Walisongo Journal of Information Technology
Vol 7 No 2 (2025): 29-38. DOI: 10.21580/wjit.2025.7.2.18698
ISSN 2715-0143 (online) | 2714-9048 (printed)

Data Mining for Football Performance Analysis and Prediction: A Systematic Literature Review

Ikhlasul Aulia Rahman[1], A'ang Subiyakto[2]
[123] Universitas Islam Negeri Syarif Hidayatullah Jakarta
Corresponding Author : ikhlasul.rahman21@mhs.uinjkt.ac.id

Abstract: *Data mining has become an important approach to analyzing football-related data to support performance evaluation and decision-making. However, the application of data mining techniques in football covers various analytical objectives, ranging from match outcome prediction to player performance analysis and physical load monitoring. This study aims to systematically review the application of data mining techniques in football, identify the methods used, and synthesize the main findings of previous studies. A Systematic Literature Review (SLR) was conducted following the guidelines proposed by Kitchenham. The literature search covered four databases: Google Scholar, ScienceDirect, IEEE, and SpringerLink. Studies were selected based on predefined inclusion and exclusion criteria, resulting in 15 articles published between 2003 and 2023. The reviewed studies were examined in terms of their analytical objectives, data mining techniques, and reported findings. The results indicate that the reviewed literature addresses four analytical categories: estimation, forecasting, classification, and clustering. The techniques discussed include neural networks, linear regression, support vector machines, decision trees, k-nearest neighbors, random forests, and K-means clustering. These techniques have been applied to different football-related tasks, including match outcome prediction, attacking effectiveness analysis, player performance assessment, and training load monitoring. The findings highlight the diverse applications of data mining in football analytics while indicating that reported model performance must be interpreted in the context of each study's dataset, prediction target, and evaluation procedure. This review provides an overview of existing applications and identifies the need for more consistent methodological reporting and comparative synthesis in future research.*

Keyword : *Data Mining; Football Analytics; Performance Analysis; Prediction; Systematic Literature Review.*

Copyright © 2025 Walisongo Journal of Information Technology

Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

1. Introduction

The increasing availability of football-related data has created opportunities to examine tactical behaviour and individual player performance through systematic data analysis. In professional soccer, position-tracking data can be used to investigate tactical behaviour, although extracting meaningful insights requires appropriate data management, analytical

methods, and interpretation (Goes, Meerhoff, et al., 2021). Similarly, research on footballer attributes has examined how static characteristics, such as age and height, and dynamic characteristics, such as pass completions and shots on target, can contribute to individual player performance analysis and prediction (Wakelam et al., 2022). These developments illustrate the importance of data-driven approaches in football analytics and provide a relevant context for examining how data mining techniques are applied to football performance analysis and prediction.

Data mining and machine learning techniques have been applied to football analytics to identify patterns in match data and develop predictive models. For example, (Berrar et al., 2024) presented a data- and knowledge-driven framework for developing machine learning models to predict soccer match outcomes, demonstrating the importance of structuring the prediction task and incorporating relevant domain knowledge. Similarly, *Prediction of Football Match Results with Machine Learning* (Rodrigues & Pinto, 2022) investigated predictive models using historical match statistics and player attributes as input variables. These studies illustrate how machine learning can be used to model football match outcomes from multiple sources of information. However, the methods and results of individual studies need to be interpreted in relation to their datasets, prediction objectives, and evaluation procedures rather than assuming that one technique is universally superior.

Beyond match outcome prediction, data mining techniques can also be used to classify match results and identify groups of players with similar performance characteristics. For example, (Fuadi & Alamsyah, 2023) applied the C4.5 classification algorithm, combined with discretization and particle swarm optimization, to predict English Premier League match results. Their study reported that the combined approach achieved an accuracy of 71.05%, compared with 57.24% for the original C4.5 algorithm, within their experimental setting (Fuadi & Alamsyah, 2023). In a different application, (Akhanli & Hennig, 2023) used clustering methods to group football players according to their performance data, producing groups that could support team composition analysis and the identification of players with similar profiles. These studies demonstrate that data mining applications in football extend beyond prediction alone to include the classification of match outcomes and the exploration of similarities among players.

Previous systematic reviews have examined the application of machine learning and data-driven methods in soccer from different perspectives. (Rico-González et al., 2023) reviewed machine learning applications in soccer and categorized the identified studies into injury prediction, performance prediction, and talent prediction. (Goes, Meerhoff, et al., 2021) focused on the use of big data to support tactical performance analysis in professional soccer, while (Wakelam et al., 2022) examined the collection, analysis, and use of footballer attributes for individual player performance analysis and prediction. These reviews provide valuable insights into specific aspects of football analytics. Nevertheless, their differing research objectives and scopes highlight the importance of clearly defining the analytical categories and selection criteria when synthesizing studies on data mining in football. Accordingly, the present review examines the application of data mining techniques in football through four analytical categories: estimation, forecasting, classification, and clustering, based on the scope and categorization adopted in this study.

In light of these considerations, this study aims to systematically review research on the application of data mining techniques in football, identify the methods used, and synthesize the findings reported in the selected studies. Guided by two research questions, the review examines which data mining techniques have been applied to football performance analysis and what findings have been reported by previous research. The selected studies are analyzed according to four categories adopted in this review: estimation, forecasting, classification, and clustering. By organizing the available evidence into these categories, this study seeks to provide

a structured overview of existing applications, clarify the analytical purposes of the methods identified, and highlight potential directions for future research in football analytics.

This study employed a Systematic Literature Review (SLR) to identify, examine, and synthesize research on the application of data mining techniques in football performance analysis and prediction. The review was conducted with reference to the guidelines proposed by (Kitchenham, 2007) and (Kitchenham et al., 2009), which organize the review process into three main phases: planning the review, conducting the review, and reporting the results. During the planning phase, the research questions and study selection criteria were established. The conducting phase involved searching the selected databases, screening the retrieved publications against the eligibility criteria, and examining the relevant studies. Finally, the reporting phase synthesized the findings of the selected studies to address the research questions. Based on the selection results reported in this study, 15 articles published between 2003 and 2023 were included in the review.

This article is divided into four sections. The first section is an introduction that outlines the background of the research problem. The second section describes the research method. The third section provides an overview of the obtained research results and there is an in-depth interpretation and discussion of the research results. The end of this paper closes with a conclusion section.

2. Method

In this study, the Systematic Literature Review method was employed to formulate research questions and conduct a literature review based on guidelines provided by Kitchenham (Kitchenham et al., 2010; Kitchenham & Charters, 2007). According to these guidelines, there are three phases: planning the review phase, conducting the review phase, and reporting the review phase.

Planning the Review Phase

In this phase, research questions (RQ) are formulated to aid in gathering all the necessary information for analyzing various journal articles. The research questions include:

RQ 1. What data mining techniques are employed to conduct performance prediction analysis in football?

RQ 2. What are the research findings from the analysis of football performance predictions?

Conducting the Review Phase

The next stage involves the process of gathering the required data for this research by determining the previously established keywords to be used in searching through databases. The keywords employed are "Prediction," "Performance," "Football," and "Data Mining." In the search process, the databases will be meticulously analyzed to uncover information related to the research topic. The search results will then be extracted and utilized as materials for this study. The following databases are utilized to facilitate the search for relevant journals.

1. Google Scholar
2. Science Direct
3. IEEE
4. Springer Link

Criteria and limitations for input to determine whether data is eligible for use in this study. The following criteria meet the requirements of this research:

Table 1. Inclusion and Exclusion Criteria

Inclusion	Journal published in 2003-2023
	Journal discussing football
	Data processing uses data mining techniques
	Football predictions discuss performance in football.
Exclusion	Journals published outside 2003-2023
	The journal discusses things other than football
	Data is not processed using data mining techniques
	Predictions address other issues

Reporting the Review Phase

Referring to the research questions, information is gathered to identify the application of data mining for predictions in football. Subsequently, each candidate journal article is evaluated against inclusion and exclusion criteria to ensure its relevance. The reported research quality is also assessed by examining the accuracy of technical descriptions in each publication. After this stage, 15 publications have been selected.

Table 2. Total number of scientific articles selected

Google Scholar	Science Direct	IEEE	Springer Link	Total Articles
8	4	2	1	15

3. Result and Discussion

Results of the Search Process and Inclusion and Exclusion Criteria, there are 15 articles that have undergone the selection process with data criteria from 2003 to 2023, and the data used only includes journals related to the application of data mining for predictions in football. In Table 3 below are the selected articles.

Table 3. List of Article Titles

Source	Journal Article	Year
(Campbell et al., 2021)	Analysing the Predictive Capacity and Dose-response of Wellness in Load Monitoring	2021
(Bauer & Anzer, 2021)	Data-driven Detection of Counterpressing in Professional Football: A Supervised Machine Learning Task Based on Synchronized Positional and Event Data with Expert-based Feature Extraction	2021
(Goes, Brink, et al., 2021)	The Tactics of Successful Attacks in Professional Association Football: Large-scale Spatiotemporal Analysis of Dynamic Subgroups Using Position Tracking Data	2021
(Rajesh et al., 2020)	A Data Science Approach to Football Team Player Selection	2020
(Li et al., 2020)	Data-driven Team Ranking and Match Performance Analysis in Chinese Football Super League	2020

(Pantzalis & Tjortjis, 2020)	Sports Analytics for Football League Table and Player Performance Prediction	2020
(Leser et al., 2019)	Finding Efficient Strategies in 3-versus-2 Small-sided Games of Youth Soccer Players	2019
(Bilek & Ulas, 2019)	Predicting Match Outcome According to the Quality of Opponent in the English Premier League Using Situational Variables and Team Performance Indicators	2019
(Young et al., 2019)	The Relationship Between Match Performance Indicators and Outcome in Australian Football	2019
(Zago et al., 2019)	Use of Machine Learning and Wearable Sensors to Predict Energetics and Kinematics of Cutting Maneuvers	2019
(Pappalardo & Cintia, 2018)	Quantifying the Relation Between Performance and Success in Soccer	2018
(Jaspers et al., 2018)	Relationships Between the External and Internal Training Load in Professional Soccer: What Can We Learn from Machine Learning?	2018
(Brooks et al., 2016)	Using Machine Learning to Draw Inferences from Pass Location Data in Soccer	2016
(Montoliu et al., 2015)	Team Activity Recognition in Association Football Using a Bag-of-Words-based Method	2015
(Joseph et al., 2006)	Predicting Football Results Using Bayesian Nets and Other Machine Learning Techniques	2006

RQ1: What data mining techniques are employed to conduct performance prediction analysis in football?

This study involves the analysis of 15 selected journal articles, with a focus on the application of data mining methods such as estimation, forecasting, classification, and clustering. After conducting a search and selecting the articles, the researchers delved into them to understand how these data mining methods have been implemented in the context of football.

Table 4. *List of journals that use Data Mining methods*

Method		Source
Estimation	Neural Network	(Zago et al., 2019)
	Linear Regression	(Zago et al., 2019)
Forecasting	SVM	(Li et al., 2020; Montoliu et al., 2015)
	Neural Network	(Jaspers et al., 2018)
Classification	Decision Tree	(Bilek & Ulas, 2019; Pantzalis & Tjortjis, 2020; Rajesh et al., 2020; Young et al., 2019)
	KNN	(Brooks et al., 2016; Joseph et al., 2006; Montoliu et al., 2015; Pantzalis & Tjortjis, 2020; Rajesh et al., 2020)
	Random Forest	(Bauer & Anzer, 2021; Campbell et al., 2021; Pantzalis & Tjortjis, 2020; Pappalardo & Cintia, 2018; Rajesh et al., 2020)
Clustering	K-means	(Bilek & Ulas, 2019; Goes, Brink, et al., 2021; Leser et al., 2019)

1. Estimation

The study described in (Zago et al., 2019) adopts an estimation method that combines neural network and linear regression algorithms to construct a predictive model. This model is designed with the aim of accurately predicting a specific target, namely the mechanical load that occurs when participants undergo running tests, especially during turns. The research utilizes a sample consisting of 13 professional female players, providing relevant empirical data for the development of this model.

The results of the study are highly impressive, with an accuracy rate exceeding 98.4%. This indicates that the developed model is highly effective in predicting mechanical loads during turns. The success of this model in achieving a high level of accuracy reflects the quality of the data used in the study as well as the capability of the algorithms employed in processing that data. These findings hold significant potential for practical applications in understanding and improving athlete training, particularly for professional female players undergoing running tests. Moreover, the study underscores the strength of using data-driven algorithms and artificial intelligence in scientific research.

2. Forecasting

The forecasting method was utilized in the studies conducted by (Jaspers et al., 2018; Li et al., 2020; Montoliu et al., 2015) employing Support Vector Machine (SVM) and neural network algorithms to build predictive models. In the research by (Montoliu et al., 2015), the SVM algorithm was applied to a sample of 60 matches from the German Bundesliga, encompassing 69,667 individual actions. This study aimed to classify various aspects of the game, including ball possession, fast attacks, and set-piece tactics. The results indicated that the model achieved an impressive accuracy level, ranging from 67% to 93%. These findings offer valuable insights into professional football, aiding in a better understanding of team strategies and performance.

On the other hand, in the study by (Jaspers et al., 2018), the Neural Network algorithm was employed with a sample of 38 professional players from the Dutch first division league over two seasons. The objective of this research was to determine the Rating of Perceived Exertion (RPE), a subjective measurement of the perceived physical effort by players during matches. The Neural Network algorithm was used to process data from the 38 players. The results of this study provide a deeper understanding of how players experience physical exertion during matches, which can be a crucial factor in the treatment and management of players' physical conditions.

3. Classification

The classification method was employed in the studies conducted by (Bauer & Anzer, 2021; Bilek & Ulas, 2019; Brooks et al., 2016; Campbell et al., 2021; Joseph et al., 2006; Montoliu et al., 2015; Pappalardo & Cintia, 2018; Rajesh et al., 2020; Young et al., 2019) utilizing Decision Tree, K-nearest Neighbor (KNN), and Random Forest algorithms to build models for prediction.

In the research by (Brooks et al., 2016), the KNN algorithm was employed with a sample of 20 teams from the Spanish First Division, covering 80% of matches over one season. The objective of this study was to predict attacking success, achieving an accuracy of 87%. Similarly, in the study by (Montoliu et al., 2015), the KNN algorithm was used with a sample of 4 teams from the Spanish First Division over two seasons, focusing on the classification of ball possession, fast attacks, and set-piece tactics. Another study by (Joseph et al., 2006) utilized the KNN algorithm with a sample of 76 matches from the English First Division over two seasons, aiming to observe match outcomes with an accuracy of 50-59%.

In the study by (Campbell et al., 2021), the focus was on predicting training loads. Using a sample comprising 7918 observations over two seasons, this research presented a Random Forest model capable of predicting training loads with an impressive accuracy of over 89%. These results reflect the significant value in understanding how to manage training to enhance athlete performance and reduce injury risks.

On a broader scale, the study by (Pappalardo & Cintia, 2018) focused on match and league ranking outcomes across various professional football leagues in Europe. In this research, the Random Forest algorithm was employed to process data from 6396 matches and 10 million competitions. Although the accuracy results were less than 80%, they still provided valuable insights into understanding factors influencing match outcomes and league rankings in a highly competitive football environment.

In the study by (Bilek & Ulas, 2019), focusing on the English Premier League, a sample of 20 professional teams over one season was used to predict match outcomes. With the Decision Tree algorithm, this research yielded accuracy rates ranging between 68-78%. These findings can assist coaches and team managers in planning strategies and decision-making during the competitive season.

4. Clustering

The clustering method was employed in the studies conducted by (Bilek & Ulas, 2019; Goes et al., 2021; Leser et al., 2019) using the K-means algorithm to build models for prediction. In the research by (Goes et al., 2021), he focus was on understanding the success of attacking tactics in football from a collectivity perspective. This study combined data from 26 professional teams over four seasons to identify successful tactical patterns in attacking strategies. The use of the K-means algorithm helped cluster this data into relevant groups, providing in-depth insights into successful attacking strategies in the world of professional football.

On the other hand, in the study by (Leser et al., 2019), the focus was on young players in the under-15 age category. In this research, the K-means algorithm was used to analyze the efficiency of tactical patterns used by young players. With a sample of 20 players, this study helped identify the most efficient tactical patterns for the development of young players, which could have a significant impact on the development of future football players.

In the study by (Bilek & Ulas, 2019), a different approach was taken, focusing on the English Premier League and match outcomes. In this research, the K-means algorithm was used to analyze data from 20 professional teams over one season. The goal was to predict match outcomes, with accuracy ranging between 68-78%. These results provide important insights into the factors influencing match outcomes in top-tier leagues such as the English Premier League.

RQ2: What are the research findings from the analysis of football performance predictions?

In the journal articles (Goes, Brink, et al., 2021), (Leser et al., 2019), (Bilek & Ulas, 2019) (Leser et al., 2019), (Bilek & Ulas, 2019), (Pappalardo & Cintia, 2018), (Brooks et al., 2016; Li et al., 2020), (Joseph et al., 2006; Young et al., 2019) that focus on match performance using data mining, there are five journal articles that can accurately predict match performance. The journal articles (Pappalardo & Cintia, 2018), (Bilek & Ulas, 2019), and (Joseph et al., 2006) aim to predict match outcomes, while the studies by (Goes, Brink, et al., 2021) and (Brooks et al., 2016) predict the success of attacks. On the other hand, the studies by (Campbell et al., 2021; Jaspers et al., 2018) are related to training loads, health, and match performance. Additionally,

the study by (Zago et al., 2019) focuses on identifying the energy required by players during matches.

1. **Match Result Prediction**

In the studies by (Bilek & Ulas, 2019; Joseph et al., 2006; Pappalardo & Cintia, 2018) consideration is given to the variables used to predict match outcomes. Meanwhile, in (Bilek & Ulas, 2019; Montoliu et al., 2015; Pappalardo & Cintia, 2018), the focus is on the analysis of frequent occurrences (such as ball possession, passing, and shooting). On the other hand, (Joseph et al., 2006) concentrates on player presence, playing location, and the quality of the opposing team.

The study by (Bilek & Ulas, 2019) employs K-means clustering with situational variables considered as determinants in 22 out of 380 matches in the English Premier League (EPL). This method can predict match outcomes with an accuracy of 68% - 78%. In the research by (Pappalardo & Cintia, 2018), which covers 6,000 out of 10 million matches in 6 European leagues, it was found that victories and defeats can be influenced by team performance and final league standings. In the study by (Brooks et al., 2016), classical notation data (such as passes, ball possession, and pass locations) from Spanish football teams were utilized. Passes were categorized into heat maps, identifying patterns used by these teams for analysis.

2. **Identification of Attack Effectiveness**

The study conducted by (Leser et al., 2019) focuses on analyzing the efficiency of tactical patterns and detecting situations with high potential for success. In the research by (Bauer & Anzer, 2021; Pantzalis & Tjortjis, 2020; Rajesh et al., 2020), the emphasis is on team performance and the individual performance of players in matches. Another study, carried out by (Goes et al., 2021), categorizes successful attacks by observing 11,187 attacks (both successful and unsuccessful), and the results indicate that successful attacks tend to occur when players in the opposing team collaborate less effectively.

Research aimed at predicting match outcomes has two main focuses. Firstly, predicting match outcomes using frequently occurring variables (such as ball possession, passing, and shooting) (Bilek & Ulas, 2019; Brooks et al., 2016; Joseph et al., 2006; Montoliu et al., 2015; Pappalardo & Cintia, 2018) [24], [27], [29], [31]. Secondly, identifying the effectiveness of attacks using player position data (Bauer & Anzer, 2021; Goes et al., 2021; Leser et al., 2019; Pantzalis & Tjortjis, 2020; Rajesh et al., 2020). Additionally, data mining can be employed to identify the most influential factors in explaining health and training loads. This can help reduce the amount of data that needs to be collected in the future and enhance monitoring efficiency. In other words, data mining can assist us in a more efficient understanding and management of health and training aspects.

4. Conclusion

Research on the application of data mining in football has explored various analytical techniques to understand player performance, team strategies, and match outcomes. These techniques include estimation, prediction, classification, and clustering, which are applied to player, team, and match data to identify patterns and support evidence-based decision-making. Estimation techniques, including neural networks and linear regression, have been used to estimate mechanical loads associated with players' turning movements. One study reported an accuracy exceeding 98.4%, indicating the potential application of these techniques in athlete monitoring and training, particularly among professional female football players. Prediction methods have also been applied to analyze match-related variables, including ball possession,

fast attacks, and set-piece situations. Reported accuracy levels range from 67% to 93%, although these results should be interpreted in the context of the specific datasets, prediction targets, and evaluation methods used in each study. Other research has examined the Rating of Perceived Exertion (RPE) to assess players' perceived physical effort.

Classification algorithms, such as Decision Trees, k-nearest neighbors (KNN), and Random Forest, have been employed to analyze match outcomes, attacking success, and factors associated with team and player performance. These approaches can provide information to support coaches and team managers in evaluating performance and developing tactical strategies. Meanwhile, clustering techniques, including K-means, have been used to identify patterns in attacking tactics and examine tactical effectiveness. Such analyses may also support player development and investigations into factors associated with match outcomes.

Overall, data mining offers opportunities to analyze football performance from physical, technical, and tactical perspectives. Its application can support performance evaluation, strategic planning, and player monitoring. Nevertheless, the usefulness and generalizability of the resulting insights depend on data quality, the suitability of the analytical methods, and the validity of the evaluation procedures. Further research should identify the limitations of existing approaches and establish how alternative data mining techniques can address specific analytical challenges in football.

5. References

Akhanli, S. E., & Hennig, C. (2023). Clustering of football players based on performance data and aggregated clustering validity indexes. *Journal of Quantitative Analysis in Sports*, *19*(2). https://doi.org/10.1515/jqas-2022-0037

Bauer, P., & Anzer, G. (2021). Data-driven detection of counterpressing in professional football: A supervised machine learning task based on synchronized positional and event data with expert-based feature extraction. *Data Mining and Knowledge Discovery*, *35*(5), 2009–2049.

Berrar, D., Lopes, P., & Dubitzky, W. (2024). A data- and knowledge-driven framework for developing machine learning models to predict soccer match outcomes. *Machine Learning*, *113*(10). https://doi.org/10.1007/s10994-024-06625-9

Bilek, G., & Ulas, E. (2019). Predicting match outcome according to the quality of opponent in the English premier league using situational variables and team performance indicators. *International Journal of Performance Analysis in Sport*, *19*(6), 930–941. https://doi.org/10.1080/24748668.2019.1684773

Brooks, J., Kerr, M., & Guttag, J. (2016). Using machine learning to draw inferences from pass location data in soccer. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *9*(5), 338–349.

Campbell, P. G., Stewart, I. B., Sirotic, A. C., Drovandi, C., Foy, B. H., & Minett, G. M. (2021). Analysing the predictive capacity and dose-response of wellness in load monitoring. *Journal of Sports Sciences*, *39*(12), 1339–1347. https://doi.org/10.1080/02640414.2020.1870303

Fuadi, M. B. A., & Alamsyah, A. (2023). Optimization of the C4.5 Algorithm Using Particle Swarm Optimization and Discretization in Predicting the Results of English Premier League Football Matches. *Journal of Advances in Information Systems and Technology*, *4*(2). https://doi.org/10.15294/jaist.v4i2.59531

Goes, F. R., Brink, M. S., Elferink-Gemser, M. T., Kempe, M., & Lemmink, K. A. P. M. (2021). The tactics of successful attacks in professional association football: large-scale spatiotemporal analysis of dynamic subgroups using position tracking data. *Journal of Sports Sciences*, *39*(5), 523–532. https://doi.org/10.1080/02640414.2020.1834689

Goes, F. R., Meerhoff, L. A., Bueno, M. J. O., Rodrigues, D. M., Moura, F. A., Brink, M. S., Elferink-Gemser, M. T., Knobbe, A. J., Cunha, S. A., Torres, R. S., & Lemmink, K. A. P. M. (2021). Unlocking the potential of big data to support tactical performance analysis in professional soccer: A

systematic review. In *European Journal of Sport Science* (Vol. 21, Number 4). https://doi.org/10.1080/17461391.2020.1747552

Jaspers, A., De Beéck, T. O., Brink, M. S., Frencken, W. G. P., Staes, F., Davis, J. J., & Helsen, W. F. (2018). Relationships between the external and internal training load in professional soccer: what can we learn from machine learning? *International Journal of Sports Physiology and Performance*, *13*(5), 625–630.

Joseph, A., Fenton, N. E., & Neil, M. (2006). Predicting football results using Bayesian nets and other machine learning techniques. *Knowledge-Based Systems*, *19*(7), 544–553. https://doi.org/https://doi.org/10.1016/j.knosys.2006.04.011

Kitchenham, B. (2007). Guidelines for performing systematic literature reviews in software engineering. *Technical Report, Ver. 2.3 EBSE Technical Report. EBSE*.

Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering - A systematic literature review. In *Information and Software Technology* (Vol. 51, Number 1). https://doi.org/10.1016/j.infsof.2008.09.009

Leser, R., Hoch, T., Tan, X., Moser, B., Kellermayr, G., & Baca, A. (2019). Finding efficient strategies in 3-versus-2 small-sided games of youth soccer players. *Kinesiology*, *51*(1.), 110–118.

Li, Y., Ma, R., Gonçalves, B., Gong, B., Cui, Y., & Shen, Y. (2020). Data-driven team ranking and match performance analysis in Chinese Football Super League. *Chaos, Solitons & Fractals*, *141*, 110330. https://doi.org/https://doi.org/10.1016/j.chaos.2020.110330

Montoliu, R., Martín-Félez, R., Torres-Sospedra, J., & Martínez-Usó, A. (2015). Team activity recognition in Association Football using a Bag-of-Words-based method. *Human Movement Science*, *41*, 165–178. https://doi.org/https://doi.org/10.1016/j.humov.2015.03.007

Pantzalis, V. C., & Tjortjis, C. (2020). Sports Analytics for Football League Table and Player Performance Prediction. *2020 11th International Conference on Information, Intelligence, Systems and Applications (IISA*, 1–8. https://doi.org/10.1109/IISA50023.2020.9284352

Pappalardo, L., & Cintia, P. (2018). Quantifying the relation between performance and success in soccer. *Advances in Complex Systems*, *21*(03n04), 1750014.

Rajesh, P., Bharadwaj, Alam, M., & Tahernezhadi, M. (2020). A Data Science Approach to Football Team Player Selection. *2020 IEEE International Conference on Electro Information Technology (EIT)*, 175–183. https://doi.org/10.1109/EIT48999.2020.9208331

Rico-González, M., Pino-Ortega, J., Méndez, A., Clemente, F. M., & Baca, A. (2023). Machine learning application in soccer: a systematic review. In *Biology of Sport* (Vol. 40, Number 1). https://doi.org/10.5114/biolsport.2023.112970

Rodrigues, F., & Pinto, Â. (2022). Prediction of football match results with Machine Learning. *Procedia Computer Science*, *204*. https://doi.org/10.1016/j.procs.2022.08.057

Wakelam, E., Steuber, V., & Wakelam, J. (2022). The collection, analysis and exploitation of footballer attributes: A systematic review. *Journal of Sports Analytics*, *8*(1). https://doi.org/10.3233/jsa-200554

Young, C. M., Luo, W., Gastin, P., Tran, J., & Dwyer, D. B. (2019). The relationship between match performance indicators and outcome in Australian Football. *Journal of Science and Medicine in Sport*, *22*(4), 467–471. https://doi.org/https://doi.org/10.1016/j.jsams.2018.09.235

Zago, M., Sforza, C., Dolci, C., Tarabini, M., & Galli, M. (2019). Use of machine learning and wearable sensors to predict energetics and kinematics of cutting maneuvers. *Sensors*, *19*(14), 3094.