Global Medical and Health Communication

Online submission: https://ejournal.unisba.ac.id/index.php/gmhc
DOI: https://doi.org/10.29313/gmhc.v12i1.12788

GMHC. 2024;12(1):1–11
pISSN 2301-9123 | eISSN 2460-5441

RESEARCH ARTICLE

AI-SPOT: a Novel Artificial Intelligence-enabled Sport Optimization Tracker to Enhance Performance and Prevent Injury in Elite Footballers

Aaron Chen Angus,[1] Dinesh Sirisena[2]
[1]School of Sports, Health and Leisure, Republic Polytechnic, Singapore,
[2]Department of Sports and Exercise Medicine, Khoo Teck Puat Hospital, Singapore

Abstract

This study introduces AI-SPOT, a novel artificial intelligence tool for optimizing performance and preventing injuries in elite footballers. Data were collected from four Singapore Premiere League clubs and the National Team, encompassing 68 male footballers over two seasons (2021–2022). The comprehensive dataset included diverse metrics, injury records, and automated live match data from established databases. AI-SPOT employs Python's scikit-learn for predictive analytics, using techniques like logistic regression and XGBoost, and was further developed with TensorFlow. Its effectiveness in injury prediction and performance assessment was validated with extensive local and international data sources. The system's potential for broader sports applications was underscored by user experience assessments, indicating a significant shift towards AI-driven strategies in sports management. Despite its reliance on high-quality, sport-specific data, AI-SPOT's adaptability highlights its role as a transformative tool in sports analytics, paving the way for advanced, data-driven approaches in sports management and strategy formulation.

Keywords: Artificial intelligence, data-driven decision making, machine learning, sport medicine

Introduction

Elite football clubs channel substantial resources into player recruitment, training, and sustenance. While detrimental to on-field performance, player injuries also pose considerable financial implications. Contemporary football sees a burgeoning inclination towards adopting data-driven decision-making mechanisms such as player performance enhancement and injury prevention. Integrating artificial intelligence (AI) offers promising prospects in predicting player injuries and ascertaining their value.

In elite football echelons, the frequency of injuries remains alarmingly elevated. Muscle strains, notably hamstring strains, are the most commonplace, succeeded by ligament sprains and injuries to the meniscus or cartilage.[1,2] Training load, about both volume and intensity, is a significant determinant for these injuries.[3,4] The aftermath of such injuries transcends mere player health. The absence of crucial players may culminate in match losses, which have a cascading effect on league standings and can instigate substantial economic setbacks.[5] Chronic injuries can exacerbate player downtime, leading to a dip in match-day earnings, a drop in player asset value, and escalated medical expenditures.[6] The consequential losses impact diverse aspects, from league standings and ticket revenue to sponsorships and player market evaluations.[5,7]

Historically, elite football clubs' determinations regarding player recruitment and participation in pivotal matches hinged on the understanding of coaches and scouts. Yet, data analytics proffers a more grounded viewpoint. Tools designed to gauge player value, epitomized by the expected goals (xG) metric, have garnered extensive traction in contemporary times.[8,9] The xG paradigm gauges shot quality predicated on various determinants, affording coaches a lucid understanding of player output.[10,11] The infusion of such data-centric instruments enables clubs to adopt enlightened choices, potentially amplifying football standards at both club and national tiers.[12,13]

The realm of sports analytics has witnessed rapid strides in the potentialities of AI. Cutting-

Received: 15 October 2023; Revised: 21 November 2023; Accepted: 22 December 2023; Published: 30 April 2024

Correspondence: Aaron Chen Angus. School of Sports, Health and Leisure, Republic Polytechnic. 9 Woodlands Avenue 9, Singapore 738964. E-mail: aaron_chen_angus@rp.edu.sg

edge algorithms now hold the prowess to forecast a player's imminent performance trajectory and their susceptibility to injuries.[14] Training these algorithms on expansive datasets augments their precision.[15] Envisioning a consolidated framework, spearheaded by AI, materializes as the potential panacea for clubs striving to hone player output while concurrently curtailing injury risks.[16] Contemporary algorithms cater to athlete surveillance and prognosticate injury vulnerabilities by tracking players' physical exertions juxtaposed against benchmarked metrics.[17] Furthermore, the pervasive xG model in football analytics evaluates shot quality drawing on diverse parameters, encompassing shot positioning, the anatomical region utilized, and the nature of the assist.[8] The present discourse seeks to refine these algorithms by capitalizing on an exhaustive dataset representing elite football talent.

This study aims to develop AI-SPOT, an advanced artificial intelligence tool designed for optimizing performance and injury prevention in elite footballers. It focuses on enhancing injury prediction algorithms and evaluating athlete performance using extensive data from top-tier football players. The objectives include implementing machine learning techniques for efficient athlete load management, resource allocation, and real-time tactical decision-making. The research intends to contribute to sports medicine, sports science, and performance analytics, showcasing the efficacy of AI in professional sports.

Methods

The study cohort comprised 68 male elite footballers systematically selected from the Singapore Premiere League (SPL) and the Singapore National Football Team. These subjects actively participated in the SPL's two consecutive competitive seasons (2021–2022). The inclusion criteria mandated active involvement in the team lineup for the respective seasons, ensuring all participants met the rigorous health and fitness standards for elite-level national football. Participant demographics included an age range of 18 to 33 years, with a mean age of 25.4 years. Bodyweight varied between 75 and 98 kilograms (mean: 87.5 kg), and height ranged from 175 cm to 190 cm, averaging 178 cm. These physical attributes and performance metrics were categorized and analyzed in alignment with established industry practices.[18] The selection process was stringent, focusing on athletes who represented the pinnacle of their profession in terms of skill and performance and met the comprehensive physical and health criteria characteristic of top-tier footballers. This rigorous selection ensured the reliability and relevance of the data collected for this study.

A comprehensive dataset was curated for each subject throughout the two SPL seasons. The parameters included various aspects of training load data,[19] summarised in Table 1.

Injury data was systematically collected from the sports medicine departments of the respective football clubs, encompassing the following data summarised in Table 2.

To maintain confidentiality, measures were implemented to anonymize participants, ensuring the absence of any identifiable information. In line with ethical research practices, subjects participated voluntarily, retaining the right to withdraw at any point without repercussions. Additionally, participants provided informed consent by standard ethical research procedures on human subjects.[20,21]

The research methodology, including the informed consent form, received approval from both the Institutional Review Board of Republic Polytechnic and the Football Association of Singapore's Medical Board (ethics approval reference code: HSR-SHL-F-2021-021).[20,21]

Performance analysis and player value data were primarily sourced from the InStat Analytics Database, Wyscout Database (courtesy of the Football Association of Singapore), and a custom Python-based web scraper extracting public match and player data. The exported XML file from the InStat platform needed minimal preprocessing, primarily structural adjustments.[22]

While data was readily available, it was evident that football clubs varied in their interpretation of the Wyscout and HUDL InStat datasets. Established methodologies typically involve video analytics, subsequently validated via the InStat platform.[23] The overarching goal of this research was to harness machine learning for faster, more efficient team play analyses, complementing conventional video analytics.

Advanced Python libraries facilitated extensive data wrangling of football match data. Cross-drill analyses were conducted during the exploratory data analysis (EDA) phase, harnessing

Table 1 Athlete Load Data Fields

Field No.	Field Name	Data Type
1	Date of record	Date
2	Day type (match, training, rest)	Nvarchar(50)
3	Athlete ID (masked name)	Nvarchar(255)
4	Position (goalkeeper, defender, midfielder, forward)	Nvarchar(50)
5	Gender	Char(1)
6	Date of birth	Date
7	Nationality	Nvarchar(100)
8	Ethnicity	Nvarchar(100)
9	Height	Decimal(5,2)
10	Weight	Decimal(5,2)
11	Body fat %	Decimal(5,2)
12	Lean mass	Decimal(7,2)
13	Basal metabolic rate	Int
14	Resting heart rate	Int
15	Maximum heart rate	Int
16	Average heart rate	Int
17	VO_2max	Decimal(6,2)
18	Blood lactate	Decimal(5,2)
19	Energy intake	Decimal(8,2)
20	Energy expenditure	Decimal(8,2)
21	Energy deficit	Decimal(8,2)
22	Hydration	Decimal(5,2)
23	Maximum speed	Decimal(5,2)
24	Average speed	Decimal(5,2)
25	Total distance	Decimal(8,2)
26	Vertical jump test score	Decimal(5,2)
27	Yo-yo test score	Decimal(5,2)
28	Sit and reach the test score	Decimal(5,2)
29	Perceived tissue damage (soreness)	Nvarchar(255)
30	Perceived effort (RPE)	Decimal(4,2)

visualizations to discern underlying patterns, trends, and potential outliers. Such outliers, once verified with sport science experts and deemed incongruent, were excluded, ensuring a more robust machine learning model.[24]

We are utilizing Python libraries, including sci-kit-learn, and the model-building phase encompasses feature reduction and model selection.

Feature reduction, a vital machine learning process, entails optimizing the number of features for improved learning outcomes. It can be achieved by removing redundant or irrelevant features in the early EDA stages or during modeling for better generalization and accuracy.[25] Principal component analysis (PCA) is vital for reducing dimensions in machine learning, especially with complex datasets like our project. PCA is used in this study to transform data into uncorrelated variables, simplifying the dataset and maintaining key information, thus aiding in efficient analysis and enhanced interpretability of intricate data.[26]

Classification models, either simple or complex, were employed to predict potential injuries. While simpler models, like decision trees, are resource-efficient and beneficial in certain situations, complex models like the extreme gradient boosting tree model offer nuanced parameter tuning for enhanced accuracy in high-dimensional datasets.[27] Model efficacy was evaluated using key machine learning

Table 2 Injury Record

Data Point	Data Field Type
Date of record	Date
Athlete ID (masked name)	Nvarchar(100)
Type of injury reported (based on list) □ Immunodeficiency (flu symptoms) □ Fever □ Tendon/muscle stiffness □ Tendon/muscle strain □ Tendon/muscle tear □ Ankle sprain □ Knee injury □ Quad, hamstring, groin strain □ Hip pointers □ Shoulder dislocation □ Acromioclavicular sprain □ Wrist and hand injury □ Meniscal damage □ Bone fracture □ Muscle atrophy □ Cartilage degeneration □ Concussion □ Others: _____	Nvarchar(100)
Number of days of medical certificate	Int
Missed match (date of match)	Date

metrics: accuracy score, receiving operator characteristics-area under the curve (ROC-AUC), precision, recall, and f1-score.[28,29]

Results

This study undertook an intensive exploration of multi-dimensional data derived from a league comprising 380 matches, averaging 1,800 events per match, encompassing 352 distinct action types. Specific actions, such as failed shot saves, goal kicks, and penalty kicks were discerned to bear a heightened likelihood of influencing game outcomes. Yet, their inherent randomness limits their tactical exploitability. As the focus shifted towards tactical actions and amenability to manipulation, constraints imposed by machine learning model assumptions arose, especially concerning logistic regression, which mandates feature independence and a degree of linearity.

A pivotal step in the data wrangling phase, feature engineering was employed to ensure the dataset aptly supported our project objectives. Important relationships derived from the raw data include a high correlation between passes and shots, an inverse moderate correlation between duels and passes, and a significant correlation between shots and game outcomes. We recognized a clear differentiation in the number of shots and passes between stronger and weaker teams. Additionally, midfielders emerged as the most frequently substituted player role. We categorized team actions into three sections: defending pitch, midfield, and attacking pitch, anticipating these divisions to offer insights into team strategies and player performance.

The initial model design was to craft distinct models for various teams, given the unique characteristics within each team's data. A challenge was ensuring the models' robustness, especially considering each team only had a sample size of 38 games per season. The data presented discrepancies in accuracy rates, with some models reflecting high accuracy for stronger or weaker teams, likely influenced by imbalanced data.

In a 7-fold cross-validated study on stratified data, a top-performing team in the 2022 SPL

Season resonated best with the Gaussian Naïve Bayes classifier, while a lower-tier team aligned more with the random forest classifier. These models, however, showed considerable variance when trained and tested over 10,000 iterations due to the limited sample size.

To enhance the model, data augmentation techniques like the synthetic minority oversampling technique (SMOTE), adaptive synthetic sampling (ADASYN), and deep learning neural networks were employed. However, the conventional augmentation methods, such as SMOTE and ADASYN, were ineffective. The conditional transformative generative adversarial networks (CTGANs) model, a variant of GANs tailored for tabular data, was introduced. Despite some instances of optimism, the generated synthetic data often needed to mirror real-world data, leading to potential inaccuracies.

PCA was leveraged to decipher the high-dimensional data. The initial PCA showcased that actions predominantly in the attacking pitch area play a pivotal role in match outcomes. Post-feature selection, a more refined dataset was

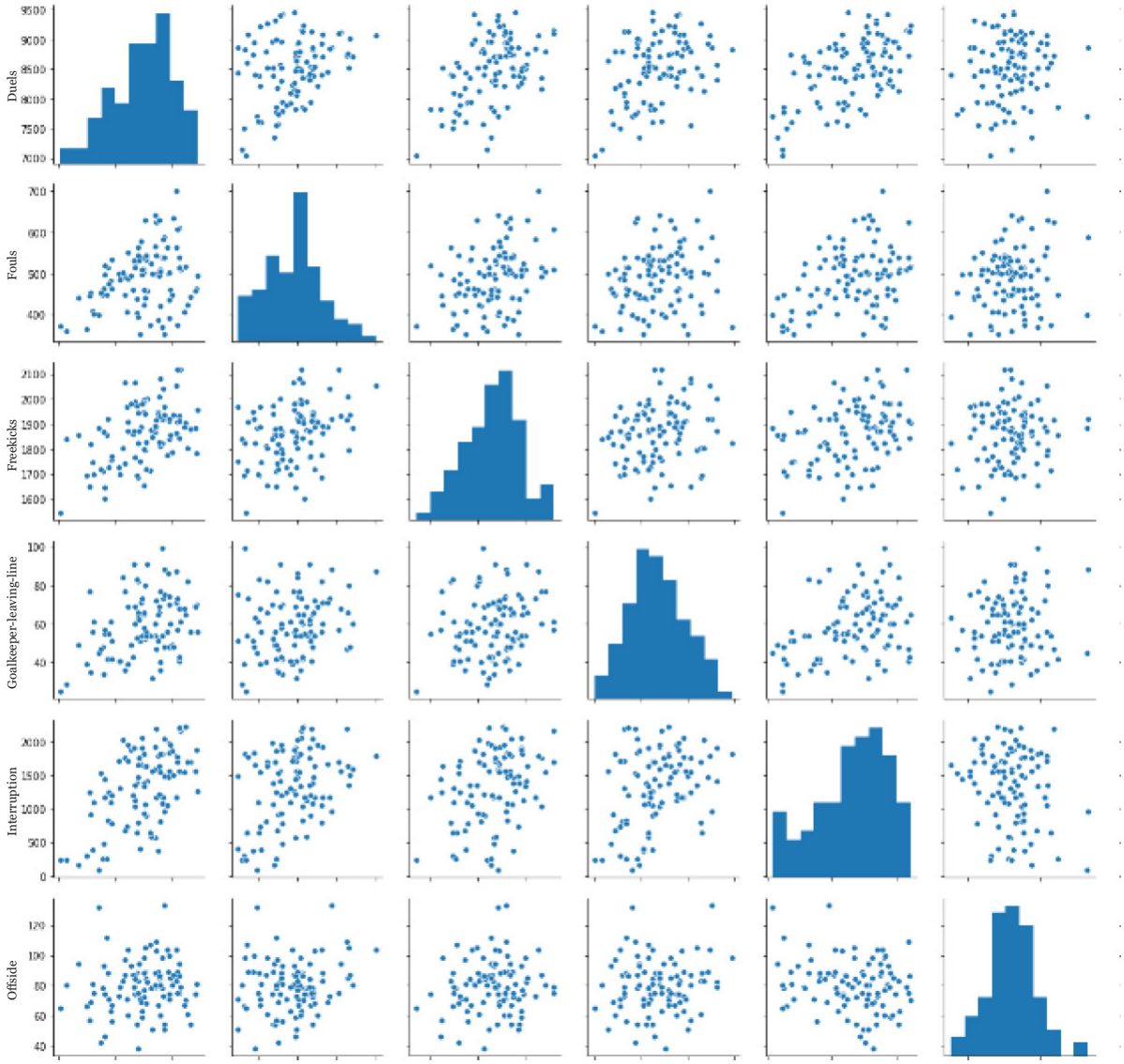

Figure 1 Pairs Plot Sample from PCA to Visualise Feature Relationships

Note: The pairs plot from PCA allows the identification of critical features for dimensionality reduction based on the comprehensive dataset used in this study by displaying correlations between feature pairs in machine learning model building

employed for a second PCA, where the primary component accounted for over 40% of the data variance, emphasizing the significance of midfield action and passing networks (Figure 1).

The iterative approach in this study led to a model that emphasizes frequently occurring and impactful actions. Simplifying the dataset is anticipated to yield more robust results, especially considering the limited sample size. The final design targets three pivotal actions across the three pitch areas: passes, duels, and shots. League-wide data will be employed for model training instead of individual team data.

Subsequent feature selection was streamlined by leveraging model fitting to pinpoint pivotal features. The logistic regression model, with its commendable accuracy of 83%, provided a reliable metric for feature importance. The gradient boosting classifier displayed a marginally superior accuracy, suggesting its potential to yield even more refined outcomes with hyperparameter tuning, and this was subsequently adopted for the model's development. The performance metrics for the various (Table 3 and Table 4).

The Shapley additive explanation (SHAP) method is rooted in cooperative game theory and has been adapted to measure feature importance in machine learning. Distinguishing from other models, SHAP offers localized feature importance evaluations, making it more adaptable to dynamic situations such as individual football matches where data from one game might vary considerably.

Our analyses revealed insightful patterns. In a swamp plot generated from logistic regression using the SHAP explanation, the most significant feature was 'team1 accurate shots'. A higher value for this feature correlates with a more significant positive impact on match outcomes. Contrastingly, the 'team1 pass non-accurate in the attack pitch' feature, although holding a mix of high values, demonstrated varied effects on the game, underscoring the unique nature of each football match. Comparative analyses of various models (like random forest and XGB classifiers) using SHAP yielded consistent results on feature importance. However, the magnitude of SHAP values varied, with tree models typically showing smaller values. Among the models evaluated, logistic regression emerged superior due to its rapid training speed, high predictive accuracy (74–78%), and substantial SHAP values. The XGB classifier, although slightly more accurate in prediction, was more time-consuming and exhibited minimal SHAP values. Given these findings, this study designated the logistic regression model for this system, but other models remain relevant for future data comparisons.

For player evaluation, leveraging the SHAP values from the logistic regression, we calculated the feature importance for individual games and extended this to assess player contributions. We sought to represent intangible match aspects like team morale, synergy, and skill through statistical evaluations. By leveraging SHAP values, we distilled these concepts into measurable actions on the field. We introduced two new metrics for player evaluation: 'player value', derived from the product of action count and its corresponding SHAP value, and a normalized metric accounting

Table 3 Performance Metrics for Classifiers Used for Model Design

Classifier	Accuracy	Log Loss	Recall	ROC_AUC
LogisticRegression	80.246914	0.410564	0.792952	0.891807
KNeighborsClassifier	62.55144	3.724202	0.572687	0.661635
SVC	53.292181	0.65454	0	0.616978
DecisionTreeClassifier	72.222222	9.594105	0.722467	0.722237
RandomForestClassifier	81.069959	0.459312	0.753304	0.897428
XGBClassifier	83.950617	0.448677	0.837004	0.909889
AdaBoostClassifier	82.304527	0.666443	0.792952	0.906077
GradientBoostingClassifier	82.098765	0.397874	0.814978	0.901944
GaussianNB	75.925926	1.443781	0.784141	0.83555
LinearDiscriminantAnalysis	84.156379	0.372274	0.845815	0.913476
QuadraticDiscriminantAnalysis	61.522634	10.989866	0.77533	0.633664
CatBoostClassifier	81.481481	0.38089	0.819383	0.910074

Table 4 Top Features in Feature Engineering

Coefficient	Features
0.727745	Shot Shot Opportunity Accurate
0.384252	Home Ground
0.35633	Pass Cross Assist Accurate
0.325931	Offside
0.249253	Pass Simple pass Assist Accurate
0.213237	Shot Shot Counter attack Opportunity Accurate
0.19711	Free Kick Throw in Not accurate
0.194009	Pass Simple pass Interception Not accurate
0.186302	Pass Smart pass Assist Accurate
0.183734	Duel Air duel Counter attack Accurate
−0.167229	Others on the ball Acceleration Not accurate
−0.180898	Duel Ground attacking duel Interception Accurate
−0.182379	Duel Ground attacking duel Counter attack Not accurate
−0.215683	Pass Cross Key pass Accurate
−0.24314	Pass Cross Accurate
−0.257963	Pass Simple pass Key pass Accurate
−0.298885	Pass Simple pass Dangerous ball lost Not accurate
−0.347183	Free Kick Free kick cross Not accurate
−0.370614	Save attempt Reflexes Accurate
−1.438092	Save attempt Reflexes Not accurate

for play duration, obtained by dividing the player value by the total minutes played on the pitch (Figure 2 and Figure 3).

The model was deployed to a client-based functional application developed using TensorFlow and hosted on a cloud for beta testing with the participating football clubs. This product is named AI-SPOT, which is short for Artificial Intelligence-enabled Sports Performance and Optimisation Tracker. Some sample screens in this paper use English Premiere League data run on this system to demonstrate and generate predictive outcomes (as the Singapore Premiere League data is classified and provided only for model building). A sample workflow in the system is shown in Figure 4.

Discussion

This study leveraged advanced machine learning to enhance football team management, specifically XGBoost and logistic regression classifiers. This approach addressed critical aspects such as athlete load, injury prediction, match performance, and player valuation. Demonstrating robustness and accuracy, the model's effectiveness was affirmed using diverse datasets, including those from the Singaporean football teams and the extensive English Premier League data, thereby showcasing its potential for broad application in sports analytics.

Utilizing TensorFlow, AI-SPOT was developed into a functional platform, prioritizing robust data security and an intuitive user interface refined through beta tests with various football teams. This development process ensured the platform's adaptability to the specific needs of team managers and coaches and accommodated the dynamic nature of football data.

AI-SPOT's success in football demonstrates its potential for broader applications in sports, marking a shift from traditional methods to data-driven strategies. Utilizing AI for predictive analytics, this approach could significantly enhance the quality and performance across various sports disciplines, not limited to football. The adaptability of the system's algorithms allows for recalibration to suit both team-based and individual sports, indicating a new era of technology-driven sports performance, injury prevention, and strategic decision-making.

In this study, AI-SPOT has demonstrated

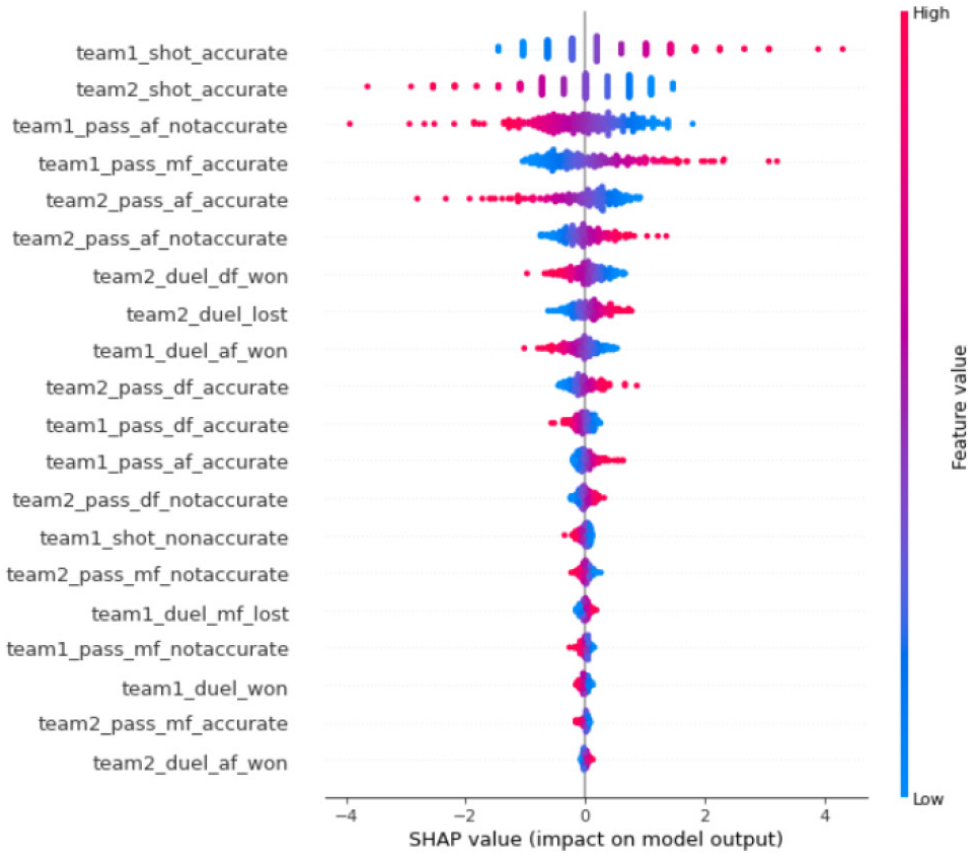

Figure 2 Swamp Plot Sample Using Logistic Regression for SHAP Values

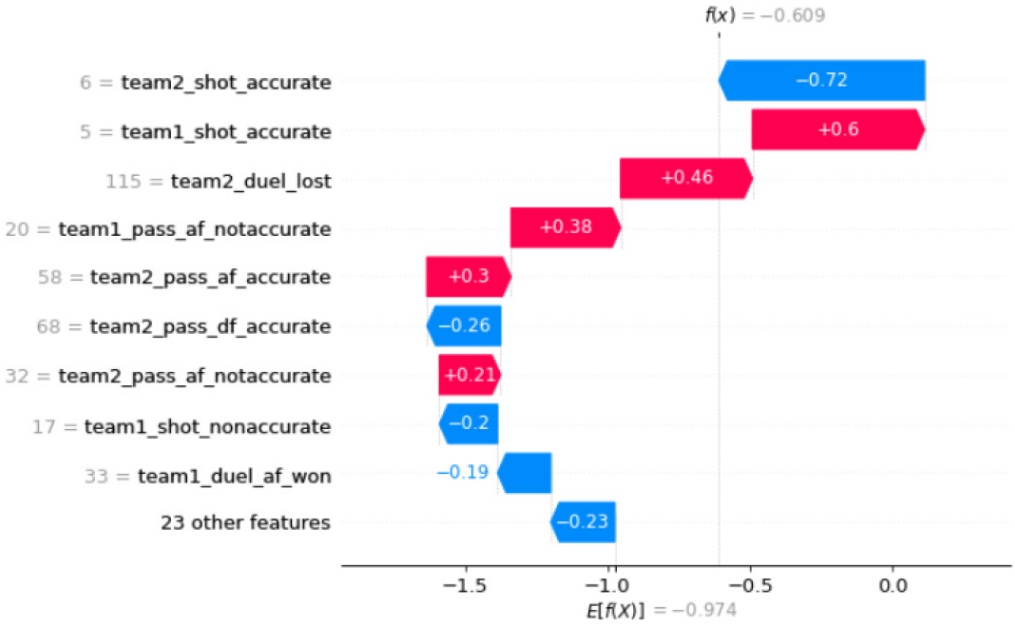

Figure 3 Waterfall Plot on Sample Match Trained on Logistic Regression

Note: The plot shows a function f(x) on the prediction probability of the game. At a negative value, the model predicts the playing team is unlikely to win the game. The blue and red ribbons represent features causing negative and positive impacts on the game outcome

Figure 4 Sample Workflow of the AI-SPOT System Using EPL Data

Note: SPOT system's workflow using EPL data showcases the systematic visualization process for predictive analytics, which aids in evaluating team performance indicators for tactical decision-making

significant advancements in sports analytics and injury prevention, yet it also encounters limitations. Compared to previous methods, AI-SPOT's advanced machine learning algorithms integration offers more accurate injury prediction and performance assessment, addressing gaps identified in earlier research. However, its reliance on extensive, high-quality data from professional leagues may limit applicability in lower-tier or amateur sports settings. Additionally, while AI-SPOT shows promise in enhancing decision-making, translating its insights into practical coaching strategies requires further exploration. These considerations suggest areas for ongoing research and development, emphasizing the need for adaptable, versatile tools in sports analytics.

Conclusions

This study harnessed advanced machine learning techniques to optimize football team management by integrating athlete load, injury prediction, match performance, and player valuation using the XGBoost and Logistic Regression classifiers, resulting in a robust and accurate model. The successful transition from model development to real-world application culminated in AI-SPOT, a functional platform powered by TensorFlow that holds promise for broader applications in sports beyond football, emphasizing the potential of data-driven decision-making in sports management and performance optimization.

Conflict of Interest

None declared.

Acknowledgment

The authors would like to express their appreciation for the support from the Football Association of Singapore (FAS) as well as Singapore Premiere League Teams who have actively supported this study, Hougang United, Lion City Sailors, Geylang United and Tampines Rovers football clubs.

References

1. López-Valenciano A, Ruiz-Pérez I, Garcia-Gómez A, Vera-Garcia FJ, De Ste Croix M, Myer GD, Ayala F. Epidemiology of injuries in professional football: a systematic review and meta-analysis. Br J Sports Med. 2020;54(12):711–8.
2. Della Villa F, Buckthorpe M, Grassi A, Nabiuzzi A, Tosarelli F, Zaffagnini S, et al. Systematic video analysis of ACL injuries in professional male football (soccer): injury mechanisms, situational patterns and biomechanics study on 134 consecutive cases. Br J Sports Med. 2020;54(24):1423–32.
3. Gabbett TJ. The training-injury prevention paradox: should athletes be training smarter and harder? Br J Sports Med. 2016;50(5):273–80.
4. Windt J, Zumbo BD, Sporer B, MacDonald K, Gabbett TJ. Why do workload spikes cause injuries, and which athletes are at higher risk? Mediators and moderators in workload–injury investigations. Br J Sports Med. 2017;51(13):993–4.
5. Dellal A, Lago-Peñas C, Rey E, Chamari K, Orhant E. The effects of a congested fixture period on physical performance, technical activity and injury rate during matches in a professional soccer team. Br J Sports Med. 2015;49(6):390–4.
6. Rössler R, Junge A, Chomiak J, Dvorak J, Faude O. Soccer injuries in players aged 7 to 12 years: a descriptive epidemiological study over 2 seasons. Am J Sports Med. 2016;44(2):309–17.
7. Lee EC, Fragala MS, Kavouras SA, Queen RM, Pryor JL, Casa DJ. Biomarkers in sports and exercise: tracking health, performance, and recovery in athletes. J Strength Cond Res. 2017;31(10):2920–37.
8. Liu H, Hopkins W, Gómez MA, Molinuevo JS. Inter-operator reliability of live football match statistics from OPTA Sportsdata. Int J Perf Anal Sport. 2020;13(3):803–21.
9. Jha D, Rauniyar A, Johansen HD, et al. Video analytics in elite soccer: a distributed computing perspective. Proc IEEE Sens Array Multichannel Signal Process Workshop. 2022;2022:221–5.
10. Fernandez-Navarro J, Fradua L, Zubillaga A, McRobert AP. Influence of contextual variables on styles of play in soccer. Int J Perf Anal Sport. 2018;18(3):423–36.
11. van Maarseveen MJJ, Oudejans RRD, Mann DL, Savelsbergh GJP. Perceptual-cognitive skill and the in situ performance of soccer players. Q J Exp Psychol (Hove). 2018;71(2):455–70.

12. Rein R, Memmert D. Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. Springerplus. 2016;5(1):1410.
13. Tenga A, Sigmundstad E. Characteristics of goal-scoring possessions in open play: comparing the top, in-between and bottom teams from professional soccer league. Int J Perf Anal Sport. 2011;11(3):545–52.
14. Bittencourt NFN, Meeuwisse WH, Mendonça LD, Nettel-Aguirre A, Ocarino JM, Fonseca ST. Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition—narrative review and new concept. Br J Sports Med. 2016;50(21):1309–14.
15. Gonçalves B, Coutinho D, Travassos B, Brito J, Figueiredo P. Match analysis of soccer refereeing using spatiotemporal data: a case study. Sensors (Basel). 2021;21(7):2541.
16. Sarmento H. From traditional scouting to a data-driven approach: soccer analytics in the era of big data. J Hum Sport Exerc. 2020;15(3):678–88.
17. Gabbett TJ. Debunking the myths about training load, injury and performance: empirical evidence, hot topics and recommendations for practitioners. Br J Sports Med. 2020;54(1):58–66.
18. Smith J. Elite football player characteristics and performance metrics: an analysis. Sports Med. 2020;50:1231–42.
19. Djaoui L, Haddad M, Chamari K, Dellal A. Monitoring training load and fatigue in soccer players with physiological markers. Physiol Behav. 2017;181:86–94.
20. Wiesing U, Parsa-Parsi R. The World Medical Association launches a Revision of the Declaration of Geneva. Bioethics. 2016;30(3):140.
21. Bjärsholm D, Gerrevall P, Linnér S, Peterson T, Schenker K. Ethical considerations in researchingsportandsocialentrepreneurship. Eur J Sport Sci. 2018;15(3):216–33.
22. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care . Lancet Oncol. 2019;20(5):e262–73.
23. Kubayi A. Analysis of goal scoring patterns in the 2018 FIFA World Cup. J Hum Kinet. 2020;71:205–10.
24. Berrar D. Data wrangling and exploratory data analysis in bioinformatics. Front Genet. 2020;11:512036.
25. Cai W, Yang Z, Wang Z, Wang Y. A new compound fault feature extraction method based on multipoint kurtosis and variational mode decomposition. Entropy (Basel). 2018;20(7):521.
26. Martin RK, Pareek A, Krych AJ, Maradit Kremers H, Engebretsen L. Machine learning in sports medicine: need for improvement. J ISAKOS. 2021;6(1):1–2.
27. Marynowicz J, Lango M, Horna D, Kikut K, Konefał M, Chmura P, et al. Within-subject principal component analysis of external training load and intensity measures in youth soccer training. J Strength Cond Res. 2023;37(12):2411–6.
28. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016 August 13–17; San Francisco, CA, USA. San Francisco, CA, USA: Association for Computing Machinery, Inc.; 2016. p. 785–94.
29. Rico-González M, Pino-Ortega J, Méndez A, Clemente FM, Baca A. Machine learning application in soccer: a systematic review. Biol Sport. 2023;40(1):249–63.