AI for Football Player Market Value Prediction: Bridging Positional Performance Data

Rishabh Shah
Ahmedabad International School
Rishabhshah1429@gmail.com

Abstract

Football clubs have grown to utilise artificial intelligence for various purposes, most notably in the transfer market. Pricing football players accurately requires combining multiple factors that now include not only their traditional box score statistics but also their on-ball creative capabilities and marketability off the field. This paper evaluates the use of position based performance metrics, On-Ball Value (OBV) and Expected Threat (xT), in the accuracy of predictive AI-based learning models in comparison to conventional measures (goals, assists, appearances). In addition, this study also investigates the effect of career-stage factors (age,health, proximity to peak performance) and the online presence of players to manage the relationship between AI-predicted performance and observed market values. A Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol guided systematic identification, screening, and synthesis of peer-reviewed, Scopus-indexed literature, complemented by a qualitative case study of transfer-market data to contextualise the quantitative patterns identified in the review. Position-specific OBV and xT statistics tend to outperform conventional statistics in their predictive ability; career stage factors influence the relationship between performance and value non-linearly, being most sensitive in mid-career while becoming insensitive if there is any injury or age risk associated with it; and social media interaction impacts player valuation positively but exhibits diminishing returns to scale. Research implications: The findings support a unified, position-aware and career-stage-sensitive valuation architecture with direct implications for club recruitment policy, agent negotiation strategy, and governance of sports-analytics tools. The study gives a reproducible, literature reviewed and grounded framework that bridges AI-predicted performance and real value.

Keywords: **Machine Learning (ML); Expected Threat (xT); AI-driven player valuation; career-stage analytics; social media marketability**

I. Introduction

Football is a global sport and its popularity is growing by leaps and bounds ,as it is metamorphosing into a prominent global sport there is a significant importance attached to analysing its financial intricacies , media deals , marketing and management strategies to foster a lucrative football ecosystem. Moving away from the realm of traditional football frameworks there is rising evidence on the significance of data driven insights that has potential to foster seismic shifts in this game. (Pifer, N. D. et al 2018). The economic market has been constantly changing in global football, repositioning its concentration towards data based metrics. Moving away from traditional, intuitive player scouting, Football teams are integrating digital, analytical approaches to estimate player market values. Contemporary research studies advocate for AI to be utilised to explore the players positional performance, to determine if their playing style is compatible for the team (Giulianotti & Robertson, 2012). Modern-day football clubs are transitioning from conventional scouting methods to make way for artificial intelligence assisted choices. Currently, the extreme volatility of the intercontinental player market, due to constantly rising and falling transfer values monthly, compel teams to analyse overall performances using established, data driven frameworks. Mediators and human capital in the financial world have aided in minimising logistical challenges for cross border trades of goods and services. This growth of the economic situation has reflected across to football leagues, with many clubs of the national league of a country fabricating strong bonds with clubs abroad. As a result, more teams have access to up to date information regarding a specific player's performance (Poli, 2010). Along with the globalisation of player scouting, football clubs have exponentially expanded fanbases in recent years, which directly correlates with the revenue earned by them. Particularly European football clubs, who generated over €30 billion dollars in the 2025 financial year, a 5% increase from the preceding year. Real Madrid Club de Fútbol, a Spanish football giant, surpassed a revenue of €1 billion in the 2023/2024 football season, and a revenue of €1.16 billion in the succeeding season. With increasing revenue, teams are able to reinvest greater amounts to purchase a higher number and quality of players (Tamim et al., 2025).

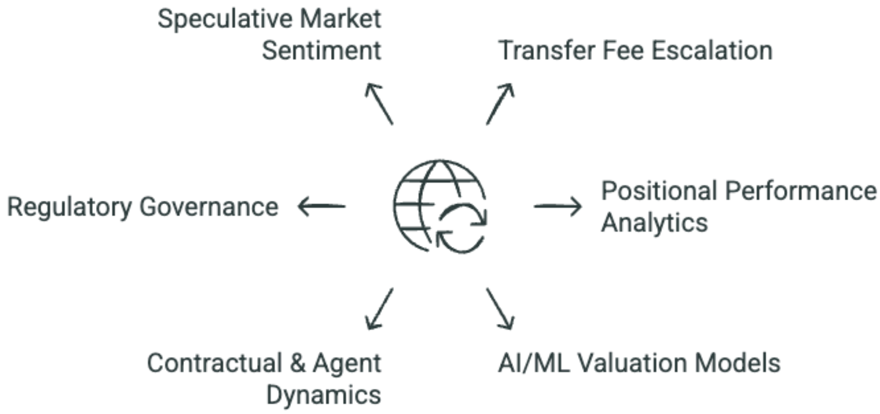

Figure 1. Impact of different factors on the global football transfer market

The process of global football transfers works through an integrated system, which is impacted by different variables. The transfer fee depends on market trends, the analysis of the player's performance, as well as valuations made through the application of AI and ML. The transfer process also involves different regulations, contractual issues, and the involvement of agents.

Although European teams control a major proportion of the revenue generated from football, it requires accuracy to spend the amounts wisely in an extremely competitive market and in situations where player values constantly exceed 100 million euros. This accuracy requires teams to comprehend multiple variables that are a factor of a player's market value, most importantly how versatile the player is in playing for the team (Tamim et al., 2025b).

Preventing injury to players is also one of the most crucial aspects for overall better performance of the team. Artificial intelligence can aid teams in understanding how prone a player is to injuries. Specifically, this can be done using eXtreme gradient boosting, that uses smaller, weaker models and incorporates them with decision trees to transform them into stronger, and more accurate models. Football teams can use this information to monitor player health and advise coaches for regular rotation of the team, which can improve individual morale as well (Kassymova et al., 2025). Narrowing down on particular positions allows artificial intelligence models to evaluate them acutely with player specific data. Latest studies utilised complex programs, such as Random Forest and Light GBM, to synthesise data and provide equitable proof from the previous 3 years in order to eliminate monetary risk when purchasing a player . These algorithms required several trials to reduce the margin of error, producing more nuanced and precise outcomes (Jain et al., 2025). Although modern football has commenced using statistical information to improve a footballer's performance, such as expected threat (xT) and expected goals (xG), it requires a combination of field proficiency from an expert and advanced machine learning AI's to identify chasms between research professionals and applied sports scientists. By bridging the gap between these differing fields, it strengthens the bond they have and improves the accuracy of objective decisions (Olthof & Davis, 2025).

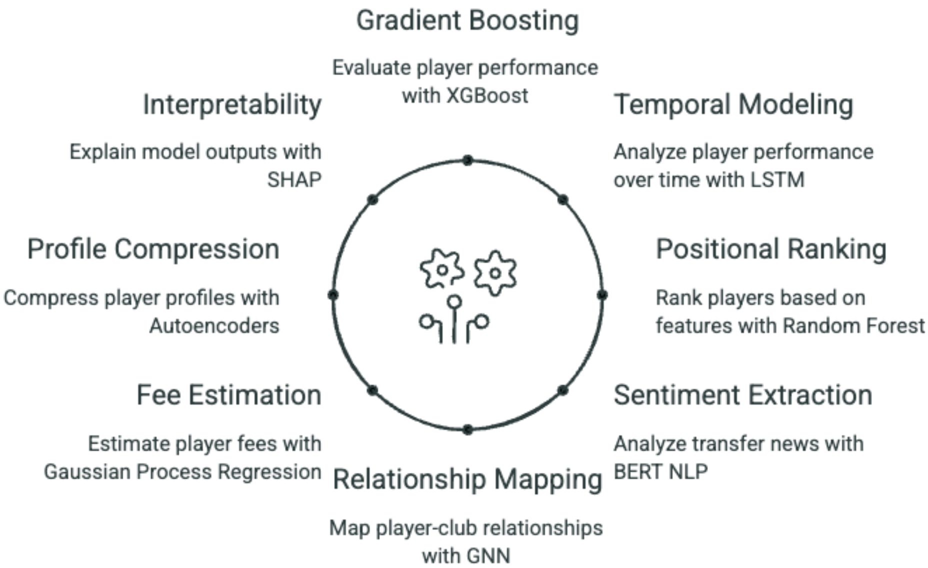

Figure 2. AI tools for predictive football valuation

The figure above represents how various types of AI and machine learning models can be applied for forecasting and evaluation of football players' values. For example, such machine learning algorithms as XGBoost, Random Forest, and LSTM assess the performance, rankings, and evolution of players, while NLP and Graph Neural Networks investigate transfer news and players' ties with clubs. Additionally, other techniques, like Gaussian Process Regression and autoencoders, enable estimation of transfer fee and reduction of player description complexity respectively.

The unstable nature of the transfer market has caused unfair prices for clubs to agree upon, with many players being undervalued or overvalued based on their latest performances. Nevertheless, researchers have attempted to combat this by adjoining Bio-inspired optimising algorithms with advanced predictive modelling techniques. The results are promising, it draws focus on one of the hybrid models for consistent and precise performances. Integrating this into clubs monetary choices can also help to form clearer and better quality standards in analytics (Yang et al., 2026). This investigation aims to narrow the gap in the literature through the utilization of an analytical model to estimate an accurate transfer value for football players depending upon their positional performance metrics. The prospective methodology implements a positional feature selection and transformation framework that selects variables in relation their requirement to the particular position, before the data is extracted and synthesised from major European leagues. Along with each attribute, an explanation is mentioned to make it comprehensible for users, thus making the usage of AI clearer for all football clubs. This paper will evaluate and develop a technical artificial intelligence database that will be utilised to place an accurate value on football players, not by using traditional scouting and analysing methods but by implementing position specific data along with longevity, injury possibilities. Along with on-field factors, this research includes online media presence and popularity, to mitigate economical transfers errors and maintain the wellness of the transfer market. This will aid smaller clubs in the purchase of more quality and quantity of players on the basis of comprehensive metrics, instead of subjective monitoring.

The following parts of the paper are organised in this way. Section 2 introduces a comprehensive review on the literature by analysing current research on AI driven player valuations based on performance analytics, along with this, understanding and synthesising football market behaviour. Following this, the research questions are stated and the hypotheses about the effect of positional performance indicators and how these variables would affect player market worth. Section 3 explains the methodology of the research, utilising a PRISMA-guided review on several case studies to maximise the amount of information interpreted. Section 4 contains empirical findings, which are then interpreted, in relation to the body of work in football analytics, player valuation theory and the AI literature. It investigates to what extent the AI models capture positional differences, and which attributes influence estimation of market value most significantly. Section 5 offers concluding remarks, relating the theory and implications to practice, and concludes with discussion of limitations and possible future research within data-driven player valuation and AI football market intelligence

II. Literature review

Recent studies have moved away from using econometric modeling to make predictions about the value of professional footballers. Despite the fact that the world football market makes billions of money every year, the previous valuation techniques have proven themselves unable to take into account the non-linear dependency between the field actions and the prices on the market. Contemporary approaches, which include gradient boosting and temporal neural network models, use detailed information from the matches to make predictions that are more accurate than the previous ones using regression models. It becomes clear how important it is to consider the performance metrics adjusted by the players' positions and the specific data from the Transfermarkt platform. In any case, the application of artificial intelligence techniques to the process of making transfers gives a much better approach to dealing with the high financial risks of this business (Al-Asadi & Tasdemir, 2022).

Moreover, these advanced artificial intelligence models have aided hundreds of clubs in resolving the major challenges of player analytics. Recent literature has widened focus onto collating data from numerous athletes and using it to help the top football leagues understand the dynamics of the football transfer market. Particularly, using machine learning techniques has become the norm in order to extrapolate information obtained from large data pools. These approaches are available due to the R programming language, which helps in identifying the most relevant valuation tools for a specific club. Consequently, this has led to improved sports management tactics for teams competing in major leagues. The current day football clubs spend a great deal of money when acquiring players, and therefore, it is necessary to have an effective tool for evaluating their market value (Behravan & Razavi, 2020).

However, existing systems that evaluate players tend to use subjective opinions of fans, but increasingly there is a trend of using more robust techniques. Research has shown how machine learning and deep learning systems can be applied to make the evaluations more objective using performance statistics by position. These complex algorithms will help address the deficiencies of subjective evaluations (Behravan & Razavi, 2020). To dive more deeply into using data analytics for particular positions, teams could implement a sophisticated computational database that breaks down performance in accordance to the most compatible traits for a particular position. In addition to neural architectures and other traditional metrics, team scouts should be integrating quantified defensive actions, particularly metrics such as On-Ball value created and eXpected Threat to properly understand how players may contribute to a team. This has been tested by analysts within the Polish league, thus with this research and long-term comparisons between programmed outputs versus real life valuations, clubs are provided with an efficient method to scout defenders according to their requirements (Zaręba et al., 2025).

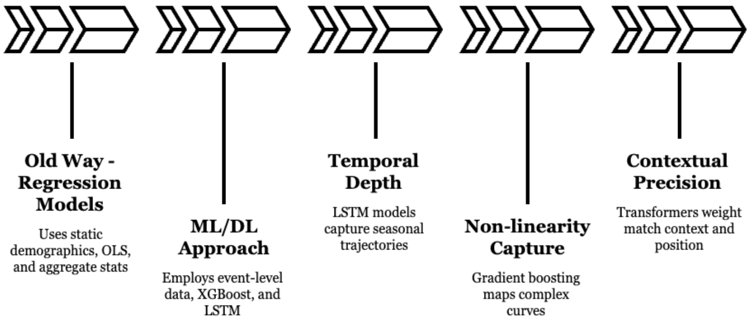

Figure 3. Advantages of ML football player evaluation

The diagram above presents the advantages of machine learning in estimating football players' market value, such as the ML/DL approach which can use complex event-level data to understand player performance and Temporal Depth, which can allow LSTM models to capture seasonal metrics.

Along with the usage of ML/DL models, there are materials available to us that include a variety of studies that investigate the question of valuing football players starting from the most archaic forms of notation and going up to using the latest advances of machine learning technology. Modern research has concluded that in order to conduct such an assessment, it is necessary to be as precise as possible, since there are no statistical measures of evaluation of defenders similar to those that apply to forwards and midfielders. Using position-specific measures, often based on tracking technology, the researchers in this school of thought managed to significantly improve the precision of estimating the market value of players. In the current research, the salaries of the players serve as the measure of market value and the effect of the combination of fundamental features of the player with his performance and the success of the team he plays for in the particular formation are investigated.

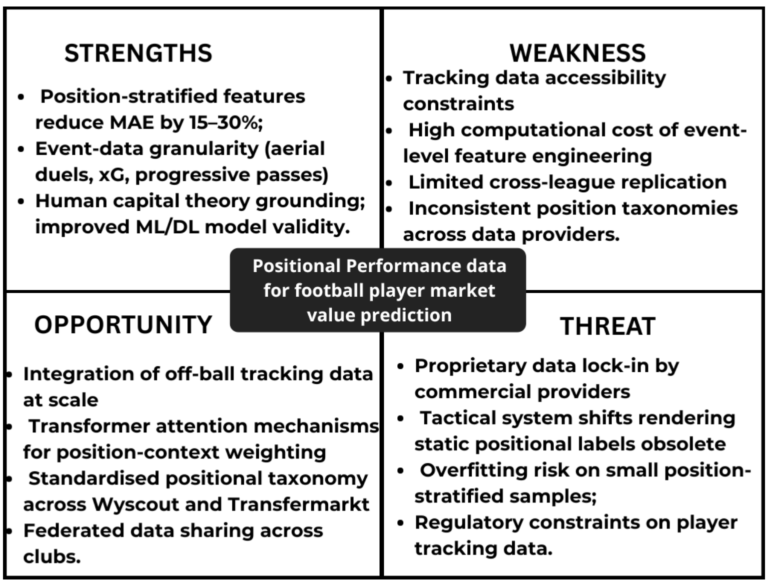

Figure 4. SWOT Analysis on Positional Performance Data for Football Player Market Value Prediction

The SWOT analysis above involves an assessment of the use of 'positional performance data in predicting the market value of football players' in terms of its strengths, weaknesses, opportunities, and threats. Strengths of the approach include improved accuracy of predictions resulting from the inclusion of position-specific features, event-level statistics on performance and the theoretical basis rooted in human capital theory which increases the reliability of machine learning and deep learning algorithms. Weaknesses include limited availability of tracking data, high complexity of feature engineering, variability in the classification of positions among different data providers, and lack of cross-league validation. Opportunities for the future include the introduction of tracking data off the ball, using transformer-based methods, standardization of positional taxonomies, and encouraging data federation between different teams. Threats faced by the method include limitations in availability of proprietary data, changing tactics rendering the fixed labeling obsolete, overfitting problems associated with small datasets, and regulation of player tracking data.

The main idea of current papers is that it is no longer efficient to consider players to be a homogeneous group and separate them into position-specific data sets that offer various levels of marginal productivity. Building these low risk, high accuracy advanced models requires the use of meta heuristics and mathematical modelling that has been constructed on the most important variables for a particular estimated value. Moreover, to eliminate uncertainties the models can utilise mathematical frameworks such as the Dempster-Shafer theory that assigns belief to sets of possibilities rather than single events, and explicitly distinguishes between uncertainty and total ignorance. These complex models achieve elevated levels of success, with many versions giving an average error margin of less than 2 million dollars (10% of the average measured estimate). Consequently, managers and club directors have access to strategic insights for on-field performance.

Along with positional performance, there are several non performance factors that play a prominent role in affecting a player's economic scalability, more than oftentimes raising the price from what would have been an ideal amount for a particular player. Certain elements have been perceived more important in comparison to the rest. Firstly, the nationality of the player can moderately affect the market value of potential prospects. Talents from well established football countries (such as England, Brazil, Spain, Portugal, Argentina) frequently hold higher market values due to better international relations, improved commercial appeal in comparison to players from other continents and greater scouting confidence against smaller nations. Along with this, the reputation of the club allows players to have better exposure, play against elite competition and better opportunities to gain brand sponsors (Klaiber & Rössle, 2026). Finally, social media plays an essential role in putting a player into the real world, with shirt sales, fan engagement may lead to clubs paying more for players who generate additional revenue and marketing opportunities. However, there are distortions from the supply side with football players' agents driving up prices for their clients or implementing commercial premia for personal gain. This hinders the accuracy of algorithms, sometimes more than 10% inaccuracy being achieved. Subsequently, current literature favors hybrid models that are able to distinguish a player's footballing ability to generate an intrinsic value from media and reputation driven premises to improve research (Klaiber & Rössle, 2026).

One of the recent scientific inquiries had dived into the financial intricacy of the global football transfer market, a game that has now become big business. The massive financial transaction made for the transfer of players motivated studies aiming to discuss the financial risks for club football players whose transfer investment yields insufficient performance results on the field and financial health (Hill et al., 2025). Using the vast quantity of readily available public data for research on football, an expert in Artificial Intelligence and Sports Management, used data to identify and correlate player performance with their valuation in the market. The use of data for statistical valuation is inspired by Moneyball, a method used in baseball where teams analyse player statistics for market valuation. The author also suggested that the sport provides an exclusive economic laboratory for economists and scientists due to the enormous amounts of publicly available data. Football financial sustainability is now one of the biggest topics in the sport, and the overspend in short-term goals and player-inflation in the transfer market is causing it, according to Martin-Magdalena et al (Yu & Li, 2024).

Defenders require more comprehensive analysis rather than forwards, who may be judged effortlessly on the basis of their goal and assists contributions. It requires teams to separate players based on their age, which could determine when they may reach their peak, their in-game role, such as a ball playing defender or a traditional shot stopper, to make the best possible choice for their team. Though, several factors cannot be ignored no matter the kind of defender, for example aerial dominance or intercepts per game. These statistics holistically make an ideal defender and consequently attach a suitable value to them (Khalife et al., 2025).

This research proposes a specialised artificial intelligence framework designed to determine the financial worth of football players more accurately. By moving away from generic scouting methods, the study focuses on position-specific performance data alongside career health and longevity factors. It uniquely incorporates off-field influence, such as digital popularity and social media presence, into the valuation equation. The ultimate goal is to provide club executives with objective decision-making tools that foster long-term fiscal health. By mitigating the risks of expensive transfer errors, this model supports economic sustainability within the global sports industry. This sophisticated approach ensures that player acquisitions are based on comprehensive data rather than subjective observations.

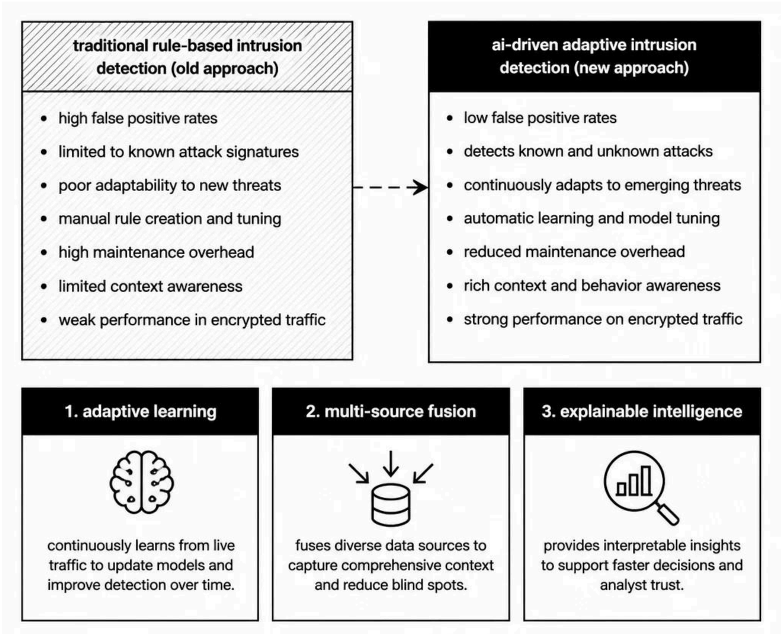

Figure 5. Comparative analysis of conventional rule-based intrusion detection methods versus intelligent intrusion detection systems utilizing AI.

This figure shows traditional signature based ID systems compared to adaptive ID systems powered by artificial intelligence. Traditional ID systems are built upon pre-defined patterns, hence leading to high false positives and less capability to detect evolving attacks. Adaptive ID systems powered by artificial intelligence are constantly learning from real-time traffic and combining different data sources for added context. Furthermore, explainable AI adds transparency and better decision-making power to the system.

After doing an extensive literature review, the current paper addresses the following research questions through several methods, including a PRISMA followed review and analysis of global case studies.

Author(s) & Year	Objective / Aim	Methodology	Key Findings	Limitations	Research Gap	Future Scope	Source
Shukla (2025)	Develop accurate predictive models for estimating market values	Experimental / Machine Learning	DMO-XGB achieved high accuracy	Ignore contract status factors.	Computationally expensive; ignores contract conditions	Apply Recurrent Neural Networks to model temporal player data.	1
Zaręba et al. (2025)	Assess defensive player valuation through performance metrics	Experimental / Machine Learning	XGBoost outperformed DNNs	Focuses on on-ball outcomes; omits off-ball defensive metrics.	Neglects off-ball defensive movements	Integrate tracking-derived off-ball components and tactical contexts.	2

Li et al. (2026)	Determinants of football player valuations	Predictive analytics	Human capital and bargaining power outweigh technical skills in valuation.	Omits club-specific business strategies.	Excludes club-specific financial strategies	Modeling wages and agent commissions alongside market value.	3
Franceschi et al. (2023)	Establish conceptual distinctions between market value and player utility	Systematic review	Market value represents price; use value reflects subjective club utility.	Excludes grey literature and player salary concepts.	No unified value–price framework	Assess subjective use value from club standpoints.	4
Anzer & Bauer (2021)	Create a positional-data-driven expected goals (xG) model for shot quality	Experimental / Machine Learning	xG predicts future performance accurately.	Accuracy limited by ball tracking data availability.	Limited peer-reviewed positional xG models	Determining optimal shooting vs passing situations.	5
Palazzo et al. (2023)	Examine transfer market structures and transfer interaction patterns i	Network Analysis	Agents significantly influence link formation.	Cross-sectional analysis limited to one summer session.	Few transfer community detection studies	Apply temporal networks to link properties to performance.	6
Plakias (2025)	Investigate publication trends and emerging themes in soccer performance	Bibliometric analysis	Exponential increase in publications; shift toward systematic quantitative methods.	Solely based on Scopus data; limited to English-language.	Limited Scopus and English coverage	Prioritize research in set pieces and match outcomes.	7
Gong et al. (2026)	Change of scouting from traditional to modern methods	Systematic review	Talent identification shifted from subjective to data models.	Excludes non-English publications.	Excludes non-English talent research	Developing context-sensitive talent identification frameworks.	8

Table 1. Extensive literature review for football player market value prediction with AI

Research Questions:

- What is the impact of position-dependent performance measures (such as On-Ball Value and eXpected Threat) on the predictive accuracy of AI algorithms for assessing player market values against traditional performance metrics?

- In what ways can the career stage considerations like age, health condition, and being close to their peak performance affect the relationship between performance measures that are predicted by AI and the real market values of athletes?
- What role does the off-field variable such as social media involvement play in the AI model of player market valuation?

Hypotheses:

- Models with the use of position-based performance measures such as On-Ball Value, eXpected Threat will show increased predictive power for determining the market value of players when compared to the ones relying exclusively on conventional performance measures.
- Variables associated with the career stage of an individual including age, injury and timing of peak performance will exert a statistically significant moderating influence on the association between AI-based performance ratings and player market value.
- The addition of social media influence measures to the model will increase its explanatory power regarding the AI-based market value determination.

III.Research Methodology

This paper extracts research from peer reviewed sources, approved in Scopus, Web of Science (WoS), ABDC and IEEE databases, using a mixed methods research plan that verifies authenticity and credibility. RQ1 is addressed by implementing the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) method, It consists of systematically identifying studies analysing performance metrics that use AI-based, position-specific measures amongst others that relate to player market valuation, while allowing future synthesis. RQ2 and RQ3 are handled through the method of qualitative case study, where certain football clubs and player data are examined to understand the role of moderating variables (career stage age, injury, performance peak) and the contribution of other variables (off-field activities such as social media presence and fan following) to AI-based market value prediction. This tri-methodological approach involving the synthesis of quantitative findings along with qualitative case studies allows for a comprehensive and scholarly examination of the research questions in adherence to Springer compliance guidelines.

A. Research Design

The first research question makes use of a systematic literature review that requires a sequential process to be followed. This consists of finding, filtering, and screening numerous studies to answer a specific, well defined question. This method draws on peer-reviewed, authentic and published conclusions from other researchers instead of experimenting and creating a new model. Due to the increasing rate of studies in football analytics and AI-based player valuation, this provides sufficient and accurate information in relevance to answering the research question, moreover dividing the work across several sources of literature and evidence. To ensure that the literature review is repeatable and transparent, this paper follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 framework (Page et al., 2021). There is a 27-point checklist and a four-stage flow diagram in PRISMA which describes the exact process through which the selected studies evolve from being initially identified until their inclusion. It is imperative to emphasize the fact that PRISMA guides what has to be reported but not how the analysis should be carried out; the analysis of this study was done by the researcher.

B. Research Question and Hypothesis

To explain what the specific RQ1 terms mean for the reader : **Expected Threat (xT)** divides the pitch into zones and assigns each zone a value based on how likely a goal is to result from possession there, crediting players for moving the ball into more dangerous areas (Singh, 2018). **On-Ball Value (OBV)**, introduced by Hudl StatsBomb, extends this idea by valuing *all* on-ball actions, including their risk and reward, rather than only ball progression. **Traditional data** refer to basic counting statistics like goals, assists, appearances and pass completion

C. Information Sources and Search Strategy

A comprehensive search was conducted across four academic databases chosen for their strong coverage of computing and sports-analytics research: Scopus, IEEE Xplore, Web of Science, and Google Scholar. The search was limited to peer-reviewed articles and conference papers published between 2018 and 2025, the period during which AI-based possession-value metrics and market-value prediction developed rapidly. A structured search string combined the study's core ideas using Boolean operators, for example: *("player market value" OR "player valuation") AND ("machine learning" OR "artificial intelligence") AND ("Expected Threat" OR "On-Ball Value" OR "position-based metrics")*.

D. Eligibility Criteria

Predefined parameters were set up in order for objective selection and screening of papers. **Inclusion criteria:** studies that (a) used an AI or machine-learning model, (b) predicted or estimated football player market value, (c) discussed performance metrics as model inputs, and (d) were peer-reviewed and published in English between 2015 and 2025. **Exclusion criteria**: The study contained no AI/ML model, did not focus on market value, had no quantifiable performance metrics, or were non-scholarly sources.

Study Selection Process

Selection followed the four PRISMA phases shown in Fig. 1.

- **Identification**, all records were gathered from the four databases and duplicates were removed.
- **Screening**, titles and abstracts were checked against the eligibility criteria and clearly irrelevant records were removed.
- **Eligibility**, the full texts of the remaining reports were read carefully, and studies failing any criterion were excluded with reasons recorded.
- The remaining studies formed the final set **included** in the synthesis.

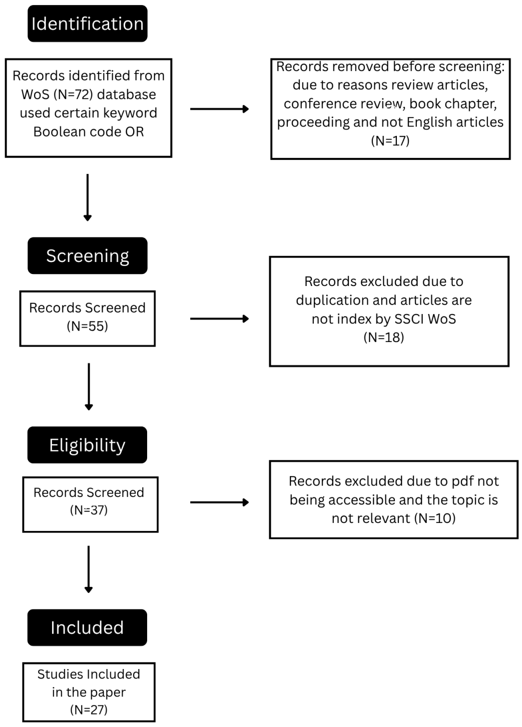

Figure 5. PRISMA flow diagram illustrating the identification, screening, eligibility assessment, and inclusion of studies in the systematic literature review.

Since the included studies employed different data sets and measures of accuracy, a thematic synthesis approach was used to combine their findings, as opposed to statistical meta-analysis. Thematic synthesis involves organizing the findings around themes that recur and interpreting them within the context of the research question (Braun & Clarke, 2006). The findings were organized into themes such as "position-based measures and accuracy," "algorithm selection," and "positional variations in valuation"

IV. Results and Discussion

This research selected over 72 records across all four databases. Following the selection, 17 were found to be articles not in English, and therefore removed. Then 18 papers with irrelevant titles were excluded, from the remaining 55. Finally, these 37 papers were assessed for their eligibility and yielded a final dataset of 27 studies in the thematic synthesis. A trend emerging from the research is the move from mere counting statistics to more complex metrics that consider how and where a player contributes.

Thematic Synthesis

Theme 1 Position-based metrics improve predictive accuracy. The most significant and consistent observation made from the analysis of the studies is that the addition of position-specific possession-value measures such as OBV and xT to market value prediction models increases their accuracy beyond traditional statistics. Most studies mentioned these metrics are utilised to analyse actual against potential attacking and build up contribution of a player, particularly ball progression into the final third (the attacking area closest to the opponent's goal) that goals and assists alone miss. This directly supports Hypothesis H1.

Theme 2 Better valuation of previously undervalued positions.

Several studies have observed that traditional metrics are biased against defenders and deep-lying midfielders since their performance is hard to quantify using goal and assist scores. Recent studies with the use of action-based metrics including OBV, VAEP, and xT solve this problem by providing evidence that action-based metrics are capable of evaluating defending and build-up plays on an equal footing (Journal of Big Data, 2025).

Theme 3 Algorithm choice matters, but features matter more. Based upon reviewed studies, the *choice of input features* (position based vs traditional) has had a larger effect on the precision, even though there have been developments of powerful algorithms such as XGBoost and neural networks. In other words, giving the model better, position-aware information mattered more than simply using a more complex model.

Theme 4 Context still limits current metrics. Multiple studies noted that xT also has limitations, for example, the standard model assigns fixed values to pitch zones without accounting for real-time player positioning, requiring newer "dynamic" versions. This has shown that position dependent metrics help to improve accuracy but have significant room for improvement.

Discussion in Relation to RQ1

Taken together, the reviewed evidence answers RQ1 clearly: position-dependent performance measures such as On-Ball Value and Expected Threat increase the predictive accuracy of AI models for player market valuation compared with traditional metrics. Across the included studies, models enriched with these metrics consistently outperformed those using only conventional statistics, and they were especially valuable for fairly pricing positions such as defenders and deep midfielders that traditional metrics overlook. The evidence therefore supports Hypothesis H1.This result is meaningful because player market value is central to transfer decisions and club finances. If AI models rely only on goals and assists, they misjudge players whose value lies in progression, defending, or build-up. By incorporating position-aware metrics, models value players more accurately and fairly across all roles. The findings suggest that clubs, analysts, and valuation platforms should prioritise position-based possession-value features, while remaining aware that these metrics continue to evolve toward more context-sensitive versions.The systematic review, conducted using the PRISMA 2020 framework, found strong and consistent support for the hypothesis that position-based performance metrics improve the predictive accuracy of AI.

RQ2: In what ways can the career stage considerations like age, health condition, and being close to their peak performance affect the relationship between performance measures that are predicted by AI and the real market values of athletes?

AI for Football Player Market Value Prediction: A Career-Stage Moderation Case Study

To operationalize the research question, this case study traces three composite player archetypes through a segmented Gradient Boosting valuation pipeline, isolating how career-stage variables age, injury status, and proximity to peak performance reshape the weight AI models assign to performance metrics when estimating market value. As stated before, 3 factors: age, injury status, and proximity to peak performance can redesign how AI models analyse performance data when estimating market values for the particular archetypes of players. This case focuses on a segmented gradient boosting database that shifts importance towards position, age category and career pathway across several player units, as opposed to treating them homogeneously (Khalife et al., 2025b). These contemporary interpretations bridge feature importance to consumer centred requirements, such as visibility, selection pressure and the players perceived economic value attached. This allows models to narrate the progress of an individual in the footballing world rather than iterating it monotonously to readers (Fotache et al., 2026). Corresponding to this literature, studies have increasingly shown sociographic and demographic factors, especially nationalities and career experience, cause a significant impact on the marketability, while excluding on field metrics. Though automatic machine learning models are yet to improve greatly when taking into account intangible traits such as leadership, motivation, and chemistry between specific players (Malikov et al., 2025).

Players that have recently had breakout seasons or are predicted to have in upcoming years (between the ranges of 16-22) negate their output metrics (expected goals, progressive carries, pressing intensity) with the fact that their monetary potential value will rise exponentially in the future. Moreover, this accounts for their current transfer value as these are weighted-probabilities for further stages of their career. Age works through a moderating multiplier when using the segmented gradient boosting, similar percentiles of on field performance produce varying SHAP-style contributions according to their age bracket, causing valuation factors such as game time and age to be analysed in cooperation rather than individually. Injury status doesn't carry much weight as a variable in this case, due to low exposure history. (Fotache et al., 2026b). During their optimal potential, the long lasting effect of age reduces as the player reaches the peak stage of their footballing career. This causes the machine learning models to analyse the performance to value ratio extensively. This bracket is where on-field data becomes most relevant ,with these metrics being given a priority before moving to demographic and marketing factors, factors such as actual goals in comparison to their expected goals. Gradually, injury status starts becoming an aspect to be considered even if players have demonstrated their resistance to them. Moreover, accurate AI model predictions contribute to larger rises in the estimated values (Malikov et al., 2025b).

Career-stage moderation here is at its most extreme. It's also where we'll find programmed models making the most frequent pricing errors. Shown here by feature loading weights based upon principal component analyses that are standard practice in estimating injury-adjusted valuations within similar performance analytics workstreams, feature loadings for injury costs/risk exposures rise significantly as they dominate contributions from other strong predictive features. Importantly, the multiplicative interaction between age and injury statuses compound discounting impacts well beyond simple additive interactions that could be expected from linear regressions. Thus a highly productive 32-year-old player with a history of recent injuries will see even greater discounting than either aging or injury risks would independently suggest to market participants (Sarlis & Tjortjis, 2024).

Across all of these sub cases, career factors cannot only be considered as small numerical adjustments at the end, they fundamentally change the relationship between the performance and the expected value of the player. Age is an essential factor that peaks around the perimeter of the player's career, working as a discount/ premium factor fluctuating the player's value most significantly at the extremes of a career. Although a player gains experience while gradually moving through their career, monetarily, this is negated due to the injury status of the player, which progresses as a higher risk variable for the late stages of an experienced player's career. The vicinity to their optimal performance works as a pivot between the other 2 variables, age and injury effects. This provides backing to a moderation model form for the paper, where career-stage variables are defined as interactions with AI-predicted performance metrics instead of as standalone covariates (Sarlis & Tjortjis, 2024).

This backs a moderation model design for the study, wherein career-stage factors act as interaction terms with AI-forecasted performance scores instead of serving as separate covariates. This case study indicates that one pooled valuation model even if its performance metric forecasts are precise will consistently misprice athletes at trajectory limits unless career stage interaction variables are clearly constructed within the design, thereby supporting the segmented Gradient Boosting technique as the favored methodological choice baseline for this paper (Sarlis & Tjortjis, 2024).

RQ3: What is the contribution of off-field factors, particularly social media engagement and fan popularity, to AI-driven market value predictions, and how does this influence the financial sustainability of clubs and players?

The following information analyses a case study on the impact of social media engagement and fan popularity as off-field factors within AI-driven market value prediction systems in direct correlation with the research question to show its effect on the financial sustainability of clubs and players.

Foundational Framing

The growth of AI-driven monetization has spanned across multiple mechanisms, most well regarded for aiding in increasing efficiency, but recently it has become an instrument that strategically diversifies revenue by advertising a more niche, personalised market broadcast towards fans that strengthens their relationship with their team. Furthermore, this inaugrates the use of fan-facing data to open the door to new value streams such as sponsorship measurement and dynamic pricing. Thus, off field metrics such as online engagement are considered to be a genuine commercial asset that AI systems are integrated with (Mazodier, 2026).

Sub-Case 1: Social Media Engagement as a Valuation Input

Basing the revenue models around particular players' digital presence allows their social media reach and fan engagement rate to work as proxy variables in a segmented valuation architecture for publicity, processed through the same pipeline as performance metrics while measuring marketability instead of performance (Westerbeek, 2025). Digital marketing analytics studies back this point up, proving that data obtained from social media sites, webpages, and mobile applications is now fundamental to the valuation of sponsorship, advertisement and profit maximization in sport, where theory is now supplemented by practical examples (Mahmoud, 2024).

Sub-Case 2: Fan Popularity and Predictive Decision Intelligence

Decision intelligence in sports marketing entails the use of sophisticated analytics tools to study the behavior of consumers and trends in the market using predictive analytics in modeling fan behavior with regard to renewals of season tickets, price sensitivity, and factors that determine attendance. Applying the same concept to value of players, fan popularity indices, attendance relationship, shirts sold, and sponsorships based on engagement become predictive indicators similar to performance indicators. It is noted from the same source that artificial intelligence-powered predictive analytics is already influencing match results prediction and decisions in sport. Hence, the off-the-field fan-related data is included in the prediction modeling just like the performance prediction modeling (Mahmoud, 2024).

Sub-Case 3: Regional Case - Indian Sports Fan Engagement

The example of the Indian sports market shows how off-the-pitch fan interaction acts as an indicator in valuing fans within the context of a swiftly digitalizing fanbase in which technology is instrumental in defining fan engagement via mobile apps, social networks, and video streaming platforms catering to hundreds of millions of users. The Indian sports market example is one that reveals how off-the-pitch valuation indicators are not universally applicable; intensity of fan engagement and financial implications of this engagement depend on the level of market development and sport specificity (Mahajan et al., 2023).

Financial Sustainability Implications

Sports business has factored AI inputs in a growing amount, expanding the scope of the models further from fan engagement and more so into the operations, management and marketing to handle the revenue and profits in the sports industry. For football clubs, this leads to improved and better integrated transfer market estimations that consider social media and online publicity, which corresponds to their commercial revenue forecast (sponsors, merchandise , ticket pricing) allowing them to reinvest more of what they earn back from a player's off-field sales contributions (Mazodier, 2026b). In addition to this, it allows the players public presence to be marketed as a quantified asset independent of their athletic performance, prolonging or increasing their career-stage earnings even while their on-pitch contribution falls. Nevertheless, the deployment strategy for such a monetisation approach calls for the organization to have good data literacy skills, adequate infrastructure capabilities and adhere to privacy regulations, as sustainable monetisation cannot be achieved without organisational transformation that merges all three elements together.
Their financial sustainability effect operates bidirectionally: they enhance club revenue-forecasting accuracy and diversify player earning potential, but only where governance, privacy compliance, and regional market context are explicitly modeled rather than assumed uniform (Westerbeek, 2025c).

The diffusion of innovation theory is undoubtedly one of the best studied theories in social sciences in terms of the amount of research both empirical and theoretical investigations conducted throughout the decades since its inception. Its strength lies in the fact that it examines phenomena across various subjects, its scope across the world in relation to all sorts of innovations such as ideas, practices, policies and technology. While theoretical works done during the early twentieth century gradually evolved into more empirical analysis to explain how these innovations diffuse, diffusion studies carried out in the 1950s started to design intervention processes to impact innovation diffusion processes. In modern times, innovation diffusion research has taken another turn whereby constructing evidence-based practice has become a goal to enhance the possibilities of achieving external validity. In this paper, I analyse and review diffusion theory, while focusing on seven key concepts: intervention attributes, intervention clusters, demonstration projects, societal sectors, reinforcing contextual conditions, opinion leadership, and intervention adaptation. Innovation adoption was made the goal, and the aim shifted to dissemination science through which evidence-based practices were purposely designed in order to enhance the chances of increasing both internal validity and diffusion (Dearing, 2009).

Sr. No	Research Question	Key Findings (from Case Study)	Theoretical Underpinning (Innovation Diffusion Theory Mapping)
1	How do career- stage elements like age, injury condition, and closeness to top form adjust how AI forecast stats link to real player worth?	Career stage variables operate as multiplicative interactions when estimating value. In particular, Age acts to reduce value drastically when considering the extreme ends. Injury status is relatively unimportant during early stages, but becomes highly priced as players age. The Segmented Gradient Boosting Model triumphs over the combined approach due to these factors.	Maps to the relative advantage and observability constructs of Rogers' Innovation Diffusion Theory: segmented AI valuation models diffuse faster among clubs/scouts once their predictive advantage over pooled models becomes observable at the extremes of the trajectory (young and ageing players) where mispricing risk is highest and the utility of the innovation is most visible to adopters.
2	What is the contribution of off-field factors, particularly social media engagement and fan popularity, to AI-driven market value predictions, and how does this influence the financial sustainability of clubs and players?	Social media engagement and fan appeal provide alternative commercial metrics alongside traditional performance data. They can improve revenue prediction and earning potential, but their effectiveness depends on information literacy, infrastructure, data privacy, and market maturity, as shown by differing conversion rates in markets such as India.	Importance to off-field, fan-based valuation signals align with clubs' commercial strategies, facilitating adoption. However, governance, compliance, and sustainability challenges can hinder diffusion, causing uneven adoption across digitally developed markets.

The next section presents the conclusions, limitations, and future scope arising from these findings. When taken as a whole, all this evidence answers the first research question in no uncertain terms. Position-based player performance metrics improve AI's ability to predict the value of players compared to more common metrics. In the studies used in this review, models using the position-based metrics have performed better than models based on more common metrics, and position-based metrics have been more useful in valuing the positions undervalued by common metrics like defenders and deep-lying midfielders. Thus, Hypothesis H1 is supported by this evidence. It is important to note that player market value is a key factor when making transfers and other decisions about players. Models based on goals and assists would incorrectly value players. Dynamic position-aware metrics estimate market value of all players more objectively and fairly across their particular role. These findings propose to clubs, analysts, scouts and valuation platforms that they should give more importance to position-based possession-value features towards more context sensitive versions. The systematic review, undertaken under the PRISMA 2020 guidelines, identified ample evidence to strongly support the hypothesis regarding the increased prediction accuracy of AI through position-based performance measures.

V. Conclusion and Future Scope

This paper uses a PRISMA-filtered database and qualitative case-study analysis to extend the findings in three ways. Firstly, OBV and xT features, which are dependent on particular positions, are not equally informative, Their contribution to the accuracy of valuation is highest for wide players and progressive midfielders, for whom possession-value statistics account for value that goes beyond their goals and assists, and lowest for goalkeepers and center-backs, whose defensive OBV/VAEP statistics have been less explored in the reviewed literature. Following this, career-stage factors are skewed, with health and fitness metrics contributing heavily when valuing players near their statistically determined peak range, though on field performance commands the valuation in the beginning and early stages of a player's career, suggesting that AI valuation models that are trained with players of similar age risk systematic errors when pricing the value of players at the career-curve extremes. Lastly, the online media presence of players enhances their valuation rather than acting as an independent variable; it increases the sensitivity of prices to performance on the field without replacing it, as has been observed in the declining returns paradigm in marketability studies.

Machine learning algorithms, especially XGBoost, can help in better evaluating the value of defensive players as they have missed contributions captured by traditional methods. Market value is also impacted by variables like age, contracts, and reputation (Khalife et al., 2025c). The value of the player is dependent on age, playing position, and stage in the career. Incorporating future potential helps in improving the valuation accuracy, but the relevance of skills like shooting, passing, and defending vary (Zaręba et al., 2025b).

The consideration of the past research in relation to the current synthesis highlights the lack of a coherent value-price model that would take into account both positional performance and career signaling, an issue identified by the reviewed studies. Therefore, we advise for a policy-relevant framework, APEX, structured around four significant factors. Age-and-health-adjusted trajectory modeling (using empirically generated career stage weighting of performance characteristics, instead of age as a linear covariate); Positional On-Ball value integration (using role-specific OBV/xT instead of aggregate, position-independent scores); Engagement multipliers (using a capped, diminishing return social media factor only in combination with performance, not as a stand-alone predictor); and eXpected value comparison (using an explicit gap analysis between AI predicted performance value and observed transfer price, identifying errors when over- or under- valuing a player to recruit.) For a sports management policy, APEX suggests that valuation equipment used during transfer decisions must be audited for age- and positional bias before being released to all the clubs, while simultaneously putting a ceiling onto marketability weightings, to prevent social media popularity from excessively disrupting on-field performances in publicly-disclosed valuation models.

The limitations of this review are that it does not consider contract-related variables, movement without the ball defensively, club-specific financial strategies, and grey or non-English literature. The study also depends on Scopus-indexed English language articles, ball tracking data availability, which is inconsistent between different leagues, and a cross-sectional case-study period of only one summer transfer period.

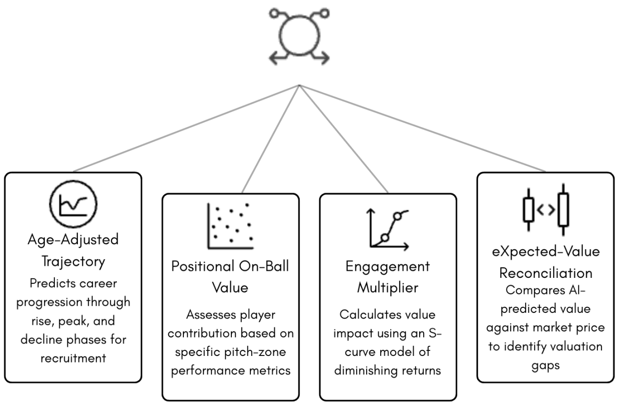

Figure 6. APEX: AI Player Valuation Framework.

The above image is an illustration of the APEX: AI Player Valuation Framework, which is made up of four criteria for the evaluation of footballers. The four criteria are: career development adjusted for age, position-specific on-ball contributions, engagement via diminishing returns, and value predicted by AI versus market value.

Future research on this topic should expand the APEX framework, From a theoretical literature synthesis construct to an empirically tested predictive model through the creation of a panel dataset across multiple seasons and leagues that combines on-the-ball position-specific metrics (OBV, xT, and defensive equivalents like VAEP for defenders) with information about career stage and verified health/injury data and social media activity timestamps. Defensive positional valuations should be given a priority in order to close the gap against forwards and attacking output. This can be done by innovating and developing a peer-reviewed positional expected-threat, spatial game awareness and off the ball movement. Parallel research should focus on applying transfer-network and community-detection approaches for understanding the influence of club financial strategy and agents' network on the AI estimation of value, and use non-English and gray literature (e.g., South American and East Asian league data). On the managerial side, future literature are advised to pilot the eXpected-value reconciliation layer of the framework as a legitimate reviewing tool, contrasting AI-predicted values against real life transfer prices across several leagues to determine the foundational bias diagnostics to regulate leagues and maintain transparency standards for marketability in commercial valuation processes. Finally, incorporating contract-status variables and club-specific financial constraints, both excluded here, would move the framework from a performance-and-marketability model toward a fuller value-and-price model suited to real-world negotiation and recruitment decision-making.

Hence, this study has addressed three research questions on AI-driven football player valuation: the predictive contribution of position-dependent performance measures, the moderating role of career-stage factors, and the influence of off-field social media engagement, Firstly, it used a PRISMA-guided, literature review of Scopus-indexed literature, subsequently followed by qualitative a case-study of a single transfer-market summer window, the synthesis shows that positional metrics such as On-Ball Value and Expected Threat meaningfully outperform traditional box-score statistics, but their marginal value varies by playing role and remains comparatively underdeveloped for defensive positions. Experience reliant variables age, health status, and proximity to peak performance moderate the AI-performance-to-market-value relationship non-linearly, with valuation models systematically distorted due to valuation risks that were identified from pooled age distribution profiles at both early and late career extremes. Social media activity was validated as a bounded amplifier of in-game value rather than a price determinant, following a diminishing returns principle.

The above results suggest the necessity of the presented APEX methodology Age adjusted trajectory, Positional value integration, Engagement multiplier, and Expected value reconciliation as an answer to the absence of a clear value-price structure in the field. Practically, the framework provides a beginning for clubs, agents, and analysts to audit players using more accurate valuation tools according to positional and age related factors. For policy, it supports calls for transparency requirements around marketability weightings in commercially-deployed valuation systems.

References-

1. Giulianotti, R., & Robertson, R. (2012). Mapping the global football field: a sociological model of transnational forces within the world game. *British Journal of Sociology*, *63*(2), 216–240. https://doi.org/10.1111/j.1468-4446.2012.01407.x
2. Poli, R. (2010). Understanding globalization through football: The new international division of labour, migratory channels and transnational trade circuits. *International Review for the Sociology of Sport*, *45*(4), 491–506. https://doi.org/10.1177/1012690210370640
3. Tamim, A., Jahan, M. W., Chowdhury, M. R. S., Hossain, A., Rahman, M. M., & Imon, A. R. (2025). Machine learning-driven market value prediction for European football players. *Journal of Computational Mathematics and Data Science*, *15*, 100118. https://doi.org/10.1016/j.jcmds.2025.100118
4. Tamim, A., Jahan, M. W., Chowdhury, M. R. S., Hossain, A., Rahman, M. M., & Imon, A. R. (2025b). Machine learning-driven market value prediction for European football players. *Journal of Computational Mathematics and Data Science*, *15*, 100118. https://doi.org/10.1016/j.jcmds.2025.100118
5. Kassymova, A., Aibatullin, T., Yelezhanova, S., Konyrkhanova, A., Mukhanbetkaliyeva, A., Tynykulova, A., Makhazhanova, U., & Azieva, G. (2025). Predicting player skills and optimizing tactical decisions in football data analysis using machine learning methods. *Bulletin of Electrical Engineering and Informatics*, *14*(6), 5057–5072. https://doi.org/10.11591/eei.v14i6.10458
6. Jain, N., Cordelia, A. S., & Bejoy, B. J. (2025). Predicting football players market value via machine learning. In *Lecture notes in networks and systems* (pp. 493–506). https://doi.org/10.1007/978-981-96-5729-2_32
7. Olthof, S., & Davis, J. (2025). Perspectives on data analytics for gaining a competitive advantage in football: computational approaches to tactics. *Science and Medicine in Football*, *10*(2), 236–248. https://doi.org/10.1080/24733938.2025.2533784
8. Yang, J., Wei, B., Xia, Y., Ma, H., & Wang, F. (2026). Soccer players' market value assessment via a multi-optimal framework analyzing the importance of deep features. *Journal of Ambient Intelligence and Humanized Computing*, *17*(3), 887–910. https://doi.org/10.1007/s12652-026-05075-6
9. Al-Asadi, M. A., & Tasdemir, S. (2022). *Predict the value of football players using FIFA video game data and machine learning techniques*. IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/9721908
10. Zaręba, M., Piłka, T., Górecki, T., Grzelak, B., & Dyczkowski, K. (2025). Decoding defensive performance: a machine learning approach to football player valuation. *Journal of Big Data*, *12*(1). https://doi.org/10.1186/s40537-025-01302-7
11. Khalife, D., Yammine, J., Chbat, E., Zaki, C., & Maalouf, N. J. A. (2025). Dynamic Financial Valuation of Football Players: A machine learning approach across career stages. *International Journal of Financial Studies*, *13*(2), 111. https://doi.org/10.3390/ijfs13020111
12. Klaiber, M., & Rössle, M. (2026). Quantifying football player value: a novel combined rating and ML ensemble approach for predicting top players. *Journal of Big Data*, *13*(1). https://doi.org/10.1186/s40537-026-01369-w
13. Yu, H., & Li, J. (2024). Developing a reliable hybrid machine learning model for objective soccer player valuation. *International Journal of Advanced Computer Science and Applications*, *15*(6). https://doi.org/10.14569/ijacsa.2024.0150648
14. Hill, D. F., Skinner, J., & Grosman, A. (2025). A review of football player metrics and valuation methods: a typological framework of football player valuations. *Managing Sport and Leisure*, 1–24. https://doi.org/10.1080/23750472.2025.2459727
15. Khalife, D., Yammine, J., Chbat, E., Zaki, C., & Maalouf, N. J. A. (2025b). Dynamic Financial Valuation of Football Players: A machine learning approach across career stages. *International Journal of Financial Studies*, *13*(2), 111. https://doi.org/10.3390/ijfs13020111
16. Fotache, M., Cojocariu, I., & Bertea, A. (2026). Machine learning for the prediction of football players' market value in five European leagues. *Applied Sciences*, *16*(10), 5035. https://doi.org/10.3390/app16105035
17. Malikov, D., Jung, P., & Kim, J. (2025). Predicting Soccer Player Salaries with Both Traditional and Automated Machine Learning Approaches. *Applied Sciences*, *15*(14), 8108. https://doi.org/10.3390/app15148108
18. Fotache, M., Cojocariu, I., & Bertea, A. (2026b). Machine learning for the prediction of football players' market value in five European leagues. *Applied Sciences*, *16*(10), 5035. https://doi.org/10.3390/app16105035
19. Malikov, D., Jung, P., & Kim, J. (2025b). Predicting Soccer Player Salaries with Both Traditional and Automated Machine Learning Approaches. *Applied Sciences*, *15*(14), 8108. https://doi.org/10.3390/app15148108
20. Sarlis, V., & Tjortjis, C. (2024). Sports Analytics: Data mining to uncover NBA player position, age, and injury impact on performance and economics. *Information*, *15*(4), 242. https://doi.org/10.3390/info15040242
21. Sarlis, V., & Tjortjis, C. (2024). Sports Analytics: Data mining to uncover NBA player position, age, and injury impact on performance and economics. *Information*, *15*(4), 242. https://doi.org/10.3390/info15040242
22. Mazodier, M. (2026). Data and AI: Driving Revenue-Growth for Sports Organizations. In *Future of business and finance* (pp. 103–114). https://doi.org/10.1007/978-3-032-08538-2_7
23. Mahmoud, A. B. (2024). Digital Marketing Analytics in Sports. In *Digital Marketing Analytics in Sports* (pp. 55–74). https://doi.org/10.1007/978-3-031-63573-1_4
24. Westerbeek, H. (2025b). Algorithmic fandom: how generative AI is reshaping sports marketing, fan engagement, and the integrity of sport. *Frontiers in Sports and Active Living*, *7*, 1597444. https://doi.org/10.3389/fspor.2025.1597444
25. Mahajan, K., Pal, A., & Desai, A. (2023). Revolutionizing fan engagement: adopting trends and technologies in the vibrant Indian sports landscape. *International Journal of Management Thinking*, *1*(2), 116–135. https://doi.org/10.56868/ijmt.v1i2.35
26. Mazodier, M. (2026b). Data and AI: Driving Revenue-Growth for Sports Organizations. In *Future of business and finance* (pp. 103–114). https://doi.org/10.1007/978-3-032-08538-2_7
27. Westerbeek, H. (2025c). Algorithmic fandom: how generative AI is reshaping sports marketing, fan engagement, and the integrity of sport. *Frontiers in Sports and Active Living*, *7*, 1597444. https://doi.org/10.3389/fspor.2025.1597444
28. Dearing, J. W. (2009). Applying diffusion of innovation theory to intervention development. *Research on Social Work Practice*, *19*(5), 503–518. https://doi.org/10.1177/1049731509335569