Graduate Journal of Interdisciplinary Research, Reports & Reviews
(Grad.J.InteR[3])
ark:/85846/JPR3.2026AR.04010001

ISSN(E): 2584-2919
Vol.04, No.01 (2026)
pp.01-11

Goal and Non-goal Impact Oriented Machine Learning Based Player Performance Estimation in Soccer Sports

Karun Madan[1], Kavita Taneja[1] Harmunish Taneja[2]

[1]*Department of Computer Science and Applications, Panjab University, Sector 14, Chandigarh, India*
[2]*Department of Computer Science and Applications, DAV College, Sector 10, Chandigarh, India*
[*]*Corresponding author: karanmadan9@gmail.com*

(Received:14-06-2026; Accepted: 12-07-2026; Published Online:25-07-2026)

ABSTRACT: Background: Many machine learning and data mining methods are available to estimate the teams' and players' performance at various levels by using the historical data.
Purpose: Estimating the future performance of the soccer players with adequate accuracy can aid in taking critical decisions while selecting an initial team line-up for the upcoming matches.
Method: In the proposed work, feature selection has been done conceptually and purely knowledge-driven by the experts to categorize the features into two groups: Non-goal impact and Goal impact. Subsequently, the conjunction of the output of four eminent machine learning based feature selection techniques, i.e., Lasso, Mutual Information, Random Forest, and RFECV, has been performed to obtain common features. These features have been presented to three machine learning models, i.e., Random Forest, Gradient Boost, and Linear regression, for three positional player categories, i.e., defenders, midfielders, and attacking forwards.
Results: The best-performing model, i.e., Random Forest, with an average accuracy of 93.45% (average accuracy of Non-goal impact and goal impact for all three playing positions), has been used to estimate two types of player performance ratings, i.e. non-goal impact based ratings (NGI ratings) and goal impact based ratings (GI ratings), to select the team line-up for the upcoming game.
Conclusion: Using mere machine learning for estimating soccer player performance is not novel, as it may result in goal-biased or position-biased ratings. Most of the previous work focused on goal impact based features, where defenders or midfielders appeared far weaker contributors in comparison to attacking players. The proposed work also gave due importance to the features having non-goal impact on the game along with traditional goal impact based features.

Keywords: Machine learning, Soccer team selection, Feature selection, Goal impact based ratings, Non-goal impact based ratings, Estimation of player performance.

1. Introduction

The huge amount of available soccer data in this digital world can be utilized to improve the performance of the team and players [1]. There are many machine learning and data mining methods available to forecast teams and players associated at various levels by using the historical data [2]. Driven by the ever-growing availability of computational resources and data, the field of sports analytics has been expanding rapidly. Hence, statisticians and researchers are attracted to these techniques and methods to use this gigantic amount of soccer data to assist coaches, players, and managers in estimating the results of future events [3].

Being one of the most data-driven and strategically convoluted sports, soccer offers a unique domain, where the estimation of player performance can bring high-dimensional insights. The conventional methods of evaluating the abilities of the players rely deeply on basic human judgment and statistics from the past performance of the players. But these traditional methods can result in biasness and imperfections. In contrast, machine learning based Intelligent computing techniques offer a fully scalable and efficient method for

© 2026 Author(s). This article is published under the CC-BY license at http://jpr.vyomhansjournals.com.

analyzing a broad range of data and recognizing patterns, which otherwise may not be directly noticeable to the managers, coaches, or other enthusiasts.

In the proposed approach, a machine learning based soccer sports management component capable of estimating the soccer player's performance has been designed and developed that can estimate the individual soccer player's performance in the form of ratings ranging from 0 to 100.

2. Related Work

Researchers have been trying to forecast player and team performance in soccer sports for a long time. Therefore, there has been a decent amount of related literature available in this field. In this segment, the summary of the related previous literature is presented with respect to soccer sports. Amdt & Brefeld [4] presented a regression-based approach to estimate the future romance of the football players by identifying the relevant features using a recursive feature elimination strategy on the dataset taken from the German Bundesliga 2016. This dataset is very sparse for identifying the player models. Malikov and Kim [5] discussed actual goals along with expected goals per game by employing ridge regression machine learning models. The reason behind choosing ridge regression over traditional regression is the ability of ridge regression to handle correlated variables and overlapping. The review of player performance has been done in five European football leagues and the top Russian Premier League. The work categorized top players into two classes, i.e., consistent performers and efficient scorers. Kulakou et al. [6] addressed the issue of estimating the soccer player's performance using a commercially accepted digital monitoring system named PmSys by collecting the self-reported wellness parameters. The dataset belongs to two seasons of Norwegian women's soccer team. The work compared two approaches, i.e., univariate and multivariate approaches, to test time series models. The effect of training the entire team has been compared with training the individual player in terms of match outcome.

Wiik et al. [7] handled the issue of deriving peaks in football players' potential by training a long short-term recurrent neural network model over the dataset of two Norwegian football teams. The machine learning method, long short-term memory i.e. LSTM has been used for time series analysis. Bergkamp et al. [8] conducted a study to investigate the process of soccer player talent identification done by 125 scouts. According to the study, during the scouting process, the early indicators of performance are not adequate. The main point of the study is that it assessed the talent recognition of the player by self-report. According to Dijkhuis et al. [9], the tracking data of the soccer players from 302 soccer matches of the Dutch Eredivisie season 2018-19 have been employed for this work. Random forest and decision tree algorithms were used to build the predictive model. The model has been devised to support the coaches and other stakeholders in making informed decisions about the players. Similarly, Carpita et al. [10] presented work that aims at investigating the role of seasonality in dropping or enhancing player performance through the clustering of the variables around the latent variables approach.

Dick et al. [11] presented two approaches, i.e., the trajectory estimation model to estimate the movement of the players based on the tracking match data and the outcome estimation model of the ball-possession phase, along with interactions between the soccer players using a graph recurrent neural network (GRNN). Stafylidis et al. [12] worked on redefining the meaning of success by analyzing the relationship between the match results and diverse performance indicators using a regression model on data from the first round of the 2020-21 season of the Greek soccer league. The study identified the crucial factors influencing the chances of victory. Pantzalis and Tjortis [13] worked on predicting team and player performance and carried out a comparison between the performance of the previous season and the current one. The study identified and recorded the personal skills while differentiating the excellent from the average defender. The study also revealed that some attacking skills and attacking match actions surprisingly have an impact on the ratings of the defenders.

© 2026 Author(s). This article is published under the CC-BY license at http://jpr.vyomhansjournals.com.

3. Methodology

The main objective of the proposed study is to design and develop a machine learning based soccer sports management component capable of estimating the soccer player's performance. The purposive sampling strategy with positional stratification has been employed to design experimental dataset by ensuring adequate representation across all four soccer playing positions, i.e., defensive, midfield, attacking and goalkeeper. The data of 423 soccer players has been taken from the publicly available FIFA22 Kaggle dataset [14] including 131 Defenders, 96 Forwards, 153 Midfielders, and 43 goalkeepers to provide adequate positional diversity.

Unlike the previous work that estimate the outcome in the form of win, lose, or draw, this study focuses mainly on estimating the individual performance of the soccer players in terms of ratings ranging from 0 to 100 using player statistics. Moreover, conventional approaches relied mainly on goal based impact, where defenders or midfielders appeared to be far weaker contributors in comparison to aggressive attacking forward players.

The proposed approach does not appreciate the blind use of customary machine learning techniques. Instead, the proposed work must guide the machine learning model. The contribution of a defender without any assists cannot be compared with that of the attacking frontline player who contributed a few successful assists in the game. Using mere machine learning is not novel; if we put all the features available in the dataset in the machine learning algorithm, then many irrelevant features would result in noisy estimations, goal-biased or position-biased ratings. Due to these weak, noisy predictions of the ratings of soccer player performance, the finally selected team would not justify its purpose in the upcoming games. So, opinion from soccer experts have been included for the initial categorization of the features into two categories, i.e., non-goal impact and goal impact categories. Initially, the features are chosen conceptually and categorized into two groups, i.e., goal impact based and non-goal impact based. Both the feature sets for non-goal impact and goal impact have been presented in Tables 1 and 2, respectively.

Most of the previous work focused majorly on the features that have goal impact on the game. The proposed work considered not only features having goal impact, but also gave due importance to the features having non-goal impact on the game.

Initially, expert-based pre-filtering has been done on the full set of features and categorized the features into two groups: Non-goal impact and Goal impact. The non-goal impact feature set comprises 30 features, but the Goal impact feature set has only 10 features. The proposed work employed a hybrid feature selection approach that used feature selection techniques only for Non-goal impact features and not for Goal impact features. Since the number of features is already on the lower side in the case of Goal impact, the scalability and reliability of the feature selection technique would be compromised for low-dimensional data [15].

As a result of the hybrid feature selection approach, only Non-goal impact features have been presented to the four feature selection techniques, and Goal impact features have been fed directly into the models based on their positional player categories, i.e., defenders, midfielders, and attacking forwards. For non-goal impact, the four feature selection techniques used in the work are Lasso, Mutual information, Random forest, and RFECV. The common features from the four feature selection techniques have been retrieved using conjunction for all three positional player categories. This selected set of features is presented to three models, i.e., Random Forest, Gradient Boost, and Linear regression, for the same three positional player categories, i.e., defenders, midfielders, and attacking forwards, to estimate the non-goal impact based player performance ratings (NGI ratings).

On the contrary, the goal impact based features for all three positional player categories, i.e., defenders, midfielders, and attacking forwards, are fed directly to three similar models, i.e., Random Forest, Gradient Boost, and Linear Regression, to estimate the goal impact based player performance ratings (GI ratings). The dataset was divided into training and testing subsets using a 90:10 ratio. To minimize the risk of overfitting,

© 2026 Author(s). This article is published under the CC-BY license at http://jpr.vyomhansjournals.com.

Table 1: List of features that offer non-goal impact on ratings of three categories of soccer players.

S. no.	Feature	Defensive Players	Midfield Players	Attacking Players
1	Pace	Less Important	Less Important	Important
2	Passing	Less Important	Important	Less Important
3	Dribbling	Less Important	Important	Important
4	Defending	Important	Less Important	Less Important
5	Defending Marking Awareness	Important	Less Important	Less Important
6	Mentality Aggression	Important	Less Important	Less Important
7	Physic	Important	Less Important	Less Important
8	Defending Standing Tackle	Important	Less Important	Less Important
9	Defending Sliding Tackle	Important	Less Important	Less Important
10	Potential	Important	Less Important	Less Important
11	Skill Long Passing	Important	Less Important	Less Important
12	Mentality Composure	Important	Less Important	Less Important
13	Weak Foot	Important	Less Important	Less Important
14	Power Jumping	Important	Less Important	Less Important
15	Power Strength	Important	Less Important	Less Important
16	Skill Dribbling	Less Important	Important	Important
17	Mentality Vision	Less Important	Important	Important
18	Skill Curve	Less Important	Important	Important
19	Preferred Foot	Less Important	Important	Important
20	Skill Fk Accuracy	Less Important	Important	Less Important
21	Movement Agility	Less Important	Important	Important
22	Movement Reactions	Important	Important	Important
23	Attacking Short Passing	Less Important	Important	Important
24	Attacking Long Passing	Less Important	Important	Less Important
25	Attacking Crossing	Less Important	Important	Important
26	Skill Moves	Less Important	Important	Less Important
27	Mentality Interceptions	Important	Important	Less Important
28	Movement Balance	Less Important	Important	Important
29	Power Stamina	Important	Important	Less Important
30	Skill Ball Control	Less Important	Important	Important

the proposed system incorporates multiple feature selection techniques to manage the number of selected features and ensures elimination of irrelevant attributes, thereby reducing the complexity of the model.

The features can have different importance in different playing positions; for instance, shots by strikers are not of equal significance as tackles for defenders. The proposed work employed separate models for different positions, i.e., three different models for defenders, midfielders, and attacking players, respectively. Employing separate models for different roles would improve the team selection process considerably. Almost all predictive systems appear the same at the basic level; the proposed system should capture intelligent computing-based strategic planning methods that other systems ignored. Finally, the proposed work provides two types of ratings, i.e., goal impact based ratings (GI ratings) and Non-goal based impact ratings (NGI ratings). Final ratings have been obtained by the fusion (averaging scores) of both the estimated NGI and GI player performance ratings. These estimated player performance ratings have been used to select the team line-up for the upcoming game.

4. Results and Discussion

Almost all the predictive systems use machine learning, but the novelty lies in the identification and categorization of the features before feeding those features to machine learning techniques for automated feature selection. Solitary measurement of performance should not be the final goals score-line or the assists for the goal score-line. Hence, the proposed estimated player

© 2026 Author(s). This article is published under the CC-BY license at http://jpr.vyomhansjournals.com.

Table 2: List of features that offer goal impact on ratings of three categories of soccer players.

S. no.	Feature	Defensive Players	Midfield Players	Attacking Players
1	Shooting	Less Important	Less Important	Important
2	Power_Shot_Power	Less Important	Less Important	Important
3	Mentality_Positioning	Less Important	Important	Important
4	Attacking_Finishing	Important	Important	Important
5	Attacking_Volleys	Important	Important	Important
6	Mentality_Penalties	Important	Important	Important
7	Movement_Acceleration	Important	Important	Important
8	Movement_Sprint_Speed	Less Important	Important	Important
9	Power_Long_Shots	Important	Important	Important
10	Attacking_Heading_Accuracy	Important	Less Important	Important

overall rating has been a combination of the goal impact based rating and non-goal impact based rating.

Goals and assists should never be the only performance indicators, but off-ball movement, Pressing range, and important pressure passes should also be considered as non-goal impact factors are equally important in the match context as the goal impact based factors. Individual sports depend primarily on the points scored by the player, but in team sports, the setup is different, as holistic team performance means much more than only a single high-performing player. Goals and assists should not be the direct measure of performance impact indicators, as they may be player-position-biased, and the frequency of goals and assists is very low compared to all the other events of a soccer match. Hence, a low-frequency event such as a goal or assist can be a partial impact indicator of the player's performance and should be in alliance with the non-goal impact events to indicate the player's performance.

In the proposed work, the feature selection has been done in two phases. Firstly, expert-based pre-filtering has been done on the full set of features and categorized the features into two groups: Non-goal impact and Goal impact. After initial pre-filtering, the non-goal impact feature set comprises 30 features, but the Goal impact feature set has only 10 features. As the number of features is already on the lower side in the case of Goal impact, in the second phase, a hybrid feature selection approach was used, where feature selection techniques were applied only for Non-goal impact features and not for Goal impact features.

As a result of the hybrid feature selection approach, only Non-goal impact features have been presented to the four feature selection techniques, and Goal impact features have been fed directly into the models based on their positional player categories, i.e., defenders, midfielders, and attacking forwards. For non-goal impact, the four feature selection techniques used in the work are Lasso, Mutual information, Random forest, and RFECV (Table 3).

The common non-goal impact features from the four feature selection techniques have been retrieved using a conjunction for all three positional player categories. Seven features finally selected in the defender category are Defending Marking Awareness, Movement Reactions, Defending Standing Tackle, Defending Sliding Tackle, Power Jumping, Defending, and Power Strength. In the same way, eight features have been finally selected for the midfielder category. These eight features are Passing, Skill_Dribbling, Skill_Fk_Accuracy, Movement_Agility, Attacking_Long_Passing, Mentality_Interceptions, Skill_Long_Passing, and Power_Stamina. Similarly, for attacking forward position, the seven selected features are Pace, Attacking_Crossing, Skill_Dribbling, Skill_Curve, Attacking_Short_Passing, Movement_Balance, and Skill_Ball_Control. This selected set of features is presented to three models, i.e., Random Forest, Gradient Boost, and Linear Regression, for the same three positional player categories, i.e., defenders, midfielders, and attacking forwards, to estimate the

© 2026 Author(s). This article is published under the CC-BY license at http://jpr.vyomhansjournals.com.

Table 3: List of non-goal impact based features selected by four feature selection techniques for three positional player categories, i.e., defenders, midfielders, and attacking forwards.

Player Position	Lasso	Mutual information	Random forest	RFECV
Defenders	• Movement Reactions • Power Jumping • Power Strength • Defending Sliding Tackle • Defending • Defending Stand Tackle • Defending Marking Awareness	• Movement Reactions • Power Strength • Defending • Defending Standing Tackle • Attacking Short Passing • Skill Long Passing • Power Jumping • Physic • Defending Sliding Tackle • Defending Marking Awareness	• Defending • Defending Standing Tackle • Power Jumping, • Defending Sliding Tackle, • Power Strength • Attacking Short Passing • Mentalilty Composure • Movement Reactions • Defending Marking Awareness	• Skill Long Passing • Defending • Defending Sliding Tackle • Power Jumping • Power Strength • Movement Reactions • Defending Standing Tackle • Defending Marking Awareness
Midfielder	• Skill Long Passing • Passing,Skill Dribbling • Skill Fk Accuracy • Movement Agility • Attacking Short Passing • Attacking Long Passing • Mentality Interceptions • Power Stamina	• Passing • Dribbling • Skill Dribbling • Skill Fk Accuracy • Movement Agility • Attacking Long Passing • Skill Long Passing • Mentality Interceptions • Power Stamina • Skill Ball Control	• Skill Fk Accuracy • Passing • Dribbling • Skill Dribbling • Skill Curve • Movement Agility • Movement Reactions • Attacking Short Passing • Attacking Long Passing • Skill Long Passing • Mentality Interceptions • Power Stamina • Skill Ball Control	• Skill Curve • Passing • Skill Dribbling , • Skill Fk Accuracy • Movement Agility • Movement Reactions • Attacking Long Passing • Skill Long Passing • Mentality Interceptions • Power Stamina • Skill Ball Control

© 2026 Author(s). This article is published under the CC-BY license at http://jpr.vyomhansjournals.com.

Contd...

Player Position	Lasso	Mutual information	Random forest	RFECV
Attacking Forward	• Pace • Skill Dribbling • Skill Curve • • • Movement Reactions • Attacking Short Passing • Attacking Crossing • Movement Balance • Skill Ball Control	• Pace • Skill Dribbling • Skill Curve • Preferred Foot • Movement Agility • Movement Reactions • Attacking Short Passing • Attacking Crossing • Movement Balance • Skill Ball Control	• Pace • Skill Dribbling • Skill Curve • Movement Agility • Movement Reactions • Attacking Short Passing • Attacking Crossing • Movement Balance • Skill Ball Control	• Pace • Skill Dribbling • Skill Curve • Movement Agility • Attacking Short Passing • Attacking Crossing • Movement Balance • Skill Ball Control

non-goal impact based player performance ratings (NGI ratings).

On the contrary, the goal impact based features for all three positional player categories, i.e., defenders, midfielders, and attacking forwards, are fed directly to three similar models, i.e., Random forest, Gradient boost, and Linear regression, to estimate the goal impact based player performance ratings (GI ratings). For defenders, the feature set comprises six features, i.e., Attacking_Finishing, Attacking_Volleys, Attacking_Heading_Accuracy, Mentality_Penalties, Movement_Acceleration, and Power_Long_Shots. Similarly, for midfielders, the feature set has seven features, i.e., Mentality_Positioning, Attacking_Finishing, Movement_Sprint_Speed, Attacking_Volleys, Mentality_Penalties, Movement_Acceleration, and Power_Long_Shots. The feature set for attacking forwards comprises ten features, i.e., Shooting, Power_Shot_Power, Mentality_Positioning, Attacking_ Finishing, Attacking_Volleys, Mentality_Penalties, Movement_Acceleration, Movement_Sprint_Speed, Power_Long_Shots, and Attacking_Heading_Accuracy.

This selected set of features is presented to three models, i.e., Random Forest, Gradient Boost, and Linear Regression, for the same three positional player categories, i.e., defenders, midfielders, and attacking forwards, to estimate the non-goal impact based player performance ratings (NGI ratings) and goal impact based player performance ratings (GI ratings). In all three positional player categories, i.e., defenders, midfielders, and attacking forwards, Random Forest outperformed the other two machine learning models, i.e., Gradient Boost and Linear regression, as shown in Table 4. Hence, the final model selected to estimate the ratings is Random Forest. Comparison of goal impact and non-goal impact has been shown in graphical form (Figs. 1 to 3) for each positional player category, respectively.



Figure 1: Accuracy achieved in the defensive player category.

© 2026 Author(s). This article is published under the CC-BY license at http://jpr.vyomhansjournals.com.

Table 4: Performance comparison of machine learning models for both non-goal impact and goal impact in all three positional player categories.

Player Position	Machine Learning Technique	Accuracy Achieved for Non-Goal Impact	Accuracy Achieved for Goal Impact
Defenders	Random Forest	96.88%	89.15%
	Gradient Boost	88.50%	79.15%
	Linear Regression	81.67%	72.43%
Midfielders	Random Forest	92.49%	89.55%
	Gradient Boost	83.76%	82.94%
	Linear Regression	77.13%	74.81%
Attacking Forwards	Random Forest	94.85%	97.78%
	Gradient Boost	88.73%	91.44%
	Linear Regression	86.12%	90.82%

There is an interesting observation from the outcome of the machine learning models. In the defenders' positional player category, the accuracy achieved by the best model, i.e., Random Forest, for Non-goal impact based features (96.88%) is significantly higher than the accuracy achieved for goal impact based features (89.15%). It reveals the significantly higher contribution of non-goal impact based features over goal impact based features for estimating the player ratings in this category of players, as shown in Fig.1.

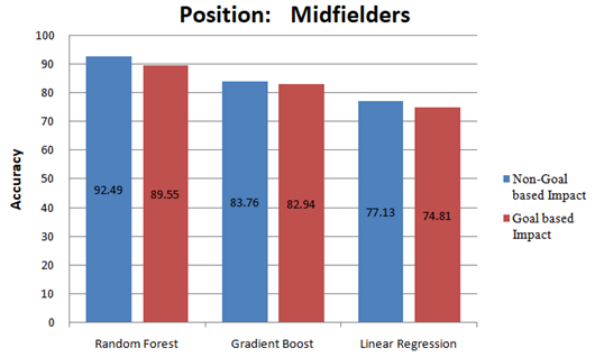


Figure 2: Accuracy achieved in the midfield player category.

Similarly, the midfielder category has also shown the same pattern. In the midfielder positional player category, the accuracy achieved by the best model, i.e., Random Forest, for Non-goal impact based features (92.49%) is comparatively higher than the accuracy achieved for goal impact based features (89.55%). It reflects the better effect of non-goal impact based features over goal impact based in this category of players, as shown in Fig. 2.

However, for the attacking forward category, there is an opposite trend, as the accuracy achieved by Random forest, for goal impact based features (97.78%) is higher than the accuracy achieved for non-goal impact based features (94.85%), as shown in Fig. 3.



Figure 3: Accuracy achieved in the attacking forward player category.

Final ratings have been obtained by the fusion (averaging scores) of both the estimated NGI and GI player performance ratings. These fusion-based ratings provide more reliability and clarity about the performance of individual players. These estimated player performance ratings have been used to select the team line-up for the upcoming game.

5. Conclusion and Future Work

Modern soccer is not only about skill-set, but also demands high-level performance analysis in terms of players and teams. The conventional machine learning based feature selection approaches used in the previous work automatically favors goal impact based features;

© 2026 Author(s). This article is published under the CC-BY license at http://jpr.vyomhansjournals.com.

as a result, the whole model is inclined towards goal impact based contributing players. Therefore, all the midfielder, who controls the tempo of the game, and the defenders both proved undervalued in contrast to overvalued attacking forward players. The proposed approach is framed in such a way as to look at the game impact not only in terms of goals or assists but also through equally important non-goal impact. Relying merely on machine learning would result in goal-biased or position-biased ratings based on only goal impact based measurement, where defenders or midfielders appeared to be far weaker contributors in comparison to aggressive attacking forward players. The proposed work uses a hybrid feature selection approach where feature selection has been done in two phases. In the first phase, this has been done purely knowledge-driven by the experts, and the features have been categorized into two groups: Non-goal impact and Goal impact.

In the second phase, feature selection techniques used are Lasso, Mutual information, random forest, and RFECV. The six selected sets of features have been presented to three models, i.e., Random Forest, Gradient Boost, and Linear regression, for the same three positional player categories, i.e., defenders, midfielders, and attacking forwards. Furthermore, the best-performing model, i.e., Random Forest, providing an average accuracy of 93.45% (average accuracy of Non-goal impact and goal impact for all three playing positions) , has been employed to estimate two types of ratings, i.e. non-goal impact based player performance ratings (NGI ratings) and goal impact based player performance ratings (GI ratings). Final player performance ratings have been obtained by the fusion of both the estimated NGI ratings and GI ratings. These estimated player performance ratings have been used to select the team line-up for the upcoming game. The work can be extended by including player injury and workload management for long-term vision. The proposed approach introduces a dual-rating player evaluation that supports more objective selection of team line-up to assist coaches in soccer sports. Selecting the appropriate unbiased team line-up based on estimating player performance is an indispensable component for managing a soccer team.

Authorship contribution: All the authors have made nearly equal contributions to the research work presented in the manuscript and have approved the final version after detailed review. The authors are accountable for the originality, accuracy, and integrity of the work.

Funding: We hereby declare that no funding has been provided by any agency.

Conflict of interest: All the authors declare that there are no conflicts of interest regarding the authorship, research, or publication of this manuscript.

Declaration: We hereby declare the originality of the data, and it has not been sent or published anywhere.

Similarity Index: The authors hereby state that there is no similarity index in the abstract and conclusion, while overall similarity is less than 10%, with individual source contribution is 2% or less.

References

[1] Madan, K., Taneja, K., Taneja, H., Intelligent Decision Making Technique for Handling Game Tactics in Soccer Sports. International Journal of Mathematical, Engineering and Management Sciences,10(5), 1604-1620 (2025). https://doi.org/10.33889/IJMEMS.2025.10.5.076

[2] Rico-González, M., Pino-Ortega, J., Méndez, A., Clemente, F., Baca, A., Machine learning application in soccer: a systematic review. Biology of sport,40(1), 249-263 (2023). https://doi.org/10.5114/biolsport.2023.112970

[3] Madan, K., Taneja, K., Taneja, H., Intelligent Strategic Planning Method based Algorithm (ISPMA) for Estimation of Soccer Sports Match Outcome. International Journal of Mathematical, Engineering and Management Sciences,10(1), 249-264 (2025). DOI:10.33889/IJMEMS.2025.10.1.014

[4] Arndt, C., Brefeld, U., Predicting the future performance of soccer players. Statistical Analysis and Data Mining: The ASA Data Science Journal, 9(5), 373-382 (2016). https://doi.org/10.1002/sam.11321

[5] Malikov, D., Kim, J., Beyond xG: A Dual Prediction Model for Analyzing Player Performance Through Expected and Actual Goals in European Soccer Leagues. Applied Sciences, 14(22), 10390 (2024). https://doi.org/10.3390/app142210390

[6] Kulakou, S., Ragab, N., Midoglu, C., Boeker, M., Johansen, D., Riegler, M. A., Halvorsen, P., Exploration of different time series models for soccer athlete performance prediction. Engineering Proceedings,18(1), 37 (2022). https://doi.org/10.3390/engproc2022018037

[7] Wiik, T., Johansen, H. D., Pettersen, S. A., Baptista, I., Kupka, T., Johansen, D., Halvorsen, P., Predicting peek readiness-to-train of soccer players using long short-term memory recurrent neural networks. In 2019 International Conference on Content-Based Multimedia Indexing (CBMI)(pp. 1-6). (2019) IEEE.

© 2026 Author(s). This article is published under the CC-BY license at http://jpr.vyomhansjournals.com.

[8] Bergkamp, T. L., Frencken, W. G., Niessen, A. S. M., Meijer, R. R., den Hartigh, R. J., How soccer scouts identify talented players. European Journal of Sport Science,22(7), 994-1004 (2022). https://doi.org/10.1080/17461391.2021.1916081

[9] Dijkhuis, T. B., Kempe, M., Lemmink, K. A., Early prediction of physical performance in elite soccer matches—a machine learning approach to support substitutions.Entropy,23(8), 952 (2021). https://doi.org/10.3390/e23080952

[10] Carpita, M., Ciavolino, E., Pasca, P., Players' role-based performance composite indicators of soccer teams: A statistical perspective. Social Indicators Research,156(2), 815-830 (2021). https://doi.org/10.1007/s11205-020-02323-w

[11] Dick, U., Tavakol, M., Brefeld, U., Rating player actions in soccer. Frontiers in Sports and Active Living,3, 682986 (2021). https://doi.org/10.3389/fspor.2021.682986

[12] Stafylidis, A., Mandroukas, A., Michailidis, Y., Vardakis, L., Metaxas, I., Kyranoudis, A. E., Metaxas, T. I., Key performance indicators predictive of success in soccer: A comprehensive analysis of the Greek soccer league. Journal of Functional Morphology and Kinesiology,9(2), 107 (2024). https://doi.org/10.3390/jfmk9020107

[13] Pantzalis, V. C., Tjortjis, C., Sports analytics for football league table and player performance prediction. In 2020 11th International Conference on Information, Intelligence, Systems and Applications (IISA(pp. 1-8) (2020). IEEE. DOI: https://doi.org/10.1109/IISA50023.2020.9284352

[14] Kaggle (2021).[dataset file]. FIFA 22 Complete player dataset. RetrievedJanuary 26, 2025, from https://www.kaggle.com/datasets/stefanoleone992/fifa-22-complete-player-dataset

[15] Nogueira, S., Sechidis, K., Brown, G., On the stability of feature selection algorithms. Journal of Machine Learning Research,18(174), 1-54 (2018). https://www.jmlr.org/papers/v18/17-514.html

© 2026 Author(s). This article is published under the CC-BY license at http://jpr.vyomhansjournals.com.

Copyright and License

[© 2026 **Karun Madan, et al.**] This is an Open Access article published in "Graduate Journal of Interdisciplinary Research, Reports & Reviews" (Grad.J.InteR^3), an **Open Access Journal** by Vyom Hans Publications. It is published with a Creative Commons Attribution - CC-BY4.0 International License. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

How to Cite?

Karun Madan, et al. (2026). Goal and Non-goal Impact Oriented Machine Learning Based Player Performance Estimation in Soccer Sports. *Graduate Journal of Interdisciplinary Research, Reports and Reviews*, 4(01), 1-11. Retrieved from https://jpr.vyomhansjournals.com/index.php/gjir/article/view/62

© 2026 Author(s). This article is published under the CC-BY license at http://jpr.vyomhansjournals.com.