *icSPORTS 2013 - Proc. Int. Congr. Sport. Sci. Res. Technol. Support*, no. September, pp. 197–205, 2013.

[13] D. Link and M. Hoernig, “Individual ball possession in soccer,” *PLoS One*, vol. 12, no. 7, pp. 1–16, 2017.

[14] J. Oberstone, “Differentiating the Top English Premier League Football Clubs from the Rest of the Pack: Identifying the Keys to Success,” *J. Quant. Anal. Sport.*, vol. 5, no. 3, 2009.

[15] P. Cintia, S. Rinzivillo, and L. Pappalardo, “Football_Network_Teams,” in *Machine Learning and Data Mining for Sports Analytics workshop (MLSA’15)*, 2015, pp. 1–9.

[16] P. Power, H. Ruiz, X. Wei, and P. Lucey, “Not All Passes Are Created Equal,” pp. 1605–1613, 2017.

[17] D. Link, S. Lang, and P. Seidenschwarz, “Real time quantification of dangerousity in football using spatiotemporal tracking data,” *PLoS One*, vol. 11, no. 12, pp. 1–17, 2016.

[18] T. Taki, J. Hasegawa, and T. Fukumura, “Development of motion analysis system for quantitative evaluation of teamwork in soccer games,” pp. 815–818, 2002.

[19] M. Kempe, M. Vogelbein, D. Memmert, and S. Nopp, “IJSS_possesion_vs_direct_play-1,” *Int. J. Comput. Sci. Sport*, vol. 4, no. 6A, pp. 35–

41, 2014.
[20] ChyronHego Coroporation, “TRACAP Optical Tracking,” 2019. [Online]. Available: https://chyronhego.com/products/sports-tracking/tracab-optical-tracking/.
[21] Y. Zheng and X. Zhou, *Computation With Spatial Trajectories*. Springer New York Dordrecht Heidelberg London, 2011.
[22] C. Reep and B. Benjamin, “Skill and chance in association football,” *J. R. Stat. Soc.*, vol. A, no. 131, pp. 581–585, 1968.
[23] Ł. Szczepański and I. Mchale, “Beyond completion rate: Evaluating the passing ability of footballers,” *J. R. Stat. Soc. Ser. A Stat. Soc.*, vol. 179, no. 2, pp. 513–533, 2016.
[24] M. Hughes and I. Franks, “Analysis of passing sequences, shots and goals in soccer,” *J. Sports Sci.*, vol. 23, no. 5, pp. 509–514, 2005.
[25] P. D. Jones, N. James, and S. D. Mellalieu, “Possession as a performance indicator in soccer.,” *Int. J. Perform. Anal. Sport*, vol. 4, no. 1, pp. 98–102, 2004.
[26] R. Rein and D. Memmert, “Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science,” vol. 5, p. 1410, 2016.
[27] C. Lago and R. Martín, “Determinants of possession of the ball in soccer,” *J. Sports Sci.*, vol. 25, no. 9, pp. 969–974, 2007.
[28] C. Collet and C. Collet, “The possession game ? A comparative analysis of ball retention and team success in European and international football , 2007 – 2010,” *J. Sports Sci.*, vol. 0414, no. May, pp. 37–41, 2012.
[29] S. Steiner, “Passing decisions in football: Introducing an empirical approach to estimating the effects of perceptual information and associative knowledge,” *Front. Psychol.*, vol. 9, no. MAR, pp. 1–11, 2018.
[30] A. Tenga, I. Holme, L. T. Ronglan, and R. Bahr, “Effect of playing tactics on achieving score-box possessions in a random series of team possessions from Norwegian professional soccer matches,” *J. Sports Sci.*, vol. 28, no. 3, pp. 245–255, 2010.
[31] J. Fernandez and L. Bornn, “Wide Open Spaces : A statistical technique for measuring space creation in professional soccer,” in *MIT Sloan Sports Analytics Conference*, 2018, pp. 1–19.
[32] P. Rudbæk, *Dansk Boldspil-Union Aldersreleteret træning 2*, no. november 2012. Dansk Boldspils-Union, 2013.
[33] D. Rosch *et al.*, “Assessment and Evaluation of Football Performance,” *Am. J. Sports Med.*, vol. 28, no. 5, pp. 29–39, 2015.
[34] F. Arruda Moura, L. Eduardo Barreto Martins, R. De Oliveira Anido, R. Machado Leite De Barros, and S. Augusto Cunha, “Quantitative analysis of Brazilian football players' organisation on the pitch,” *Sport. Biomech.*, vol. 11, no. 1, pp. 85–96, 2012.
[35] SuperStats, “SuperStats,” 2018. [Online]. Available: https://www.superstats.dk/hold/hojde. [Accessed: 01-Jun-2019].
[36] D. G. Robertson, E. G. Caldwill, J. Hamill, G. Kamen, and N. Saunders, *Research Methods in Biomechanics*, 2nd ed. Human kinetics, 2014.
[37] S. Y. Lincoln and G. E. Guba, *Naturalistic Inquiry*, vol. 75. London, UK: SAGE Publications.
[38] T. O'donoghue, *Planning Your Qualitative Research Thesis and Project - An Introduction to Interpretivist Research in Education and the Social Sciences*, 2nd ed. London, UK: Routledge, 2018.
[39] J. Cohen, “Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit,” *Psycho- Log. Bull.*, vol. 70, no. 4, pp. 213–220, 1968.
[40] J. R. Landis and G. G. Koch, “The Measurement of Observer Agreement for Categorical Data Data for Categorical of Observer Agreement The Measurement,” *Society*, vol. 33, no. 1, pp. 159–174, 2013.
[41] R. Baeze-Yates and B. A. Riberio-Neto, *Modern Information Retrieval the Concepts and Technology Behind Search*. Harlow, England: Pearson Education Ltd., 2011.
[42] A. Khan, B. Lazzerini, G. Calabrese, and L. Serafini, “Soccer Event Detection,” in *4th International Conference on Image Processing and Pattern Recognition (IPPR 2018)*, 2018, no. April, pp. 119–129.

[43] J. Christian and C. Jensen, “Event detection in soccer using spatio-temporal data,” 2015.
[44] KMD, “Spillerne efterspørger selv mere relevant data,” *SPORT & DATA*. [Online]. Available: https://www.kmd.dk/indsigter/mads-davidsen-spillerne-efterspoerger-selv-mere-relevant-data. [Accessed: 01-Jun-2019].