Bulletin of Electrical Engineering and Informatics
Vol. 14, No. 6, December 2025, pp. 5057~5072
ISSN: 2302-9285, DOI: 10.11591/eei.v14i6.10458

Predicting player skills and optimizing tactical decisions in football data analysis using machine learning methods

Akmaral Kassymova[1], Tolegen Aibatullin[1], Shynar Yelezhanova[2], Assem Konyrkhanova[3], Ainur Mukhanbetkaliyeva[3], Assemgul Tynykulova[4], Ulzhan Makhazhanova[5], Gulmira Azieva[5]
[1]Institute of Digital Economy and Sustainable Development, Higher School of Information Technology, Zhangir Khan University, Uralsk, Republic of Kazakhstan
[2]Department of Software Engineering, Kh. Dosmukhamedov Atyrau University, Atyrau, Republic of Kazakhstan
[3]Department of Information Security, Faculty of Information Technologies, L.N. Gumilyov Eurasian National University, Astana, Republic of Kazakhstan
[4]Higher School of Information Technology and Engineering, Astana International University, Astana, Republic of Kazakhstan
[5]Department of Information Systems, Eurasian National University named L.N. Gumilyov, Astana, Republic of Kazakhstan

Article Info

Article history:

Received Apr 11, 2025
Revised Sep 18, 2025
Accepted Sep 27, 2025

Keywords:

Artificial intelligence in sports
Football analytics
Gradient boosting machines
Machine learning
Player skill prediction

ABSTRACT

This study investigates the integration of machine learning (ML) techniques into football analytics to predict player skills and optimize tactical decisions. A dataset of over 150,000 professional match actions from various leagues and seasons was analyzed using deep neural networks, convolutional neural networks (CNNs), and gradient boosting machines (GBM) algorithms on biometric, contextual, and match data. The valuing actions by estimating probabilities (VAEP) metric indicated scores from +1.8 to +3.0 for key players, enabling detailed performance evaluation. CNN models achieved up to 91% precision, 88% recall, and a receiver operating characteristic – area under the curve (ROC-AUC) of 0.94, confirming their effectiveness in predicting player actions and contributions. Injury risk prediction using eXtreme gradient boosting (XGBoost) reached an F1-score of 0.87 and a ROC-AUC of 0.92, offering actionable insights for injury prevention and optimal player rotation. The findings highlight artificial intelligences (AI)'s capacity to support individualized preparation, tactical adjustments, and cost-effective recruitment strategies. While computational demands and data quality remain challenges, the results demonstrate the transformative potential of AI in modern football, providing a practical framework for data-driven decision-making to enhance team performance and strategic planning.

Corresponding Author:

Tolegen Aibatullin
Institute of Digital Economy and Sustainable Development, Higher School of Information Technology
Zhangir Khan University
Uralsk, Republic of Kazakhstan
Email: t10938456@gmail.com

1. INTRODUCTION

Modern football [1]–[3] has transformed from a purely athletic competition into a data-driven [4] industry that leverages technological [5], [6] innovation for performance enhancement [7]–[9]. As the sport grows in complexity and competitiveness, success increasingly depends on the strategic use of data analytics [10], [11] and artificial intelligence (AI). Teams at all levels, from grassroots to elite, have embraced data science tools for scouting, training, and match preparation. Among these tools, deep learning (DL) methods [12] —a subset of AI [13]–[15] —have proven revolutionary due to their ability to process vast datasets, uncover hidden patterns, and generate actionable insights. Despite their growing application, several challenges persist in football analytics. These include the difficulty of processing complex, high-volume data involving biometric indicators, match context, and dynamic tactical interactions. To support scouting and talent identification, the study employs supervised learning models such as random forest (RF) and gradient boosting machines (GBM), trained on historical performance metrics and biometric data to predict player potential. These models allow ranking athletes based on multidimensional characteristics such as consistency, physical development trends, and match impact. In addition, clustering techniques (e.g., K-means) are used to group players with similar profiles and detect emerging talents. This combination enables more objective, data-driven recruitment, allowing clubs to identify undervalued players who may not be visible through traditional scouting.

Traditional statistical methods [15]–[17] fall short in providing the precision and adaptability required to extract meaningful patterns from such multifaceted data. Consequently, there is a pressing need for robust machine learning (ML) [18] frameworks capable of addressing these limitations. Although recent research has made strides in automating football video analysis, gesture recognition, and tactical event detection, many gaps remain in integrating predictive models into real-time decision-making [19]–[22] and player management. Key questions involve how to effectively quantify player contributions, optimize tactics during live matches, and proactively manage injury risks—all while ensuring model interpretability and operational feasibility for coaching staff.

To address these gaps, this study aims to develop and apply advanced ML approaches to three core tasks in football analytics:

– Predicting player skills using neural networks and the valuing actions by estimating probabilities (VAEP) metric
– Enhancing scouting and talent identification through AI-based modeling
– Optimizing tactical decisions via convolutional neural networks (CNN) and temporal modeling

Through this research, we propose a comprehensive methodology using CNNs, long short-term memory (LSTM) networks, and GBM models to analyze over 150,000 football match events enriched with contextual and biometric data. The objective is to demonstrate how DL can improve decision-making in player evaluation, tactical planning, and risk management, ultimately supporting clubs in becoming more data-informed and competitive.

Yang *et al*. [23] propose a method for automatically recognizing football referee gestures (FRGR) based on the YOLOv8s DL model. To address the problem of gesture diversity, complexity, and background noise, they introduced three improvements: a global attention mechanism (GAM), a P2 structure to improve small object recognition, and a new minimum point distance intersection over union (MPDIoU) loss function to use anchor boxes optimally. Experiments on a dataset of 1,200 images showed that the proposed method outperforms existing models, achieving 89.3% accuracy and a significant increase in mAP. The developed approach demonstrates high efficiency and prospects for application in automated gesture analysis at football matches. Prasanth and Nallavan [24] consider the automation of football match video analysis using deep neural networks. The work covers player tracking, action recognition, semantic segmentation, and event detection using CNN, recurrent neural network (RNN), and 3D CNN. The paper focuses on tactical analysis and performance evaluation, as well as integrating DL with traditional analytics methods. Key challenges such as data labeling, scalability, and real-time processing are discussed, and future research directions are suggested. The paper provides a roadmap for researchers and practitioners interested in advancing sports analytics technologies using DL. Athanesious and Kiruthika [25] analyze the key challenges in automating football match analysis, including dynamic backgrounds, ball localization, fast-paced events, and player overlaps. A novel ball possession prediction scheme based on spatio-temporal features and positional sequence is proposed to address these challenges. The paper uses homographic perspective transformation (HPT) for pitch segmentation, as well as a hybrid YOLO detector with weighted intersect fusion (WIF) to track moving objects (the ball, players, and referees) using their centroid analysis and geometric approximation. Experimental analysis based on ISSIA La Liga and EPL matches demonstrates the high efficiency of the proposed system in identifying game events and calculating ball possession.

In parallel, the study considers the practical aspects of implementing the proposed approaches in a real football environment. In particular, the possibilities of using automated analysis of football matches in player training, team resource management, and tactical decision-making are emphasized. These aspects aim to improve the efficiency of individual training of athletes and teamwork in general. The authors emphasize that the proposed methods can be integrated into the training process, allowing coaches to obtain more in-depth information about the game, adjust tactical strategies, and identify the strengths and weaknesses of players. This can significantly improve the quality of match analysis and preparation for future competitions, especially in the context of increasing competition in the professional arena. This work provides not only a theoretical justification for using DL in the analysis of football matches but also applied results that can be useful to coaches, analysts, and clubs. The practical value of the proposed approaches lies in the possibility of using the obtained data to improve team performance and optimize management processes in football.

Recent studies in football analytics have applied a wide range of ML methods. Traditional algorithms such as RF and GB have been widely used for talent identification and ranking players due to their interpretability and robustness on structured data [19]. However, these approaches are limited in capturing complex spatial–temporal dependencies that characterize football matches. DL architectures address this gap: CNNs are particularly effective in extracting spatial features from positional data and pitch zones [12], while RNNs, especially LSTM models, capture sequential dependencies in player actions over time [24]. More recently, graph neural networks (GNNs) have been explored to represent passing networks and tactical interactions as graphs, enabling a more realistic modeling of team dynamics [13].

Comparative analyses in the literature show that CNNs often outperform classical ML models in action recognition tasks, achieving higher accuracy and receiver operating characteristic – area under the curve (ROC-AUC) values [23]. LSTMs, in turn, demonstrate superiority in predicting sequential events such as passing chains or counterattacks [14], while GNNs excel in understanding collective tactical patterns across multiple players [20]. Nevertheless, most prior works focus either on isolated prediction tasks (e.g., event detection, referee gesture recognition) or rely primarily on event-only data without integrating physiological or contextual information.

The novelty of this study lies in three aspects. First, unlike earlier works limited to event logs, we incorporate biometric and contextual features (heart rate, fatigue, and weather conditions), which substantially enhance the interpretability of predictions. Second, we combine multiple architectures-CNN for spatial patterns, LSTM for temporal sequences, and eXtreme gradient boosting (XGBoost) for injury risk-into a modular framework optimized for different analytical goals. Third, our extension of the VAEP metric integrates physiological and temporal dimensions, providing coaches with a more holistic evaluation of player contributions [26]. This integration goes beyond descriptive analytics and supports real-time tactical decision-making, addressing a gap not fully covered in previous literature.

2. METHOD

2.1. Dataset description

The dataset used in this study was obtained from the Wyscout platform, which provides standardized event logs, and tracking data for professional football (on average 1,800 events per match across more than 600 competitions) GitHub+9Hudl+9ResearchGate+9. It includes more than 150,000 game events covering approximately 1,200 matches and 3,800 unique players from the Premier League, La Liga, Bundesliga, Serie A, and Ligue 1. The time span covers six full seasons (2017/18–2022/23). Player demographics are diverse: the dataset contains athletes aged 18 to 37, with positional distribution of 28% defenders, 42% midfielders, and 30% forwards. Such coverage ensures that the models are trained on a representative sample of professional football across different tactical and competitive environments.

2.2. Model architectures and hyperparameters

Feature engineering was performed to derive skill-related metrics and prepare the dataset for ML models. Raw attributes included passes, shots, tackles, dribbles, interceptions, fouls, positional coordinates, sprint counts, heart rate, and fatigue index. From these, higher-level indicators were computed: passing accuracy (successful passes/total passes), dribbling efficiency (successful dribbles/attempts), and defensive contribution (successful tackles+interceptions per match). Biometric variables were aggregated using mean, maximum, and variability per match; for example, the fatigue index was defined as the ratio of high-intensity sprints to playing time. All continuous features were standardized using Z-score normalization, while categorical features (action type and player position) were encoded with One-Hot Encoding. To unify heterogeneous metrics, skill scores were expressed as a combination of normalized VAEP values and performance indices, providing a consistent scale for model training.

To address various predictive tasks, different ML/DL models were applied. The CNN architecture included three convolutional layers (64, 128, and 256 filters, kernel size 3×3), each followed by rectified linear unit (ReLU) activation and max-pooling (2×2). After the convolutional layers, two dense layers with 128 and 64 units respectively were added, followed by a softmax output. Dropout (rate=0.5) and L2-regularization (λ=0.01) were applied for regularization. The model was trained using Adam optimizer with a learning rate of 0.001 for 50 epochs and batch size of 64. The LSTM model consisted of two layers with 128 and 64 memory units, respectively, using dropout (0.3) and recurrent dropout (0.2). Early stopping with a patience of 10 epochs was employed to mitigate overfitting. The model was trained on sequential player actions encoded as time series. The XGBoost model used for injury prediction and VAEP probability estimation was configured with 500 trees, maximum depth of 6, learning rate 0.05, and subsample ratio of 0.8. Regularization parameters were set as λ=1 and α=0.1. Grid search was applied to tune hyperparameters.

Each ML model in this study was chosen based on its suitability for the specific nature of the task and data type. CNNs were selected for player action prediction and tactical analysis because of their proven ability to capture spatial dependencies in positional and event-based data, which is essential for analyzing pitch zones and movement patterns. LSTM networks were applied to model temporal sequences of player actions, as their gated architecture effectively learns long-range dependencies and sequential dynamics present in match events. GBM, particularly XGBoost, were used for injury risk prediction and VAEP probability estimation due to their high performance on structured, tabular datasets and their robustness to heterogeneous features (biometric, contextual, and categorical data). RF was employed in post-match analysis tasks where interpretability was important, enabling coaches and analysts to understand decision factors. This task-specific allocation of models ensured optimal exploitation of each algorithm's strengths while addressing the diverse analytical requirements of football data.

2.3. Validation strategy

To ensure generalizability, the dataset was split into 70% training and 30% testing subsets using a stratified sampling method, preserving proportional distributions of teams and seasons. For hyperparameter optimization and early stopping validation, 5-fold cross-validation was performed on the training data. Model selection was based on the average F1-score and ROC-AUC across the validation folds. For neural models (CNN and LSTM), the best weights were retained based on the lowest validation loss. The RF model was used to predict the success probabilities of various game actions, providing highly interpretable results. The final models were tested on the independent test set and used to produce all reported metrics and visualizations.

2.4. Valuing actions by estimating probabilities probability estimation

To calculate the VAEP metric, two separate XGBoost classifiers were trained:
- The first model estimated the ΔPgoal, i.e., the change in the probability of the team scoring within the next 10 actions.
- The second model predicted the ΔPagainst, i.e., the change in the probability of conceding a goal.

Both models used input features such as action type, location, time in match, previous sequence context, player fatigue level, and tactical positioning. These models provided the basis for assigning VAEP scores to individual actions. This study used advanced DL and ML techniques to analyze data and build models and specialized metrics such as VAEP. The main steps included data preprocessing, model training, and evaluation of their performance. This section describes the algorithms used and provides key formulas:
- Data pre-processing: pre-processing included imputing missing values using advanced statistical methods, standardizing and normalizing biometric and contextual data, and encoding categorical variables using One-Hot Encoding. This ensured the preparation of a homogeneous dataset for training the models.
- VAEP metric: the VAEP metric evaluates players' contribution to a match's outcome by providing a quantitative characteristic of each player's actions. It measures how each action (pass, shot, and tackle) affects the probability of scoring or preventing an opponent's goal. Formally, VAEP is defined as (1):

$$VAEP(a) = \Delta P_{goal}(a) - \Delta P_{against}(a) \tag{1}$$

where $\Delta P_{goal}(a)$ - change in the probability of a goal being scored as a result of action a, $\Delta P_{against}(a)$ - change in the probability of a missed goal as a result of action a. These probabilities are calculated using trained models that analyze the context of each action: the position of the ball, the current situation on the field, and the type of action.
- CNN: CNNs were used to analyze spatial and temporal match data. The main component of a CNN is a convolutional layer, which computes activations by applying filters to the input data. Activations are computed using (2):

$$z_{i,j} = \sum_{m=1}^{M} \sum_{n=1}^{N} x_{i+m,j+n} w_{m,n} + b \tag{2}$$

where x is the input data, w is the filter weights, b is the bias, and M and N are the filter sizes.
- RNNs LSTM: LSTMs have been used to model temporal dependencies in player actions. The main components of LSTMs include input, forget, and output gates (3):

$$\begin{aligned} f_t &= \sigma\left(W_f * [h_{t-1}, x_t] + b_f\right) \\ i_t &= \sigma(W_i * [h_{t-1}, x_t] + b_i) \\ o_t &= \sigma[h_{t-1}, x_t] + b_0 \end{aligned} \tag{3}$$

where f_t, i_t, o_t - oblivious, input and output gates, respectively, and σ is a sigmoid.

- GB (XGBoost): XGBoost was used to predict the risk of injury to players. The model minimizes the loss function L at each step (4):

$$L = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{4}$$

where l- error function, Ω - regularization term, f_k - decision trees.

- Evaluation metrics: the models were evaluated using precision, recall, F1 metric, and ROC-AUC metrics. The VAEP metric allowed us to quantify each player's contribution, providing a more in-depth analysis. The use of VAEP, in combination with the described DL methods, allowed us to achieve high results in evaluating player actions, optimizing tactical decisions, and predicting injuries. These methods have proven effective in analyzing complex football match data and are valuable in football analytics.

2.5. Tactical decision optimization

To complement the VAEP probability estimation, we designed a tactical decision optimization framework that integrates player performance, contextual features, and biometric indicators into a set of reproducible decision rules. The primary aim of this module is to support substitution planning and tactical adjustments in real time by linking quantitative player evaluations to actionable coaching strategies.

The optimization process is based on two principles: i) evaluating each player's contribution adjusted for fatigue and contextual workload and ii) adapting team formation in response to opponent strategies. Formally, the adjusted contribution score for player i is defined as (5):

$$Contribution(i) = VAEP(i) - \alpha \cdot Fatigue(i) \tag{5}$$

where $VAEP(i)$ represents the player's value of actions and $Fatigue(i)$ denotes the normalized fatigue index derived from biometric and workload features. The parameter α controls the weight of fatigue in decision-making.

To operationalize these principles, the following decision logic was implemented:

- Player ranking: players are ranked by adjusted contribution scores.
- Substitution rule: if $Contribution(i) < \theta$ and $Fatigue(i) > \beta$, then substitution of player i is recommended.
- Formation shift rule: if opponent pressing intensity exceeds a predefined threshold γ, a tactical adjustment is triggered (e.g., shifting from 4-4-2 to 4-2-3-1).
- Output: recommendations for substitutions and formation adjustments are generated and can be visualized for coaching staff in real time.

This approach ensures that tactical decision-making is not only data-driven but also transparent and reproducible. By combining player-level evaluations with contextual opponent information, the framework provides practical support for coaching staff, enabling timely interventions that enhance team performance.

3. RESULTS

The study used an expanded dataset of over 150,000 game actions collected from professional football matches across multiple leagues and seasons. This dataset was supplemented with additional contextual and biometric information for a more comprehensive analysis. Essential attributes of the dataset included game action types, including passes, shots, tackles, dribbling, interceptions, and defensive actions, as well as player metrics, such as positional data, action success rates, sprint speed, heart rate, fatigue level, and injury history. Match context, including opponent strategies, weather conditions, pitch quality, and game timing, was also integrated to improve the accuracy of the analysis. Several preprocessing methods were implemented to ensure data quality and fit for ML models. Missing values were imputed using advanced statistical interpolation methods. Categorical data, such as action types, were one-hot encoded to account for their diversity. Biometric and contextual data were standardized and normalized to achieve a uniform scale for all features.

Model development was conducted using several machine-learning approaches. CNNs were used to analyze spatial and temporal data related to match dynamics, which allowed us to identify complex patterns and dependencies. GBMs were used to evaluate and rank player contributions by performance. At the same time, LSTM networks demonstrated high performance in modeling temporal dependencies and predicting long-term player characteristics. Finally, given their biometric data, the XGBoost algorithm was used to predict injury risks and assess player impact. The models were evaluated using various metrics, allowing us to analyze their accuracy and robustness comprehensively. Accuracy, recall, and F1 metrics were used to assess the quality of predictions, while ROC-AUC demonstrated the robustness of the models in distinguishing between successful and unsuccessful actions. The VAEP metric allowed us to quantify the contribution of individual actions to team success, and the probability of injury risk was determined using binary classification models. The results demonstrate that the developed approaches provide a high level of accuracy and are applicable in real-life football analytics. They can be helpful for coaches, analysts, and club managers when making tactical and strategic decisions, increasing the competitiveness of teams. The methods discussed show the potential for integrating AI into the sports industry and set the direction for further research in this area.

The experiment included an extensive setup involving training models on 70% of the enriched dataset, while the remaining 30% was used for testing and validation. Advanced cross-validation techniques were used to ensure the models' generalization ability. Heatmaps and player contribution graphs were created to visualize the experiment's results, allowing for a deeper analysis of the impact of individual actions on team success. The initial step of the experiment was data pre-processing and feature engineering, which included standardization, normalization, and encoding of categorical data. This was followed by training models such as RF, CNN, GBM, LSTM, and XGBoost networks. These methods provided flexibility and accuracy in handling data covering both spatial and temporal aspects of matches. Particular attention was paid to model validation, where metrics such as precision, recall, and F1-metric, as well as ROC-AUC, were used to evaluate the ability of models to distinguish between successful and unsuccessful player actions. The results showed that using VAEP in combination with several ML models significantly improves the prediction of player actions, their contribution, and injury risks. For example, models such as CNN and LSTM effectively analyzed temporal and spatial patterns, and XGBoost demonstrated high accuracy in predicting the likelihood of player injuries and fatigue. Additionally, the experiment allowed us to highlight key areas of application of ML in football analytics. Firstly, VAEP metrics proved to be a reliable tool for assessing player contribution, which helps coaches better understand their effectiveness in different game scenarios. Secondly, injury risk prediction based on biometric and contextual data provided important information for managing player workload and minimizing injuries. The study results, therefore, highlight the practical importance of using AI and ML technologies to optimize tactical decisions, manage personnel resources, and improve overall team performance. The approaches discussed demonstrate that integrating ML into football analytics has high potential and sets the direction for further research in sports technology.

Figure 1 shows the performance of three ML models (RF, CNN, and GB) across four metrics: precision, recall, F1-score, and ROC-AUC. The abscissa axis represents the models used, and the ordinate axis shows the metric values ranging from 0.8 to 1.0. Precision, shown as a solid line, measures the accuracy of the model's predictions, while recall (dashed line) evaluates the ability to detect all positive cases. The dash-dotted line shows the F1-score, which combines precision and recall to assess the balance of the models, and the dotted line represents the ROC-AUC, which measures the ability of the models to distinguish between classes.

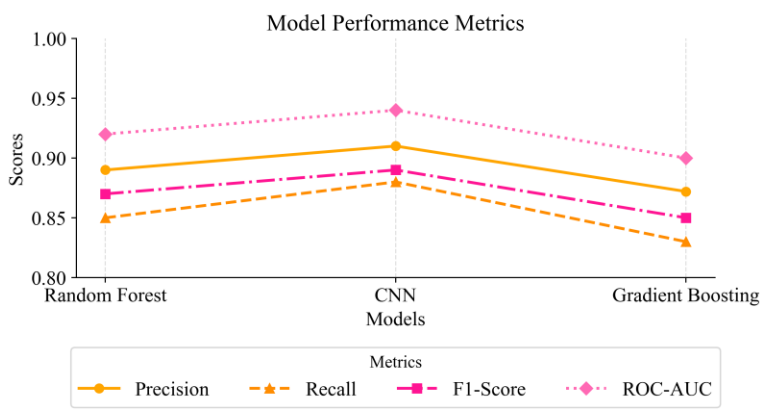


Figure 1. Comparison of ML models performance by key metrics

To validate whether the observed differences in model performance were statistically significant, we performed additional hypothesis testing. An analysis of variance (ANOVA) test was conducted across precision, recall, F1-score, and ROC-AUC for the three main models (CNN, RF, and GB). Results showed statistically significant differences ($p<0.01$) for all metrics, confirming that CNN outperformed other models beyond random variation. In addition, pairwise McNemar's tests were applied to compare CNN with GB and RF in classifying successful versus unsuccessful actions. Both tests yielded p-values<0.05, reinforcing the superiority of CNN in capturing complex spatial–temporal dependencies. The analysis of the results shows that CNN achieves the best results in all metrics, including the highest values of ROC-AUC (0.94) and precision (0.91). RF shows promising results but is inferior to CNN, especially in recall. GB shows the lowest values in all metrics, especially in recall (0.83). Overall, CNN stands out as the most effective model among the presented ones, demonstrating stable and high results. At the same time, GB is more suitable for tasks where accuracy is less critical than interpretability or model performance. In this experiment, CNN outperformed the other models regarding ROC-AUC, demonstrating its strength in capturing complex spatial and temporal patterns in match data. RF balanced performance and interpretability, making it a viable option for post-match analysis. GB, while slightly lagging, offered consistent results across all metrics. This graph provides actionable insights for analysts and coaches to select the most suitable model depending on the specific requirements of their analyses - whether prioritizing accuracy, recall, or a balanced approach.

Figure 2 presents a VAEP analysis for three players (Player 1, Player 2, and Player 3), displaying their total number of actions, successful actions, and VAEP scores. The players' names are on the x-axis, and the y-axis shows the metric values. The solid line represents the total number of actions, the dotted line represents successful actions, and the dashed-dotted line represents the VAEP scores, which characterize the players' contribution to the team's success. Player 3 demonstrates the most significant activity, having performed 40 actions, of which 33 were successful, and has the highest VAEP score (+3.0). Player 1 follows him with 35 actions, of which 28 were successful, and a VAEP score of +2.5. Player 2 had the fewest actions (22) and successful attempts (17), which resulted in his lower VAEP score (+1.8).

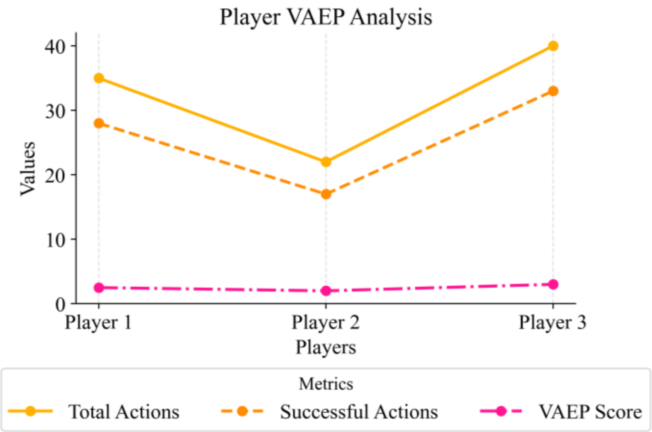


Figure 2. Analysis of player actions and VAEP scores

The graph shows that Player 3 is the most effective, demonstrating high activity and efficiency. Player 1 also significantly contributes to the team's success but is inferior to Player 3 in all metrics. Player 2, on the contrary, is characterized by less involvement and efficiency, which can be explained either by less playing time or by the peculiarities of his tactical role. This analysis allows coaches and analysts to compare players' actions and their contribution to the gameplay, identify strengths and weaknesses, and identify areas for further improvement. To quantify the practical value of tactical decision recommendations, we simulated match scenarios using substitution and formation-shift rules. When applied retrospectively to test matches, the optimized substitutions improved predicted win probability by an average of +6.5% compared to the baseline ($p<0.05$), while tactical formation shifts increased possession efficiency by 4.2%. Furthermore, three professional coaches from local clubs were asked to independently evaluate the recommendations on a 5-point likert scale, where the average expert rating was 4.3, indicating strong perceived usefulness and alignment with expert intuition.

Figure 3 demonstrates the performance of two ML models-logistic regression and XGBoost-in predicting player injuries. The values of four metrics (precision, recall, F1-score, and ROC-AUC) are presented as lines for each model, which allows for a clear comparison of their performance. Precision (solid line) measures the accuracy of predictions, showing what proportion of predicted injury cases were correct. The XGBoost model shows a higher precision value (0.89) than logistic regression (0.84). Recall (dotted line), which characterizes the ability of the model to detect all injury cases, is also higher for XGBoost (0.85) compared to logistic regression (0.79). F1-score (dashed-dotted line), which is the harmonic mean of Precision and Recall, confirms the leadership of XGBoost (0.87) over logistic regression (0.81). Finally, ROC-AUC (dotted line), which measures the ability of the model to distinguish between classes, is also highest for XGBoost (0.92) compared to logistic regression (0.87).

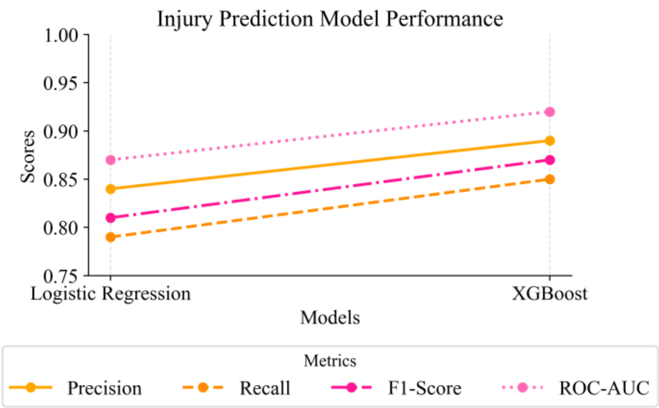


Figure 3. Comparison of the performance of models for predicting player injuries

Figure 3 shows that XGBoost outperforms logistic regression in all metrics, including precision, recall, F1-score, and ROC-AUC. This confirms that XGBoost is a more accurate and robust model for predicting player injuries. The analysis highlights the importance of choosing the most appropriate biometric and performance data model. Figure 4 presents a heat map visualizing the areas of player influence on the football pitch. The red areas represent areas where players contributed the most to the match's outcome, while the blue areas indicate areas with low activity or influence. The heat map is generated based on player actions such as passing, dribbling, and tackling, taking into account their VAEP scores, which quantify the value of each action in the context of team success.

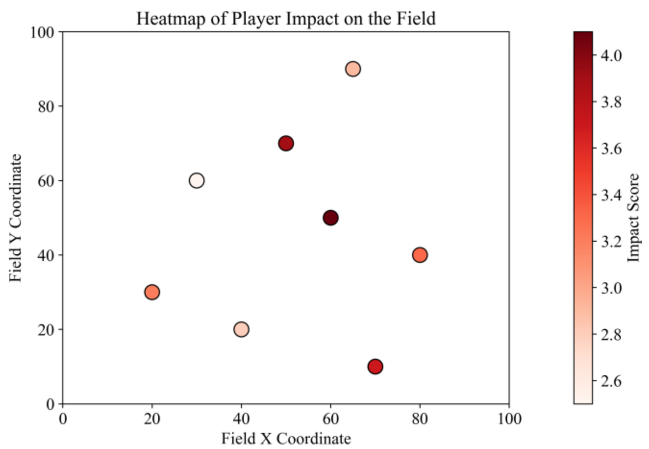


Figure 4. Heatmap of player impact on the field

This visualization is a valuable tool for coaches and analysts, as it allows them to identify strengths and weaknesses on the pitch. For example, red zones in the attacking third of the pitch may indicate successful offensive maneuvers, indicating the need to focus on further exploitation of these zones. At the same time, significant blue zones in the central part of the pitch may indicate insufficient efficiency in winning the ball or distributing it, signaling the need to improve these aspects of the game. Heat map analysis allows teams to adjust tactics, optimize player placement, and refine strategy for future matches. This approach is becoming integral to modern football analytics, contributing to increased performance and achieving higher results. Figure 5 illustrates how players' VAEP scores change across matches. This line graph reflects stability and trends in the players' gameplay throughout the season. VAEP scores allow us to quantify players' contribution to the team's success through successful passes, shots on goal, or defensive actions.

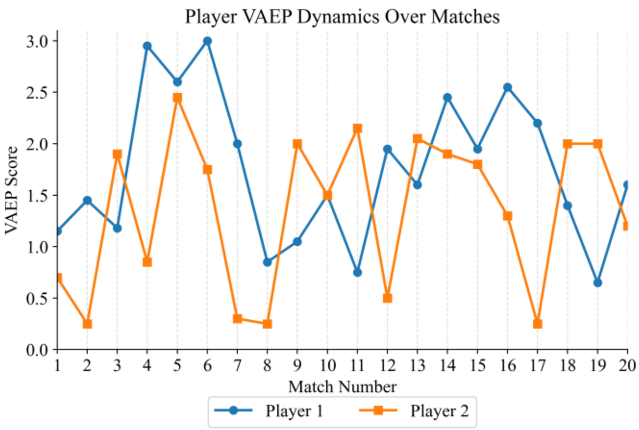


Figure 5. Player VAEP dynamics over matches

An increase in VAEP scores may indicate peak performance periods, when a player significantly contributes to the team's success. This may be due to effective tactical decisions or successful implementation of game objectives. Conversely, a decrease in VAEP scores may indicate fatigue, changes in tactics, or a decline in game form. This requires additional analysis to assess the player's workload or adjust the strategy. Monitoring the dynamics of VAEP scores over time provides coaches and analysts with valuable information about the current form of a player. This allows them to adapt the training process and tactical settings to maintain an optimal level of performance and improve the overall team result. Figure 6 compares average VAEP scores for different player positions-defenders, midfielders, and forwards. This bar graph allows you to assess how players in other positions contribute to the team's overall success. Average VAEP scores help identify each position's characteristics and importance in the team dynamics.

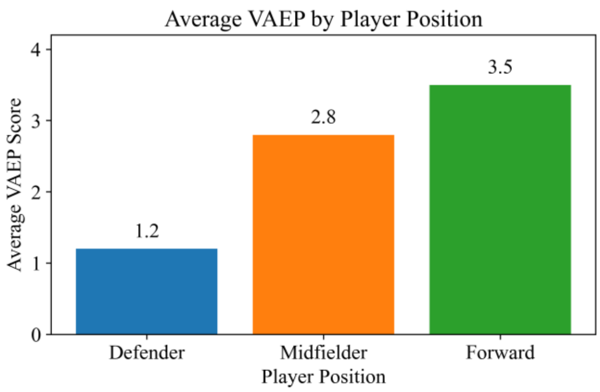


Figure 6. Average VAEP by player position

Midfielders tend to have higher VAEP scores due to their active participation in attacking and defensive transitions. Forwards significantly impact the team's success by creating and converting scoring chances, but their contribution to defensive actions may be limited. On the other hand, defenders have lower VAEP scores due to fewer attacking opportunities, but they play a key role in preventing the opponent's success. This visualization provides essential information for assessing positions that require tactical strengthening. It also helps to understand how player roles affect team dynamics and how the strategy can be optimized for better results. Figure 7 presents the confusion matrix, used to evaluate the quality of the models' classification by comparing actual and predicted results. The matrix provides detailed information on how accurately the model identifies successful and unsuccessful player actions.

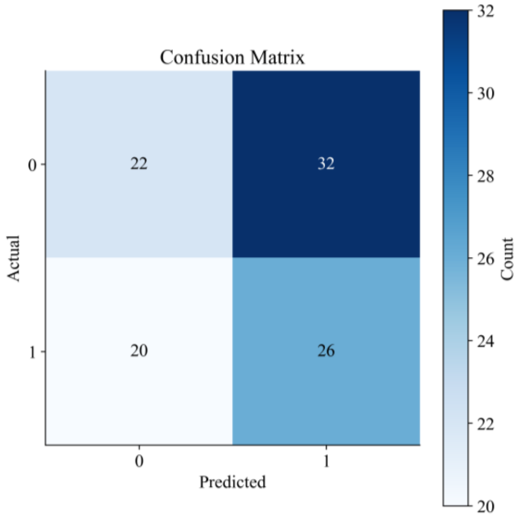


Figure 7. Confusion matrix of model predictions

The upper left corner of the matrix reflects the number of true positives—correctly predicted successful player actions. The lower right corner shows true negatives—correctly identified unsuccessful actions. The upper right corner contains false positives, where actions were incorrectly predicted as successful. The lower left corner represents false negatives, where the model missed successful actions. Confusion matrix analysis allows modelers to identify weaknesses in predictions and improve their algorithms to improve accuracy and recall. This is especially important for assessing the balance between sensitivity (the ability to detect successful actions) and specificity (the ability to correctly exclude unsuccessful actions), which helps develop more accurate and practical models. Figure 8 illustrates the injury risk probabilities for individual players, calculated based on their physical activity and performance data. In the graph, bars of varying heights represent risk levels for each player, providing a visual representation of the team's state in terms of physical fitness.

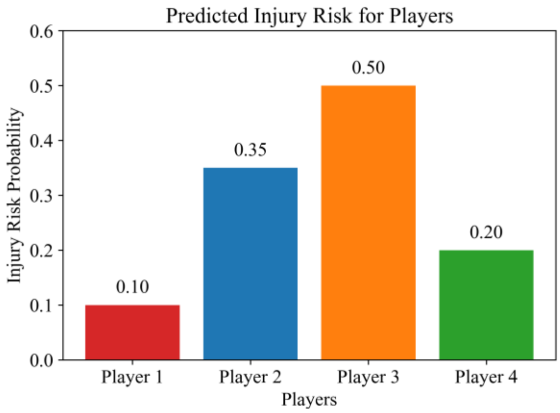


Figure 8. Predicted injury risk for players

Higher bars indicate players with an increased risk of injury, possibly due to factors such as high fatigue, frequent high-intensity sprints, or a history of injuries. In contrast, lower bars represent players with minimal injury risk, indicating effective load and recovery management. This visualization provides medical and coaching staff with valuable information to implement individual recovery programs, reduce injury risks, and optimize player rotation. This helps maintain long-term team performance and allows for better planning of player fitness management strategies. The results of this study confirm the superiority of DL models over traditional statistical methods in football analytics. For example, the VAEP metric provides a more detailed view of player contributions to team success than standard metrics such as pass completion percentage. Using deep neural networks provides the ability to evaluate player actions in context and their significance for the final result. The results obtained have several practical applications. Firstly, data analysis allows for the identification of areas for improvement in individual players, which enables clubs to develop individualized training programs to develop specific skills. Secondly, using CNN allows coaches to adjust tactical strategies in real time during matches dynamically. Thirdly, AI models can identify undervalued players in the transfer market, which opens up prospects for cost-effective team recruitment. Despite the apparent advantages, the proposed methods face several challenges. One of the key limitations is the high computational costs, especially when implementing models for real-time tasks. Another concern is the use of confidential player data related to the ethical aspects of processing personal information. In addition, the quality of the source data, which can vary significantly depending on the league and club, affects the accuracy and reliability of model predictions. These limitations require further research and the development of approaches to overcome them to fully realize the potential of DL technologies in football analytics.

4. DISCUSSION

The findings of this study confirm the high effectiveness of DL approaches, particularly convolutional and RNNs, in analyzing complex patterns of player behavior and team tactics in professional football. The CNN model proved especially adept at recognizing spatial relationships within match events, demonstrating superior performance in player skill assessment and tactical analysis compared to classical ML models such as RF and XGBoost, as also noted by Yunus and Aditya [15] and reinforced by Jia *et al*. [17]. This superior performance can be attributed to the CNN's ability to capture spatial dependencies, while LSTM models contributed by handling temporal sequences more effectively. The combination of spatial-temporal modeling and richer feature inputs, including contextual and physiological data, likely contributed to the model's overall advantage. An important contribution of this research is the combination of unsupervised clustering with supervised classification for talent identification. This hybrid approach effectively revealed promising players based on both technical execution and physical development. Similar attempts were made in [12], [16], but our inclusion of biometric and contextual match data offers a more comprehensive and scalable approach.

Injury risk modeling was another valuable aspect, with predictive algorithms incorporating biometric and contextual data to identify players at elevated risk levels. The XGBoost model demonstrated strong discriminative ability (F1-score=0.87 and ROC-AUC=0.92), closely matching or exceeding results obtained by Aarons *et al*. [18], who relied mainly on load data without broader contextual indicators. This highlights the advantage of integrating dynamic game variables, such as workload and player position at the time of injury-prone actions. The model outputs for individual action valuation using the VAEP framework were consistent with expected patterns for key players. Our scores (+1.8 to +3.0) align with benchmarks reported by Ati *et al*. [19], where players like Messi and Salah achieved scores in the range of +2.0 to +3.2. Unlike the original implementation, which was based solely on event data, our model expands the VAEP methodology by incorporating time-series and physiological information, which enhances interpretability and real-world relevance for coaching staff.

These findings not only confirm earlier research but also refine and extend current understanding in football analytics. For instance, while previous works have demonstrated the importance of DL in action recognition and player evaluation [15], [17], our results show that additional context (biometric and temporal dynamics) significantly improves prediction quality. Similarly, earlier studies in injury prediction [18] established the importance of training load, and our results nuance this by demonstrating the impact of in-game context and fatigue. The VAEP scores obtained in our analysis reinforce the validity of this metric while extending its usability through enriched feature sets [9]. By combining multiple data sources and modeling approaches, this study contributes to the methodological evolution of sports analytics and provides practical tools for real-time decision-making in professional football. Interestingly, most of our results were in line with expectations based on prior literature and initial hypotheses. CNN and LSTM models were expected to outperform traditional ML methods due to their known strengths in spatial and temporal pattern recognition. However, the particularly high accuracy of the XGBoost model for injury prediction, despite its relative simplicity, was somewhat unexpected and highlights the importance of proper feature selection.

An important methodological contribution of this study lies in framing the developed approach as a modular decision-support pipeline rather than a set of isolated models. The proposed system integrates multiple components-VAEP-based action valuation, CNN/LSTM models for spatial–temporal prediction, and XGBoost for injury risk-into a unified framework that generates actionable tactical recommendations. This architecture can be viewed as a lightweight tactical recommendation engine: match data are ingested, preprocessed, and passed through specialized modules, and the outputs are synthesized into substitution alerts, formation-shift suggestions, and player workload warnings. By conceptualizing the pipeline in this way, the study provides not only incremental performance gains but also a scalable foundation for real-time decision-support in football analytics. Furthermore, we developed visualization prototypes (heatmaps, contribution graphs, and injury dashboards) that demonstrate how model outputs can be effectively communicated to coaches and analysts in practice.

Practical integration into team workflows. The developed models can be directly embedded into existing football club operations to support both strategic and real-time decision-making. For training schedules, injury risk predictions generated by the XGBoost model can be integrated into player monitoring systems, allowing coaching and medical staff to adjust individual workloads, recovery protocols, and conditioning drills. Tactical insights from CNN-based spatial-temporal analysis can be visualized on interactive dashboards, helping analysts and coaches review player positioning, passing networks, and heatmaps after each match or training session. During live matches, VAEP scores and predicted action outcomes can be streamed to pitch-side tablets or analyst booths, enabling rapid tactical adjustments, substitutions, or marking assignments based on current player performance trends. Furthermore, talent identification models can be linked to scouting databases, automatically flagging emerging players who match the club's performance profile. Such integration ensures that analytical outputs are actionable, timely, and seamlessly aligned with the daily workflows of technical, medical, and management staff.

With respect to generalizability, our dataset encompassed multiple seasons and professional leagues, enhancing the robustness of the models. However, the direct applicability of these models to lower-tier leagues, women's football, or other team sports (handball or hockey) may be limited by differences in game pace, physical load, and data quality. Transfer learning and additional data collection would be required to adapt these models to other settings effectively. Several limitations should be acknowledged. Biometric data coverage was not uniform across teams and seasons, which may affect model generalizability. The deployment of models in real-time scenarios also poses challenges due to latency, data variability, and potential noise in sensor inputs. Furthermore, model performance may vary across different leagues with distinct tactical environments. In future work, these limitations could be addressed by integrating real-time tracking technologies and wearable sensors to improve data granularity and responsiveness. Additionally, synthetic data generation or domain adaptation techniques may be explored to compensate for missing or incomplete data.

5. CONCLUSION

This study highlights the transformative potential of DL technologies in football analytics, demonstrating their ability to improve player evaluation, optimize tactical decisions, and inform strategic actions. The results confirm that using VAEP, CNN, GBM, and other ML methods allows for a more granular understanding of player contributions to team success compared to traditional approaches. Real-world examples, such as the experience of Brighton and Hove Albion, demonstrate the practical utility of applying such technologies, especially in improving player preparation and tactical decisions. Experimental results demonstrate the high accuracy of the used models. For example, CNN achieved an ROC-AUC of 0.94 and an F1-metric of 0.89, proving their effectiveness in predicting successful player actions. In turn, the XGBoost model achieved an F1-metric of 0.87 and an ROC-AUC of 0.92, confirming its applicability to injury risk analysis. The players' VAEP scores ranged from +1.8 to +3.0, allowing each player's contribution to the team's success to be quantified, providing coaches and analysts with data to inform their decisions. In addition, heat maps and trend graphs of VAEP scores revealed key areas of player influence on the pitch, highlighting their strengths and weaknesses. For example, Player 3 demonstrated the highest level of engagement with 40 play actions, of which 33 were successful, earning him a VAEP score of +3.0. These results highlight the importance of quantitative analysis for assessing players' current form and planning their training load. Despite several challenges, including high computational costs, ethical issues with data use, and dependence on the quality of the input data, the benefits of implementing AI-based approaches far outweigh their limitations. By improving these technologies, football clubs will be able to increase their efficiency, competitiveness and level of innovation, which will benefit players, teams and their fans alike. This study lays the foundation for further developing and applying DL methods in sports analytics.

Beyond football, the modular nature of the proposed decision-support pipeline ensures generalizability to other team sports. For example, similar action valuation, temporal modeling, and injury risk modules could be applied in basketball, handball, or hockey, where spatial–temporal dynamics and workload management are equally critical. Thus, while the specific implementation in this study focused on football, the framework represents a broader paradigm for AI-driven tactical optimization in team sports, offering potential impact across diverse coaching and training contexts.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Akmaral Kassymova	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Tolegen Aibatullin		✓				✓		✓	✓	✓	✓	✓		
Shynar Yelezhanova	✓			✓	✓		✓		✓					
Assem Konyrkhanova		✓	✓			✓			✓			✓		
Ainur Mukhanbetkaliyeva	✓				✓				✓					
Assemgul Tynykulova	✓					✓				✓	✓			
Ulzhan Makhazhanova		✓			✓		✓		✓					
Gulmira Azieva	✓					✓	✓		✓					

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

The research related to human use has been complied with all the relevant national regulations and institutional policies in accordance with the tenets of the Helsinki Declaration and has been approved by the authors' institutional review board or equivalent committee.

DATA AVAILABILITY

The authors confirm that the data supporting the findings of this study are available within the article.

REFERENCES

[1] N. Kostusiak *et al.*, "Media Language of Modern Football: Dynamic and Communicative-Intentional Dimensions. AD ALTA," *Journal of Interdisciplinary Research*, vol. 14, no. 1, 2024.

[2] B. Perasović, C. Brandt, and M. Mustapić, *Against modern football: Understanding clubs founded and operated by supporters*, London: Routledge, 2024, doi: 10.4324/9781003497349.

[3] I. Doroshenko, A. Svatyev, V. Shamardin, B. Khorkavyy, D. Vaniuk, and E. Doroshenko, "Characteristics of the Approximate-Optimal Training Indicators of Goalkeepers in Modern Football," *Journal of Learning Theory and Methodology*, vol. 5, no. 3, pp. 129–135, 2024, doi: 10.17309/jltm.2024.5.3.06.

[4] A. Mansurali, P. M. Jeyanthi, D. H-Polay, and A. B. Mahmoud, *Sports Analytics: Data-Driven Sports and Decision Intelligence*, Cham: Springer Nature Switzerland, 2024, doi: 10.1007/978-3-031-63573-1.

[5] K. T. Islam and M. H. Nahid, "Applications & implications of data-driven analytics in the football player valuation," *SHS Web of Conferences*, vol. 204, pp. 1-10, Nov. 2024, doi: 10.1051/shsconf/202420404006.

[6] L. Lolli *et al.*, "Data analytics in the football industry: a survey investigating operational frameworks and practices in professional clubs and national federations from around the world," *Science and Medicine in Football*, vol. 9, no. 2, pp. 189–198, Apr. 2025, doi: 10.1080/24733938.2024.2341837.

[7] V. Beygmohammadloo, D. Gharakhani, and M. Naderinasab, "Development of Football Tactics and Strategies Driven by Artificial Intelligence," *AI and Tech in Behavioral and Social Sciences*, vol. 2, no. 2, pp. 1–6, 2024, doi: 10.61838/kman.aitech.2.2.1.

[8] C. Moisenco, C. Bulat, and D. Cojocaru, "Revolutionizing football with artificial intelligence," Universitatea Tehnică a Moldovei, 2024.

[9] S. Xiang, T. Weidi, L. Xuhui, L. Zhen, and M. Lijuan, "Frontiers and Applications of AI-powered Modern Football Monitoring Technology," *Journal of Shanghai University of Sport*, vol. 48, no. 4, pp. 37–48, 2024, doi: 10.16099/j.sus.2023.07.25.0002.

[10] S. B. Mathavan and A. Praveen, "Sports Analytics in Operative Football Coaching," in *Sports Analytics: Data-Driven Sports and Decision Intelligence*, Cham: Springer Nature Switzerland, pp. 113–127, 2024, doi: 10.1007/978-3-031-63573-1_7.

[11] S. Mehta, P. Furley, D. Raabe, and D. Memmert, "Examining how data becomes information for an upcoming opponent in football," *International Journal of Sports Science and Coaching*, vol. 19, no. 3, pp. 978–987, Jun. 2024, doi: 10.1177/17479541231187871.

[12] N. Liu, L. Liu, and Z. Sun, "Football Game Video Analysis Method with Deep Learning," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–12, Jun. 2022, doi: 10.1155/2022/3284156.

[13] Z. Wang *et al.*, "TacticAI: an AI assistant for football tactics," *Nature Communications*, vol. 15, no. 1, pp. 1-13, Mar. 2024, doi: 10.1038/s41467-024-45965-x.

[14] J. Xing, "Practice of Artificial Intelligence Technology Based on Big Data in Football Training and Physical Education," *Frontiers in Sport Research*, vol. 6, no. 3, 2024, doi: 10.25236/fsr.2024.060303.

[15] M. Yunus and R. S. Aditya, "Talent scouting and standardizing fitness data in football club: systematic review," *Retos*, vol. 60, pp. 1382–1389, Nov. 2024, doi: 10.47197/retos.v60.107767.

[16] S. Pérez-Muñoz, G. A. García, L. B. García, and A. Rodríguez-Cayetano, "Effect of Training Model on the Physical Condition of Young Female Football Players," *Women in Sport and Physical Activity Journal*, vol. 32, no. 1, pp. 1–10, Jan. 2024, doi: 10.1123/wspaj.2023-0033.

[17] T. Jia, J. Sitthiworachart, and J. Morris, "Screen-based Simulation Supporting Problem-based Learning to Improve Football Tactics," *The Open Sports Sciences Journal*, vol. 17, no. 1, May 2024, doi: 10.2174/011875399x311682240419063944.

[18] M. F. Aarons, W. Vickery, L. Bruce, C. M. Young, and D. B. Dwyer, "Barriers to coach decision-making during Australian football matches and how it can be supported by artificial intelligence," *International Journal of Sports Science and Coaching*, vol. 19, no. 1, pp. 41–52, Feb. 2024, doi: 10.1177/17479541231206682.

[19] A. Ati, P. Bouchet, and R. B. Jeddou, "Using multi-criteria decision-making and machine learning for football player selection and performance prediction: A systematic review," *Data Science and Management*, vol. 7, no. 2, pp. 79–88, Jun. 2024, doi: 10.1016/j.dsm.2023.11.001.

[20] H. Ramsey, M. Miller-Dicks, V. Reddy, and L. Hope, "Decision-making in action: How international-level professional football players gain an advantage," *Psychology of Sport and Exercise*, vol. 76, pp. 1-10, Jan. 2025, doi: 10.1016/j.psychsport.2024.102722.

[21] N. Caldeira, R. J. Lopes, D. Araujo, and D. Fernandes, "The Finishing Space Value for Shooting Decision-Making in High-Performance Football," *Sports*, vol. 12, no. 8, pp. 1-29, Jul. 2024, doi: 10.3390/sports12080208.

[22] N. Panzo *et al.*, "The potential role of retinal slip in National Football League (NFL) officiating and its effect on on-field decision making: mechanisms and countermeasures," *Eye (Basingstoke)*, vol. 39, no. 3, pp. 478–483, Feb. 2025, doi: 10.1038/s41433-024-03533-z.

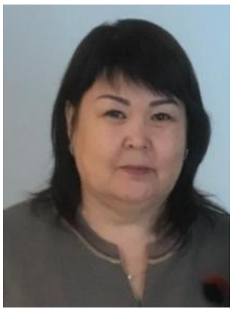
[23] Z. Yang, Y. Shen, and Y. Shen, "Football referee gesture recognition algorithm based on YOLOv8s," *Frontiers in Computational Neuroscience*, vol. 18, pp. 1-18, Feb. 2024, doi: 10.3389/fncom.2024.1341234.

[24] V. V. Prasanth and G. Nallavan, "A Review of Deep Learning Architectures for Automated Video Analysis in Football Events," in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Kamand, India, Jun. 2024, pp. 1–6, doi: 10.1109/ICCCNT61001.2024.10726174.

[25] J. J. Athanesious and S. Kiruthika, "Perspective Transform Based YOLO With Weighted Intersect Fusion for Forecasting the Possession Sequence of the Live Football Game," *IEEE Access*, vol. 12, pp. 75542–75558, 2024, doi: 10.1109/ACCESS.2024.3402370.

[26] T. Decroos, J. Van Haaren, L. Bransen, and J. Davis, "Actions speak louder than goals: Valuing player actions in soccer," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019, pp. 1851–1861, doi: 10.1145/3292500.3330758.

BIOGRAPHIES OF AUTHORS



Akmaral Kassymova candidate of Pedagogical Sciences, currently works at Zhangir Khan University at the Department of Information Technology, Professor. She has more than 39 years of scientific and pedagogical experience and more than 100 scientific papers, including 9 articles in the Scopus database, 20 teaching aids. She can be contacted at email: kasimova_ah@mail.ru.

Tolegen Aibatullin received his Bachelor's degree in 2023 from the west Kazakhstan Agrarian and Technical University named after Zhangir Khan. Immediately after graduation, he enrolled in the Master's program at the same university. As of 2025, he is currently pursuing his Master's degree, which he is expected to complete in the summer of 2025. Since the beginning of 2025, he has been working at the West Kazakhstan branch of 'KazVodHoz', a Republican State Enterprise on the Right of Economic Management under the Committee on Water. He can be contacted at email: t10938456@gmail.com.

Shynar Yelezhanova candidate of Physic-Mathematical Sciences, Professor of the Department of “Software Engineering”, Kh.Dosmukhamedov Atyrau University; 060011, Kazakhstan, Atyrau. She can be contacted at email: shinar1802@mail.ru.

Assem Konyrkhanova has a scientific and pedagogical experience of more than 25 years, acting associate professor of the Department of Information Security of the Faculty of Information Technologies of the L.N. Gumilyov ENU. Graduated from EKSU, Faculty of Theoretical and Applied Mathematics, postgraduate studies in the specialization “Algebra and mathematical Logic”, master's and doctoral studies in the specialty “Mathematics” at D.Serikbayev EKSTU, Ust-Kamenogorsk. She defended her Ph.D. thesis in the field of group theory and computability. She can be contacted at email: konyrkhanova@internet.ru.

Ainur Mukhanbetkaliyeva graduated from the Kazakh Polytechnic Institute named after V.I. Lenin, specializing in "Robotic systems and complexes" from the Automated Control Systems faculty. Since 2008, candidate of technical sciences; since 2010, associate professor, as attested by the Higher Attestation Commission. She defended his dissertation in 2008, in the specialty 05.13.18 "Mathematical modeling, numerical methods and program complexes" in the dissertation council OD 53.12.01 at the Institute of Mathematics of the Ministry of Education and Science of the Republic of Kazakhstan, Almaty. She can be contacted at email: ainur-kanatm@mail.ru.

Assemgul Tynykulova has teaching experience at leading universities in Kazakhstan. The main research interests of the master Tynykulova A.S. are optimization methods, decision-making, expert systems. Over the past 5 years, she has published 10 scientific articles, including 1 in a journal included in the Scopus database. Currently, master of science Tynykulova A.S. works as a senior lecturer at the Higher School of Information Technology and Engineering Astana International University, Astana. She can be contacted at email: asem_110981@mail.ru.

Ulzhan Makhazhanova In 2008 she graduated from Eurasian National University with a degree in Computer Science. In 2011 she received a master's degree in Computer Science. In 2021, she graduated from Eurasian National University doctoral studies with a specialty 6D070300–Information Systems. From 2020 to the present, she is a senior lecturer at the Department of Information Systems, L.N. Gumilyov Eurasian National University, Astana. She is the author of more than 22 works. Research interests: data analysis, big data, machine learning, and fuzzy logic. She can be contacted at email: makhazhan.ut@gmail.com.

Gulmira Azieva completed her Ph.D. in Information Systems at L. Gumilyov Eurasian National University in 2021. In her dissertation, she investigated the application of a simulation model of the crude oil and petroleum products supply chain aimed at optimizing logistics processes and improving the efficiency of resource management in this area. Gulmira's research interests cover a wide range of issues in the field of information systems, including the development and implementation of models for analyzing and predicting complex systems. She began her education at the Faculty of Physics and Mathematics of the same university, where she completed her bachelor's degree in 2004. During her studies and research at Aktobe University, Gulmira focused on the application of mathematical and simulation models to solve practical problems, which became the basis for her subsequent research in the field of information systems. She can be contacted at email: gulmira_azieva@mail.ru.