Football Analysis: An Automated System for Team Generation

Janvi Desai
Dwarakdas J. Sanghvi College of Engineering, Mumbai, India

Tanishq Tapiawala
Dwarakdas J. Sanghvi College of Engineering, Mumbai, India

Neha Ram
Dwarakdas J. Sanghvi College of Engineering, Mumbai, India
Corresponding author: Tanishq Tapiawala, Email: tanishq.tapiawala@gmail.com

Football Analytics has seen major growth in the past couple of years. Many of the top football clubs have started implementing various aspects of the vast field of football analytics. The utilization of data has become particularly important with respect to player development. In such a fast-paced game, keeping track of all the players at once is an extremely difficult task. In the paper, we propose the idea of an automated system that will help the manager to pick out the team strategy, with the playing line-up, for a particular match depending on the opponent's team tactics. The system will include 2 phases, first phase will predict the player's playing position using cluster analysis. In the next phase, the system will use deep learning to evaluate the performance of players based on their position. Using the output from the two phases, we will predict the team strategies.

Keywords: Match analysis, Playing position, Player performance, Team strategy, Cluster analysis, Deep learning.

2021. In Rahul Srivastava & Aditya Kumar Singh Pundir (eds.), *New Frontiers in Communication and Intelligent Systems*, 121–127. Computing & Intelligent Systems, SCRS, India. https://doi.org/10.52458/978-81-95502-00-4-15

1 Introduction

The use of information in football has gotten particularly critical to foster player abilities or match examination. The universe of team sports on a very basic level includes quick and complex plays that are distressing for viewers to follow and the manager to judge the team. In a speedy group game, monitoring every one of the players and dissecting their exhibition after each match is especially difficult. Current situations for distinguishing the presentation are scouring through long stretches of physically stamped recordings. This is a costly and careful interaction. The numbers can perceive what the mentors or players pass up a major opportunity, along these lines can be utilized to sort out how a group plays, the strategies they use in a game.

Thus, there is a need for an application that will help the managers to automate the whole process of scouring through multiple hours of match videos to determine a player's performance. The player performance should also depend on the position in which they play. This motivated us to create such a system that will automate the whole process of evaluating the player performance by suggesting the playing 11, depending on the strategies utilized by the opponent in the previous games.

2 Literature Survey

There are a couple of existing frameworks that assist an administrator with sorting out a strategy that the group will use during a match. For this the administrator should scour through the previous match recordings physically. This cycle is extremely thorough and tedious. The different helpful papers that we found during the writing audit of the current frameworks include:

(i) Team Tactics Estimation

This paper proposes a novel strategy for assessing group strategies in soccer video dependent on a Deep Extreme Learning Machine (DELM) and interesting attributes of strategies [5].The proposed technique appraises the strategies of each group from players' developments. The tactics involved are divided as:

1. Retreat (Defense)- Withdrawing from the opposing theory
2. Fore check (Defense) - Pressuring the ball
3. Set piece (Both) - Corner kick or Free kick
4. Possession (Attack) - Retaining control of the ball over longer periods of time
5. Swift Attack (Attack) - Kicking along through ball

(ii) Tactical Performance of Football Players

This investigation is intended to evaluate the specialized match execution of major parts in a drawn-out point of view [1]. It profiles the specialized exhibition of a player relying upon their playing position (focal safeguard, full back, focal midfielder, wide midfielder, forward). It assesses the genuine impacts of positional and situational factors on the player execution.

(iii) Tactical Performance Statistics for Individual Soccer Players

This paper proposes an automated framework that can distinguish, track, arrange the groups of various players and even recognize the player controlling the ball [2]. They use CNNs [10] to create three especially significant measurement of a player:

1. Span of ball possession
2. Number of successful passes
3. Number of successful takes

(iv) Predicting Player Position for Talent Identification

This paper is set to introduce another framework as indicated by the perspective for finding gifts in the sport of football enthusiastic to the players' particular brand name, physical, mental, and specific [3]. The blend of characteristics as overviewed by tutors are then used to expect the players' circumstance in a match that suits the player the best in a particular party approach. Evaluation of the proposed structure is two-wrinkle; quantitatively through request examinations to expect player position, and truly through a Talent Identification Site made to achieve a generally indistinct target. Results from the portrayal tests using Bayesian Networks, Decision Trees and K-Nearest Neighbour.

Table 1. Discussion of Recent Papers

Previous Papers	Model Used	Accuracy	Research Gap
Team Tactics Estimation	DELM	87.80%	This approach does not include the team tactics for the whole match. It only predicts what the team is doing during the moment of the play.
Predicting Player Position for Talent Identification	Bayesian Network	99%	The approach used doesn't cover every aspect. Only predicting the position of the player is included.
Tactical Performance Statistics	CNN	90.97%	The approach evaluates every player in a similar manner.
Tactical Performance	IBM SPSS 22	--	The approach used doesn't cover every aspect.

3 Proposed System

System Architecture: The system is divided into two phases:

(i) Team Tactic Estimation

A. Multi Class classification

Perhaps the most utilized abilities of managed AI procedures are for grouping content, utilized in numerous settings like telling if a given eatery survey is positive or negative or deriving in case there is a feline or a canine on a picture. This errand might be partitioned into three areas, binary classification, multiclass classification, and multi-label classification.

Multiclass classification a grouping task with multiple classes, this sort of characterization makes the supposition that each example is relegated to one and only one class. For instance, when we need to order a bunch of organic products which might be oranges or apples, then, at that point an organic product can be an apple or an orange yet not both simultaneously.

Similarly, in a football match, each of the players is assigned a particular position, depending on their strengths, to play on the field. Of the 11 players, 1 is the goalkeeper and the remaining 10 fill various attacking, midfield and defensive positions. Using a multi class classifier, we would differentiate the players into their positions based on their spatio-temporal data [20].

B. Clustering

Clustering is a type of unsupervised learning. An unsupervised learning method is a method wherein we draw references from datasets comprising of data without labelled responses or named reactions. For the most part, it is utilized as a process to discover significant design, informative fundamental cycles, and generative highlights of the given data.

It is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups. It is essentially an assortment of items based on similitude and divergence between them.

The fluid nature of the modern game symbolizes that the player positions during a game aren't as rigid as other sports. The players are only loosely defined into a position. The manager uses various formations that are closely related to each other during a game. Teams align in different formations, depending on what is happening during the play [19]. Clustering will help us to group together, these various formations and predicting which formation is employed to attack or defend the ball.

C. Team Tactic Estimation

After we have determined a player's positions and the various formation used by a manager for a game, we can determine the tactics of a team. The summation of the formations of the team during a team attack and while defending the ball tells us about the team tactics [5].

Table 2. Summary of Various Team Tactics [8]

Tactic Name	Formation	Playing Strengths
Classic	4-4-2	• Strikers excel at movement and finishing • Wingers supply crosses from wide areas • Box to box midfielders
Tiki-Taka	4-3-3	• Midfielders are fluid in passing and movement • Defenders must be good on the ball • Goalkeeper acts as a sweeper
Counter-Attack	4-2-3-1	• Wingers exploit space • Deep midfielders spring attacking moves • Defenders remain solid and compact
Park the Bus	4-3-2-1	• Sole attacker is left isolated • Wingers become default fullbacks • Defenders tuck into central areas
Long Ball Game	4-1-4-1	• Striker must be strong and good in air • Tireless wingers need to attack and defend • Defensive midfielder shields defence
High Press	4-2-3-1	• Defence starts in attack • Midfielders win the ball in opposition half • Defenders play a high line

(ii) Player Evaluation

A. Feature Extraction

Deep learning is a subset of Artificial Intelligence, which is essentially a neural association with no less than three layers. These neural organizations try to reproduce the lead of the human mind however far from planning with its ability allowing it to "learn" from a ton of data. While a neural organization with a solitary layer can regardless make harsh assumptions, additional mysterious layers can help with upgrading precision. Every playing position requires a different skill set of abilities [16]. They have a different job to do in order to keep the team working smoothly. Thus, we need to find required attributes of different players depending on their playing position to evaluate them accordingly.

B. Player Evaluation

After extracting the player positions, the required attributes and statistics of the players, we can use CNN (Convolution Neural Networks) to evaluate the player performance [10]. Doing so will help the coach to understand the strengths and weaknesses of the player. This will help the player to get insight about his match performance and suggest what areas of the field he needs to work on. Evaluating the player fairly is one of the most important jobs of the manager, this system will help the manager to solve the problem. After selecting the profitable formation, the system will select the players based on their performance and the position preferred.

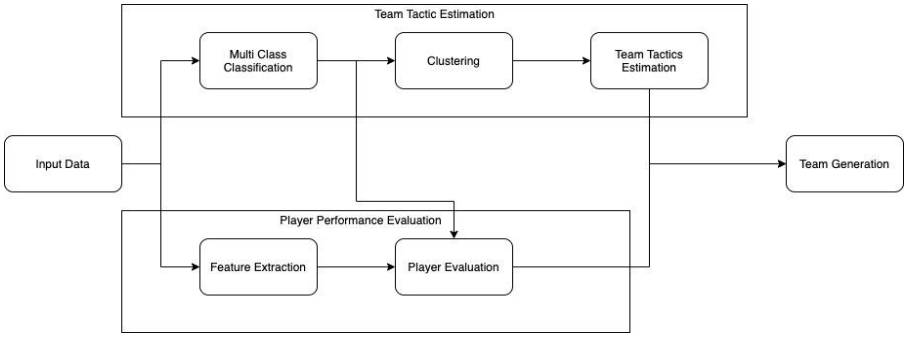

Fig 1. Overall architecture of our approach

Flow chart:

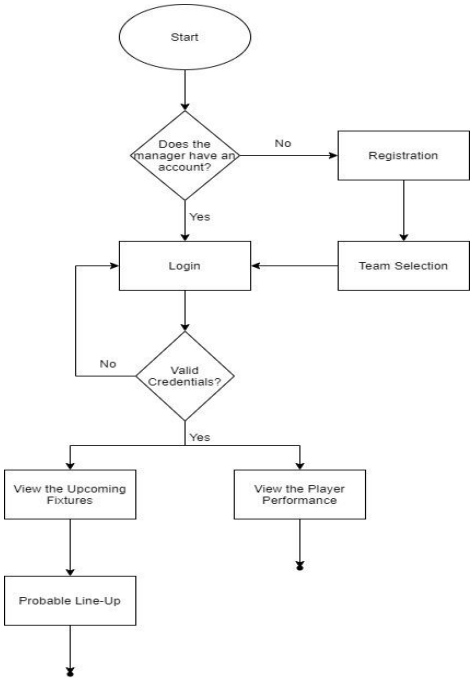

Fig 2. Flow of the system

The system starts by asking a manager to login or register. If the manager has not registered, the system will ask the manager to register and select his team. The manager will be given an option to

select the team from the top leagues of Europe like; English Premier League (England), La Liga (Spain), Bundesliga (Germany), Ligue 1 (France) and various others. Once the manager has selected the team, he will be able to view the scheduled matches as well as the player performances over the past fixtures.

For a football match, data can be mainly divided into two components:

- **Tracking Data:** It is a spatio-temporal dataset [17]. This dataset will contain X and Y coordinates of the players on the field as well as the ball, with respect to the centre of the football pitch
- **Event Data:** This is also a spatio-temporal dataset [9]. It contains information about the various events like pass, shoot and tackle that occurs during the game. The information will include stuff like where on the pitch event occurred and which of the players were involved.

Using the tracking data, the system will be able to predict the player position. We can get the team formation by knowing the position of all the 11 players present on the pitch. So, the system will help the manager to know about the various tactics the opposite team uses.

The event data will help us gather useful information about the players. This information will help us evaluate the player and tell the manager about his strengths and weaknesses.

Combining the information from the event data and the tracking data, system will formulate the tactics the team should implement and the player suitable for that formation.

4 Conclusion and Future Work

In managing a football team, a coach has the main objective of picking out his preferred formation and selects the players that fit properly into that system. There is also no special formula for the coach to evaluate and compare between different players. All the playing positions require a unique skill set and thus it is worthless to compare two different players playing in different positions. We have proposed a system that will predict the playing-11 and the tactic they should employ, while playing against a particular team. With the proposed system we plan to remove the concept of favouritism in a team. Analysing the different systems that are already present, we found out that there is no such system that will help predict the opposite team tactics and selecting our formations according to that.

The scope of this system is very wide. Various other features can be added to our system. Depending on the data from the various other small-scale leagues, the feature of scouting can be added. The manager would be able to scout for the new and upcoming talents, even though the game time of the rookies is not as much compared to the others through the feature of player evaluation. The player evaluation module can also be enhanced based on the data present in our directory as there can be many ways to evaluate a player.

References

[1] Y. Qing et al., "Evaluation of the technical performance of Football players," *Int. J. Environ. Res. Public Health,* vol. 17, no. 2, pp. 604, 2020.

[2] R. Theagarajan and B. Bhanu, "An Automated System for Generating Tactical Performance Statistics for Individual Soccer Players from Videos", *IEEE Trans. Circuit Syst. Video Tech.*, vol. 31, no. 2, pp. 632-646, 2021.

[3] N. Razali et al., "Predicting player position for talent identification in association football", in *IOP Conf. Series: Mater. Sci. Eng.*, 2017, pp. 226.

[4] V. Khaustov and M. Mozgovoy, "Learning Believable Player Movement Patterns from Human Data in a Soccer Game", in *ICACT 2020*, pp. 91-93.

[5] G. Suzuki et al., "Team Tactics Estimation in Soccer Videos Based on a Deep Extreme Learning Machine and Characteristics of the Tactics" in *IEEE Access*, vol. 7, pp. 153238-153248, 2019.
[6] C. Kenneth and R. A. Daniel, "The application of sports technology and sports data for commercial purposes," in *the Use of Technology in Sport-Emerging Challenges*, London, U.K.: Intech Open, 2018.
[7] W. G. Taylor et al., "An elite hockey player's experiences of video-based coaching: A post structuralist reading," *Int. Rev. Sociol. Sport*, vol. 52, no. 1, pp. 112–125, 2017.
[8] A. Jankovi´c et al., "Influence of certain tactical attacking patterns on the result achieved by the teams participants of the 2010 FIFA World Cup in South Africa," *Phys. Culture*, vol. 65, no. 1, pp. 34–45, 2011.
[9] B. Fakhar, H. R. Kanan and A. Behrad, "Event detection in soccer videos using unsupervised learning of spatio-temporal features based on pooled spatial pyramid model," *Multimedia Tools Appl.*, vol. 78, no. 12, pp. 16995–17025, 2019.
[10] X. Wu et al., "A light CNN for deep face representation with noisy labels," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 11, pp. 2884–2896, 2018.
[11] A. Senocak et al., "Part-based player identification using deep convolutional representation and multi-scale pooling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 1732–1739.
[12] D. Memmert and D. Raabe, *Data Analytics in Football: Positional Data Collection, Modelling and Analysis*, Routledge: Abingdon, UK, 2018.
[13] D. Memmert, K. A. Lemmink and J. Sampaio, "Current approaches to tactical performance analyses in soccer using position data", *Sports Med.*, vol. 47, pp. 1–10, 2017.
[14] D. Memmert et al., "A tactical comparison of the 4-2-3-1 and 3-5-2 formation in soccer: A theory-oriented, experimental approach based on positional data in an 11 vs. 11 game set-up", *PLoS ONE* 2019, 14, e021019.
[15] 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, pp. 172–181, 2017.
[16] M. Bush et al., "Evolution of match performance parameters for various playing positions in the English Premier League", *Hum. Mov. Sci.* vol. 39, pp. 1–11, 2015.
[17] V. Khaustov, G. M. Bogdan and M. Mozgovoy, "Pass in Human Style: Learning Soccer Game Patterns from Spatiotemporal Data," in *IEEE Conf. Games*, 2019.
[18] J. Gama et al., "Network analysis and intra-team activity in attacking phases of professional football", *Int. J. Perform. Anal. Sport,* vol. 14, no. 3, pp. 692–708, 2014.
[19] W. Qing et al., "Discerning Tactical Patterns for Professional Soccer Teams. An Enhanced Topic Model with Applications", in *Proc. 21th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining,* 2015, pp. 2197–2206.
[20] J. Gudmundsson and T. Wolle, "Football Analysis Using Spatio-temporal Tools", in *Proc. 20th Int. Conf. Advan. Geographic Inform. Syst.,* 2012, pp. 566–569.