

# Sports Prediction Model through Cloud Computing and Big Data Based on Artificial Intelligence Method

# Aws I. Abu Eid<sup>1\*</sup>, Achraf Ben Miled<sup>2,3</sup>, Ahlem Fatnassi<sup>2</sup>, Majid A. Nawaz<sup>2</sup>, Ashraf F. A. Mahmoud<sup>2</sup>, Faroug A. Abdalla<sup>2</sup>, Chams Jabnoun<sup>2</sup>, Aida Dhibi<sup>2</sup>, Firas M. Allan<sup>2</sup>, Mohammed Ahmed Elhossiny<sup>4,5</sup>, Salem Belhaj<sup>2</sup>, Imen Ben Mohamed<sup>2</sup>

<sup>1</sup>Faculty of Computing Studies, Arab Open University, Amman, Jordan

<sup>2</sup>Computer Science Department, Science College, Northern Border University, Arar, Kingdom of Saudi Arabia

<sup>3</sup>Artificial Intelligence and Data Engineering Laboratory, LR21ES23, Faculty of Sciences of Bizerte, University of Carthage, Tunis, Tunisia

<sup>4</sup>Applied College, Northern Border University, Arar, Kingdom of Saudi Arabia

<sup>5</sup>Faculty of Specific Education, Mansoura University, Mansoura, Egypt

Email: \*a\_abueid@aou.edu.jo

How to cite this paper: Eid, A.I.A., Miled, A.B., Fatnassi, A., Nawaz, M.A., Mahmoud, A.F.A., Abdalla, F.A., Jabnoun, C., Dhibi, A., Allan, F.M., Elhossiny, M.A., Belhaj, S. and Mohamed, I.B. (2024) Sports Prediction Model through Cloud Computing and Big Data Based on Artificial Intelligence Method. *Journal of Intelligent Learning Systems and Applications*, **16**, 53-79. https://doi.org/10.4236/jilsa.2024.162005

**Received:** December 4, 2023 **Accepted:** March 26, 2024 **Published:** March 29, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). http://creativecommons.org/licenses/by/4.0/

c 0 Open Access

# Abstract

This article delves into the intricate relationship between big data, cloud computing, and artificial intelligence, shedding light on their fundamental attributes and interdependence. It explores the seamless amalgamation of AI methodologies within cloud computing and big data analytics, encompassing the development of a cloud computing framework built on the robust foundation of the Hadoop platform, enriched by AI learning algorithms. Additionally, it examines the creation of a predictive model empowered by tailored artificial intelligence techniques. Rigorous simulations are conducted to extract valuable insights, facilitating method evaluation and performance assessment, all within the dynamic Hadoop environment, thereby reaffirming the precision of the proposed approach. The results and analysis section reveals compelling findings derived from comprehensive simulations within the Hadoop environment. These outcomes demonstrate the efficacy of the Sport AI Model (SAIM) framework in enhancing the accuracy of sports-related outcome predictions. Through meticulous mathematical analyses and performance assessments, integrating AI with big data emerges as a powerful tool for optimizing decision-making in sports. The discussion section extends the implications of these results, highlighting the potential for SAIM to revolutionize sports forecasting, strategic planning, and performance optimization for players and coaches. The combination of big data, cloud computing, and AI offers a promising avenue for future advancements in sports analytics. This research underscores the synergy between these technologies and paves the way for innovative approaches to sports-related decision-making and performance enhancement.

### **Keywords**

Artificial Intelligence, Machine Learning, Spark Apache, Big Data, SAIM

## **1. Introduction**

The exponential surge in data generation has ushered in an era where the amount of new information being produced daily dwarfs the entire data amassed in the past. A mere glimpse into contemporary data proliferation reveals that what was once accumulated over two decades now materializes within the confines of a single minute. In this era of information deluge, the challenges of effectively managing, processing, and deriving value from this mammoth influx of data have taken center stage. The terminology "big data" has been coined to encapsulate these vast datasets, which may exhibit structured characteristics akin to traditional relational databases or unstructured attributes akin to the dynamic content of social media platforms and organizational repositories [1]. At the heart of the concept lies the symbiotic relationship between intensive applications and the colossal datasets that drive them. This interplay has propelled researchers, engineers, and innovators to devise novel strategies to harness the potential encapsulated within big data. As data repositories swell to unprecedented scales, the confluence of cutting-edge technologies becomes imperative to unveil hidden patterns, correlations, and insights that can inform critical decision-making processes [2]. In this context, cloud computing and big data convergence emerge as a pivotal paradigms. The scalable and elastic nature of cloud computing resources provides a viable solution to the computational demands posed by big data applications.

Moreover, the fusion of artificial intelligence (AI) techniques with cloud computing and big data analytics presents an intriguing prospect, promising enhanced predictive capabilities and deeper contextual understanding. Hence, cloud computing can be conceptualized as an internet-centric framework wherein computing resources, encompassing databases, networks, software applications, storage, hardware, and even integral operating systems within information technology (IT) services, stand readily accessible to end-users precisely when needed. Although it doesn't rely on ground-breaking innovations, cloud computing significantly enhances the efficacy of IT services [3]. Particularly noteworthy is the indispensable role of data distribution and storage availability, particularly in contexts characterized by data-intensive and processor-intensive environments, acting as linchpins for achieving remarkable performance levels. The synergy between the Hadoop system, cloud computing, and the realm of big data is a pivotal nexus driving modern data analytics paradigms. Hadoop, an open-source framework designed to process and manage massive datasets, finds a natural alliance with cloud computing resources' scalable and flexible nature. This alliance enables organizations to harness the power of distributed computing, exploiting cloud environments' extensive storage and computational capabilities. As big data applications demand substantial computational resources that can scale dynamically, the cloud-Hadoop amalgamation offers an optimal solution [4]. Cloud computing's ability to allocate resources as needed aligns seamlessly with Hadoop's distributed processing, facilitating the efficient handling of vast datasets [5]. This confluence enhances processing speed and introduces cost-effectiveness by optimizing resource utilization. Thus, the Hadoop-cloud nexus exemplifies the evolving landscape where cloud-based infrastructure propels the potential of big data analytics to new frontiers, fostering insights and innovation across diverse domains. This study strongly emphasizes Machine Learning (ML) techniques, specifically neural networks, to address the challenge at hand, predicting patterns. In this context, the neural network (NN) paradigm emerges as a powerful tool for time-based predictions of movement dynamics. The utilization of neural networks in movement prediction has gained substantial attention due to their capacity to capture complex temporal relationships and intricate patterns within the data. These networks demonstrate the ability to learn and generalize from large datasets, contributing to enhanced prediction accuracy and adaptability. Integrating neural networks within the ML framework empowers the exploration of movement prediction problems with heightened sophistication. As explored by previous research [6], the NN-time paradigm holds promise for unraveling nuanced movement patterns and substantially enriching predictive capabilities in sports. Our main contribution is the Model SAIM on the Hadoop platform for predicting sports results, which is an integrated learning algorithm and comprehensive.

## 2. Literature Review

The landscape of machine learning has experienced a notable surge in the use of Artificial Neural Networks (ANNs) to predict diverse outcomes within various domains. This section focuses on elucidating the expanding role of ANNs specifically within the realm of sports outcome prediction

The application of Artificial Neural Networks (ANNs) has emerged as a ubiquitous approach within the domain of machine learning (ML) for predicting effects [7]. Consequently, this review centers on exploring the utilization of ANNs in the context of sports outcome prediction. ANNs transform input values into a desired output through interconnected processing units or neurons. The strength of ANNs resides in the nuanced regulation of neurotransmitters, which collectively contribute to the outcome, guided by weights assigned to connections. The output of an ANN is contingent upon input strength and other intrinsic network elements, notably these connection weights. A pivotal application of ANNs has been crafting prototype classification models, effectively using a tutorial database to establish and fine-tune the ANN model's functionality. A continual interplay of adjustments is undertaken to optimize forecast accuracy by optimizing the interneuron connection weights. These alterations exemplify the neural network algorithm's adaptation process to meet the predefined accuracy criterion set by the user [8].

The significance of accumulating computing resources extends beyond mere accessibility, encompassing the imperative to avoid wasting time and memory during training. One intriguing facet of artificial neural networks is their adaptability in handling class variables for diverse objectives within distinct categories. This adaptability becomes evident when considering such flexible categorizations based on the probability of success or failure, with potential variations defined by two classes [9]. These classes serve as indicators of the target being utilized. The authors Matthew Dzwil, Mia Hopman, and Katie Houskeeper in [10] discussed in Player Ratings within the NFL to improve innovation in the major playoff draft and based the focal point of this endeavor on turning positional tracking data from wide receivers during the 2022 NFL Scouting Combine into meaningful and insightful metrics. A systematic approach to data pre-processing that uses specific threshold parameters to characterize and segment workouts was developed. It also expanded the trajectories observed throughout the 2022-2023 NFL season. Metrics were created that effectively summarized the players' complex movements during training and trajectories and culminated in the project with the initiation of the trajectory prediction model, which aims to uncover the underlying factors contributing to players' performance on the pitch.

A pioneering study led by Nguyen Hoang Nguyen [11] illustrated the deployment of neural network models in predicting outcomes within the National Football League (NFL) context. Consider competition data from the initial eight rounds, possession duration, and rushing yardage to extract insights from five pivotal attributes. The study sought to distinguish high-performing from low-performing teams, employing an approach characterized by its package-oriented methodology. Regression Analysis yielded an RMSE of 2.1969 and MAE of 1.6465, while Classification Analysis displayed a Recall of 0.9368 and ROC AUC of 0.9152. Furthermore, authors' exclusive reliance on basketball statistics, the potential influence of external factors on popularity might be underappreciated.

Yao, P [12] delved deeper into the use of deep learning in basketball, using a comprehensive set of research methodologies, including literature analysis, video examination, comparative exploration, and mathematical statistics, to examine the integration of deep learning into the real-time analysis of basketball data. The system worked on recognizing and analyzing the movement of the proposed basketball position in two consecutive parts. The first stage used a bottom-up posture estimation approach to determine the joint coordinates and thus extract the target's posture sequence from the video footage. In the later stage, he presented an in-depth analysis of the action recognition algorithm based on the convolution of the space-time graph while harnessing the extracted posture sequences to classify basketball moves. Improved accuracy and 3D insights were also achieved using a multi-objective training approach. The integration of big

data has transformed strategies and performance analysis by analyzing the behavioral trajectories of opponents and players through big data and developing individual strategies tailored to enhance individual attacking prowess. Moreover, he discovered the tactical methods of other teams and the players' ideal formations, thus enhancing the formulation of offensive and defensive strategies. Indeed, the study showed remarkable developments in the technical and tactical analysis of professional basketball games using deep learning techniques.

Moreover, in basketball games, Zhu et al. [13] have harnessed the potential of big data. The integration of expansive full-angle replay technology has amplified the visual engagement for fans, concurrently opening avenues for leveraging big data's capabilities. This convergence has improved training quality, player valuation, and game equity. However, the essence of big data's contribution remains supplementary, ultimately necessitating the discernment of referees for decisive judgments [14]. Similarly, Shan et al. underscore that big data's applications encompass the comprehensive collection and analysis of draft-related metrics such as player height, bounce, and sprinting explosive power. The derived insights facilitate the holistic evaluation of players and the prognostication of their rankings. Many high-ranking predictions tend to materialize into successful players. Relative to this, the domestic league's data collection might exhibit limitations regarding precision and favorability toward certain player profiles. A synergistic merger of basketball and data becomes essential to optimize outcomes, culminating in a comprehensive data analysis system that substantiates meticulous player selection [15]. Deepening the exploration, Liu G et al. delve into real-time deep learning analysis of basketball sports data, elucidating the transformative role of data completeness, accuracy, and systematization in the current big data era. The holistic cultivation of data management practices in basketball games augments data-driven managerial strategies and is an indirect precursor for the sport's future development trajectory [16]. Thus, establishing a robust foundation in information technology data management, coupled with the adept utilization of relevant data collection tools and analytical software, becomes pivotal for ushering in a new era of basketball development [17].

The exploration and simulation of the Hadoop platform have attracted considerable scholarly attention, as evidenced by numerous studies. Notably, in the works of [18] [19] [20], diverse implementations have been elucidated to profile and predict crucial system metrics through the leverage of the Hadoop framework. A prevalent trend in these investigations involves applying Machine Learning (ML) techniques to facilitate analysis and performance forecasting. ML effectively establishes a connection between workload parameters and performance metrics, enabling predictive insights. This gamut of predictive models spans an array of ML algorithms, encompassing linear regression and the sophistication of neural networks.

## 3. The Methodological Approach

This section proposes a strategy for distributed cluster data processing and ma-

chine learning using Hadoop MapReduce, Mahout, and Spark. The key aspects include:

1) Task Fragmentation:

Divide complex tasks into independent sub-problems with clear boundaries.

Each sub-problem is processed individually on a distributed cluster node.

This approach ensures data block autonomy and avoids dependencies between nodes.

2) Optimized Data Processing:

Utilize Hadoop MapReduce and Mahout for efficient data processing.

Optimize disk storage for efficient data handling.

3) Spark Integration:

Integrate Spark and Hadoop platforms for enhanced memory loading and query facilitation.

This combined platform provides a powerful environment for machine learning algorithms.

4) Methodology:

The methodology includes sections on objectives, tool utilization, challenge and solutions, and visual representation.

The focus is on utilizing Spark's memory-based processing for superior performance compared to disk-based Mahout.

Overall, the strategy aims to leverage the strengths of different technologies to achieve efficient and scalable distributed data processing and machine learning. This amalgamation is visualized in the Hadoop platform circuit function diagram (Figure 1).

**Figure 1** shows the architecture of a data warehouse within the Hadoop ecosystem, which comprises a structured framework that leverages various components to enable efficient storage, processing, and analysis of large volumes of structured and unstructured data. The Hadoop ecosystem is a collection of open-source tools and technologies that collectively enhance the capabilities of Hadoop.

The research's focal point gravitates towards emerging technologies catering to the multifaceted challenges of big data, particularly in steering the trajectory of big data analysis. Amid the evolution transpiring within the Apache Open Source segment, advanced explorations encompass modeling, regression, and latent learning on expansive big data within Hadoop's Director, predicting product numbers and fostering burgeoning startups. In this context, Spark is a zone that underpins support for substantial data clusters. It accommodates high-level data storage for swift database retrieval, featuring an integrated learning machine-learning algorithm rooted in group learning principles.

Apache Spark is an open-source, distributed computing platform facilitating rapid and efficient large-scale data processing. Renowned for its speed, versatility, and ease of use, Spark is designed to handle diverse data processing workloads, including batch processing, interactive queries, machine learning, and real-time stream processing. **Figure 2** shows an overview of the Apache Spark platform.



Figure 1. An overview of the architecture of data warehouse Hadoop ecosystem components.



Figure 2. An overview of the Apache Spark platform.

Apache Spark's versatility, speed, and comprehensive ecosystem have made it a preferred choice for organizations seeking to process and analyze large volumes of data efficiently, derive meaningful insights, and build advanced machine-learning models—all within a single unified platform integrated by the Hadoop platform. Within The Methodological Approach Presented, the utilization of the Spark platform assumes a pivotal role in partitioning the dataset into distinct segments, each serving a distinct purpose. Firstly, the monitored aspect involves supervised learning techniques, focusing on high-quality and resilient data prediction through rigorous training. On the other hand, the unattended learning facet caters to predictions based on less robust data. The dataset in question pertains to the outcomes of the English Football League across the year 2021, encompassing 540 players and 20 teams. By harnessing the capabilities of Spark, this segmentation facilitates a nuanced approach to data analysis, enabling tailored methodologies for predictions and insights derivation.

The training database, denoted as T, encapsulates a comprehensive data repository from the English Football League. This database's construction of data packages is pivotal, forming the foundation for subsequent analysis and predictive modeling. A prudent strategy is employed to mitigate the effects of variability inherent in sports data: the aggregation of multiple scenarios. This technique involves averaging together diverse scenarios, enhancing the robustness of the predictions and insights derived from the training database. By amalgamating various scenarios, the potential influence of outliers or anomalous occurrences is diminished, enabling a more accurate and representative understanding of the underlying patterns and trends within the English Football League dataset. The training database, denoted as T, encapsulates a comprehensive data repository from the English Football League. This database's construction of data packages is pivotal, forming the foundation for subsequent analysis and predictive modeling. A prudent strategy is employed to mitigate the effects of variability inherent in sports data: the aggregation of multiple scenarios. This technique involves averaging together diverse scenarios, enhancing the robustness of the predictions and insights derived from the training database; this iterative process leads to the determination of M. By amalgamating various scenarios, the potential influence of outliers or anomalous occurrences is diminished, enabling a more accurate and representative understanding of the underlying patterns and trends within the English Football League dataset.

$$f(x) = 1/M \sum_{M=1}^{M} (fm(X))$$
(1)

The described approach employs a bootstrap model to extract data subgroups intended for learners within training platforms, specifically designed to accommodate package architecture. This methodology is a conduit for amalgamating essential learning outcomes, facilitating classifying applications suited for pocket structure.

#### **3.1. Data Extraction**

Analyzing the English Football League through extensive big data is an intricate endeavor with immense potential. Its core focus lies in examining player performance during matches, match outcomes, and the progression of rounds to anticipate and formulate corresponding strategies. It requires dissecting player behavior during games to forecast their future actions. It also entails thoroughly analyzing opponents' tactics by scrutinizing their players' positioning and behavioral shifts and developing tailored strategies. In addition, a meticulous examination of big data of individual opponent players across multiple games is instrumental in enhancing the strategies. Applying big data in the professional English Football League, particularly in player selection, is paramount for a team's success. Through the drafting process, teams aim to acquire promising players capable of transforming their fortunes. To identify such talent, prospective players undergo a series of rigorous physical fitness assessments. These assessments encompass a comprehensive array of statistics, such as performance, sprint speed, and overall stature. Subsequently, this data is meticulously analyzed alongside the player's historical performance in lower leagues to predict their potential draft position and prospects. Consequently, high draft picks naturally garner significant attention due to their exceptionally high success rates, underscoring the profound impact securing the top pick can have on a team's destiny.

One vital statistical metric that holds immense importance is "Expected Goals," or XG. Expected Goals is a statistical model that quantifies the quality of goal-scoring opportunities a team or player creates or faces during a match. It assigns a numerical value to each scoring chance, taking into account factors such as the position of the shot, the angle, and the type of chance (e.g., header, one-on-one with the goalkeeper), which is one of the most widely used scoring methods today. It is designed to present a player's contribution concretely and statistically using a specific and rational mathematical formula that can be used to evaluate a player's overall performance and is considered a standard quantitative indicator of a player's value. Its base data are specific parameters, such as the number of rebounds, points scored, and assists during a player's game. These data are weighted to obtain a player's efficiency value. It allows horizontal comparison of players during the season and vertical comparison of the whole season. It provides a more objective evaluation of the overall value of a player. The data mining system within the English Football League primarily relies on data mining principles to categorize and consolidate various types of data, including match events, team statistics, individual player data, and business-related information. This system is designed to create a comprehensive data warehouse specific to the English Football League, allowing for the extraction of valuable information and knowledge at any given time.

The event data encompasses fundamental details such as game timing, location, and team participation. Team data comprises essential statistics related to team performance, including scores, points, and offensive and defensive data. Individual data focuses on metrics like shots, goals, rebounds, assists, and personal physiological data.

Accumulating and analyzing this data serves as a foundation for providing scientific and effective support for English Football League tournaments, team development, and player progress. The football match data mining system primarily oper-

Journal of Intelligent Learning Systems and Applications

ates from an academic and data analysis standpoint, offering essential technical, data, and tactical analysis resources and case studies. These resources are leveraged to build a comprehensive platform dedicated to the English Football League. For a visual representation of this data mining system, please refer to **Figure 3**.

In **Figure 3**, you can see a visual representation of the integral components of the English Football League's data mining system. This system serves as a crucial support mechanism for the league, offering invaluable insights for strategic decision-making, player development, and overall league management

#### 3.2. Deep Learning Analytical Model

The approach for player action recognition in this system mimics the human action recognition process. It begins by identifying the posture of each individual at each moment and subsequently analyzes the continuous sequence of postures to ascertain the specific action being performed. In the context of real-time football data analysis, this model leverages a deep learning-based network structure, particularly employing a graph convolutional neural network design. Both approaches are rooted in the foundation of convolutional neural network design; transfer learning expedites this method's application for real-time football match data analysis, enabling effective results even with a limited dataset.

Convolutional neural networks (CNNs) comprise stacked convolutional layers, activation functions, and pooling layers organized in modules. Fully connected layers are added at the end, or feature mapping via  $2 \times 2$  convolutions is applied to adapt the features for the specific task. In signal processing, when g(u)and G(x) are bounded and predictable, the following definition is used

$$f1(x) = \int g(u)G(x-u)du * \sum_{n=0}^{\infty} (F(n))$$
  
$$f2(x) = \int g(u)G(x-u)du * \sum_{n=0}^{\infty} (F(n))$$
 (2)





The statement that g(u) and G(x) are bounded and predictable can be represented as follows:

• For *g*(*u*) being bounded:

A constant M > 0 exists such that  $|g(u)| \le M$  for all u values.

• For *G*(*x*) being bounded:

A constant N > 0 exists such that  $|G(x)| \le N$  for all values of x.

• For *g*(*u*) being predictable:

The function g(u) has a well-defined and predictable behavior over its domain.

• For *G*(*x*) being predictable:

The function G(x) has a well-defined and predictable behavior over its domain.

These conditions imply that both g(u) and G(x) do not exhibit extreme or erratic behavior and remain within certain finite limits, making them amenable to mathematical analysis and modeling.

In practice, the operation of convolution shares striking similarities with autocorrelation, with the primary distinction lying in the convolution kernel's alteration. When we delve into machine vision, the convolution layer operation parallels the convolution operation within the field of communication. In essence, the general concept of two-dimensional convolution can be succinctly expressed as follows:

$$g(x,y) = f(x,y) * W(x,y) = \iint_{\infty} f(\xi,n) Wx + (\xi,y+n) d\xi dn$$
(3)

This expression describes how we slide a small filter or kernel (W) over every possible position within the input image (f). At each position, we multiply the values of the input image and the kernel that overlap at that position, and then we sum up all these products. This process generates a new image (g) representing the input image's filtered output based on the kernel's characteristics. This operation is fundamental in image processing, computer vision, and signal processing for tasks like blurring, sharpening, edge detection, and more.

Digital images inherently possess a discrete nature. Therefore, we employ a discrete representation when performing two-dimensional convolution in image processing. This discrete approach involves applying the convolution operation on a grid of discrete pixels or elements, where each pixel represents a specific discrete value. This allows us to perform convolution operations in a way suitable for digital images, making it a fundamental technique in tasks like filtering, feature extraction, and digital imagery enhancement.

$$g(x,y) = f(x,y) * W(x,y) = \sum_{s=-a}^{a} \sum_{t=b}^{b} f(s,t) W(x+s,y+t)$$
(4)

The equation describes calculating the value of g(x, y) at a specific position (x, y) in the feature map. It does so by iterating over all possible positions within the convolution kernel (given by *s* and *t*) and computing the weighted sum of the corresponding pixel values from the original Map (*f*) and the convolution kernel (*W*). For each position (x, y) in the feature map, you take a neighborhood of

pixels from the original Map defined by the size of the convolution kernel (2*a* by 2*b*), and for each pixel in that neighborhood (indexed by s and t), then can multiply the corresponding pixel value in the original Map (*f*) by the value in the convolution kernel (*W*) at position (x + s, y + t). These products are then summed up to obtain the value of g(x, y) at that position.

In essence, this process calculates how the convolution kernel "slides" or "moves" over the original Map, and at each position, it computes the weighted sum of pixel values, resulting in the feature map g(x, y). This operation is fundamental in image filtering, edge detection, and feature extraction in image processing and computer vision.

Every neural network, including convolutional neural networks, necessitates the inclusion of nonlinear layers. Nonlinear layers are indispensable because they introduce complexity and enable the network to model intricate relationships within the data. In the absence of nonlinear layers, regardless of the network's depth, it essentially reduces to a single-layer neural network, incapable of capturing the nuanced patterns in the data. One commonly employed activation function to introduce nonlinearity is the sigmoid function.

$$f(x) = \frac{e^{-x}}{1 + e^{-x}}$$

$$f'(x) = f(x)(1 + f(x))$$
(5)

The sigmoid function is widely used for activation in neural networks and machine learning. This function maps any real number (x) to the range between 0 and 1. As *x* approaches positive infinity, f(x) approaches 1, and as *x* approaches negative infinity, f(x) approaches 0. The sigmoid function has a characteristic S-shaped curve.

The derivative, f'(x), measures the rate of change of the sigmoid function concerning its zero at the extremes (0 and 1) of the sigmoid function, indicating that the function has a flat slope at those points. It's highest at x = 0.5 (the inflection point of the sigmoid curve), where the curve has the steepest slope. The derivative f(x) is crucial in training neural networks because it's used in gradient descent algorithms, such as backpropagation. It helps adjust the weights and biases of the network during training to minimize errors and improve the network's ability to learn and make accurate predictions. The sigmoid function's derivative has the property that it's largest when the sigmoid output is around 0.5, which contributes to more effective learning in neural networks.

#### 3.3. Analysis System Design

The specific structural block diagram, illustrated in **Figure 3**, provides a comprehensive overview of the workflow:

1) Data Input: The initial phase involves the intake of diverse data sources, encompassing live match data, video feeds, historical records, and physiological measurements. This eclectic dataset forms the bedrock for subsequent analysis.

2) Data Pre-processing: Following data acquisition, a crucial step entails data

pre-processing. This involves data cleaning, transformation, and normalization and preparing the data for in-depth analysis.

3) Overall Analysis:

• Competition Analysis: To enable effective action recognition using graph convolution, we have revealed that the network performs optimally with a sequence of consecutive frames as input. Thus, the pose estimation network initially processes the frames, generating nodal data stored within a loop array to accommodate this specific frame length.

• Action Recognition with Graph Convolution: The system activates the graph convolutional network once the loop array contains full frames. This network plays a pivotal role in recognizing actions based on the input data. The ultimate categorization of actions is inferred through the computations of the graph convolutional neural network.

4) Loop Array Management: As new video frames become accessible, they automatically overwrite the oldest frames in the loop array. Subsequently, a sequential selection frame is extracted from the most recent video frames. These frames are then systematically fed into the graph convolutional network for inference. This iterative process persists until the video's conclusion or challenges arise in tracking the target object.

This comprehensive description provides a step-by-step insight into the system's operation, emphasizing the integration of diverse data sources and the sequential procedures that underpin the action recognition process. Utilizing cutting-edge data analytics platforms like Hadoop and Spark (Figure 1, Figure 2) has revolutionized how we evaluate player performance. These platforms have allowed us to delve deeper into the game's intricacies, providing a comprehensive understanding of a player's contribution to their team.

For football, this article sets up the statistics of Player Name (PN), Team (T), Games Played (GP), Goals Scored (GS), Minutes Played (MIN), Assists (ASST), SHOTS, and Shots on Goal (SOG). With its distributed processing capabilities, Hadoop has become instrumental in handling the vast data from football matches. It enables analysts to efficiently manage and process player statistics, ensuring accuracy and reliability in reporting. Using the Hadoop system to generate statistics involves several steps:

1) Data Storage:

- Store the prepared football match data in a distributed file system supported by Hadoop, such as the Hadoop Distributed File System (HDFS). Hadoop's distributed storage allows for scalability and fault tolerance.
  - 2) Map Reduce Jobs:
- Utilize Hadoop's Map Reduce programming model to process the stored data and calculate the desired statistics. Write Map Reduce jobs to perform specific tasks:
  - Map Phase: Extract relevant information from each match record, such as player names, teams, goals, minutes played, assists, shots, and shots on goal.

✓ Reduce Phase: Aggregate and calculate statistics for each player, including GP, GS, MIN, ASST, SHOTS, and SOG.

3) Output:

• Store the computed statistics in HDFS or another storage system for further analysis or reporting.

4) Data Analysis:

• Use Hadoop or other data analysis tools (Apache Spark) to query and analyze the generated statistics. Generate reports, visualizations, or dashboards to present the results meaningfully.

5) Optimization:

• Optimize the Hadoop jobs for performance and scalability, as large football datasets can be resource-intensive. This may involve tuning parameters and using Hadoop's cluster management features.

By following these steps and leveraging Hadoop's distributed processing capabilities, you can efficiently generate and analyze football statistics for Player Name, Team, Games Played, Goals Scored, Minutes Played, Assists, SHOTS, and Shots on Goal. This approach allows you to handle large volumes of data while ensuring accuracy and reliability in reporting.

On the other hand, Apache Spark, known for its lightning-fast data processing capabilities, has brought real-time analytics to the forefront of football statistics. Using Spark, analysts can instantly compute and analyze player metrics like goals, assists, and shots on goal, providing coaches, fans, and pundits with immediate insights into a player's performance. As we continue to witness the growing influence of technology in football, Hadoop and Spark stand as indispensable tools in shaping our understanding of the beautiful game. These platforms empower us to appreciate the nuances of every pass, shot, and goal, and they help us celebrate the talents of players who make football a thrilling spectacle to the entire team's performance, enhancing the system's utility for broader data analysis and team-level assessment.

#### 3.4. Data Visualization

This research uses Microsoft Power BI and Excel as a visualization platform that seamlessly integrates with Apache Spark and Hadoop Distributed File System (HDFS) to streamline visualizing and analyzing large-scale data sets. Using Microsoft Power BI and Excel, data analysts and scientists can efficiently write Spark code, access data stored in HDFS, and generate dynamic and informative visualizations within the same environment. This synergy allows for real-time data exploration, transformation, and interpretation, making it invaluable for uncovering meaningful insights from massive datasets. Microsoft Power BI userfriendly interface and support for various programming languages, including Scala and Python, make it an excellent choice for data professionals seeking to harness the full potential of Spark and HDFS for data analysis and visualization.

**Figure 4** shows the overview of the methodology employed, which involves a robust combination of technologies to streamline data analysis and visualization.



Figure 4. Overview of methodology support for the English Football League.

Microsoft Power BI, a versatile data science platform, is the foundation for this process. Data analysts and scientists can seamlessly write, execute, and iterate on Apache Spark code within Microsoft Power BI and Excel. Spark, a high-performance distributed data processing framework, is utilized for in-depth data analysis, offering the ability to process large datasets efficiently. Furthermore, the Hadoop Distributed File System (HDFS) acts as the data storage backbone, providing a scalable and fault-tolerant repository for the data. This integrated approach ensures that data can be processed, transformed, and analyzed in a distributed and parallelized manner. At the same time, Zeppelin's interactive environment facilitates dynamic data exploration and visualization, making it a methodology for comprehensive data-driven insights in a Hadoop-based ecosystem.

# 4. Analysis of Results

In this section, the study embarks on an in-depth analysis of English Football League sports data, employing PySpark's MLlib library to conduct clustering analysis. The data source file utilized for this analysis encompasses key attributes such as "PN, T, GP, GS, MIN, ASST, SH, SOG," presenting an opportunity to unravel nuanced insights into player performance within the league. The primary goal is to categorize players with similar attributes, unveiling distinctive patterns and correlations among footballers based on their statistical data.

# 4.1. Analysis of English Football League Sports Data

Utilizing PySpark's MLlib library to perform clustering analysis on a CSV file containing attributes like "PN, T, GP, GS, MIN, ASST, SH, SOG" offers a compelling opportunity to gain profound insights into player performance within the English Football League. The primary objective is to group players with similar attributes, revealing significant patterns and relationships among football players based on their statistical data **Figure 5**.

The chart in **Figure 5** visually represents player counts within the English Football League categorized into five distinct clusters. Each cluster is denoted by its corresponding numerical label (0 through 4) and showcases the number of players in the same cluster. These clusters offer insights into how players are

grouped based on their statistical attributes, shedding light on the league's diverse player profiles and performance levels. The Total at the end sums up the overall player count, which stands at 540 players in this analysis. This visual representation is a valuable tool for coaches, analysts, and enthusiasts to better understand players' distribution across different performance clusters within the league.

Power BI generated the chart in **Figure 6** to create an interactive dashboard that is a powerful way to visually represent and analyze player data.

The dashboard provides a dynamic platform to examine player groups' composition, distribution, and key attributes that define each cluster. Users can interact



Figure 5. Player counts by five clustering English Football League.



Figure 6. Group summation of the English Football League using Power BI.

with the data, filter by specific clusters or attributes, and gain real-time insights into player dynamics.

**Figure 7** presents a breakdown of player clusters within the league based on their average goals scored. This visual representation is essential in understanding the distribution of player performance within these clusters, specifically regarding their goal-scoring capabilities.

Each bar in the Figure represents one of the five clusters (labeled as 0, 1, 2, 3, and 4), and the bar height corresponds to the average number of goals scored by players in that cluster. Here's an explanation of what this Figure conveys:

1) Cluster Composition: The Figure provides insights into the composition of each cluster. For instance, Cluster 2 has the highest average goals scored, indicating that it consists of players who are prolific goal-scorers in the league. In contrast, Cluster 0 has the lowest average goals scored, suggesting that it contains players who are less involved in scoring goals.

2) Comparative Analysis: The Figure allows for a quick cluster comparison. We can observe that Cluster 2 surpasses all other clusters in terms of goal-scoring, while Cluster 0 lags behind. This information can be valuable for team managers and coaches deciding player selection and tactics.

3) Cluster Insights: By examining the average goals scored within each cluster, we can gain insights into players' playing styles and roles within those groups. For example, players in Cluster 2 may be forwards or attacking midfielders known for their goal-scoring prowess. In contrast, players in Cluster 0 could be



Figure 7. Average goals scored by cluster in the English Football League.

defenders or midfielders with fewer goal-scoring responsibilities.

4) Data-Driven Decision-Making: This Figure enables data-driven decisionmaking for various stakeholders, such as team managers, coaches, and analysts. They can use this information to strategically allocate players to different positions and roles within the team or to identify areas where additional recruitment might be needed.

**Figure 7**, a powerful visualization tool, provides a clear and concise overview of player performance in the English Football League. It empowers decision-makers with the data they need to optimize team composition, tactics, and strategies based on the goal-scoring capabilities of players in each cluster.

**Figure 8** presents a comprehensive overview of the sum of minutes played (MIN) by various teams in the English Football League, grouped by their respective clusters. The Figure provides valuable insights into the distribution of playing time across different teams, revealing how each cluster of teams has allocated minutes to their players throughout the season.

The Figure is organized with cluster numbers (0 through 4) and team names, including prominent teams such as Arsenal, Burnley, Chelsea, Everton, Watford, Liverpool, Aston Villa, Bournemouth, Southampton, Norwich City, Crystal Palace, Leicester City, Manchester City, West Ham United, Sheffield United, Manchester United, Tottenham Hotspur, and others.

For instance, under cluster 0, we observe that Arsenal has accumulated 1877



Figure 8. Average goals scored by cluster in the English Football League.

minutes, Burnley has 3370 minutes, Chelsea has 1726 minutes, and so on. Similarly, cluster 1 displays minute distributions for these teams and so forth for the other clusters.

This visualization helps understand how different teams manage their players' minutes and highlights patterns and disparities in player usage among clusters, contributing to more informed decisions regarding player rotation, performance assessment, and team strategies within the English Football League.

**Figure 9** provides a comprehensive view of the average assists (ASST) across different teams in the English Football League, organized by their respective clusters. This table reveals the typical level of assistance provided by players in each cluster, shedding light on the teamwork and playmaking abilities of teams within the league.

**Figure 8** is structured with cluster numbers in rows (0 through 4) and team names in columns, encompassing teams like Arsenal, Burnley, Chelsea, Everton, Watford, Liverpool, Aston Villa, Bournemouth, Southampton, Norwich City, Crystal Palace, Leicester City, Manchester City, West Ham United, Sheffield United, Manchester United, Tottenham Hotspur, and others.

For instance, within cluster 0, we observe that Arsenal has an average of 4 assists, Burnley has 0.5 assists, Chelsea has four assists, and so forth. Similarly, cluster 1 displays average assist values for these teams and the remaining clusters.





**Figure 9** insights can be instrumental in analyzing team strategies, player roles, and overall performance within the English Football League, offering valuable data for coaches, analysts, and fans to assess and appreciate the teamwork and creativity exhibited by different teams.

**Figure 10** illustrates the average number of shots (SH) taken by teams within different clusters in the English Football League. This table provides valuable insights into teams' offensive strategies and shooting tendencies, shedding light on their attacking styles.

The Figure is organized with cluster numbers (0 through 4) in rows and team names in columns, encompassing teams such as Arsenal, Burnley, Chelsea, Everton, Watford, Liverpool, Aston Villa, Bournemouth, Southampton, Norwich City, Crystal Palace, Leicester City, Manchester City, West Ham United, Sheffield United, Manchester United, Tottenham Hotspur, and others.

For instance, within cluster 0, we observe that Arsenal averages 44 shots, Burnley averages 43.5 shots, Chelsea averages 48 shots, and so forth. Similarly, cluster 1 displays the average number of shots for these teams and the remaining clusters.

Figure 10 insights can be instrumental in analyzing team strategies, assessing their goal-scoring potential, and understanding how different teams approach their offensive play in the English Football League. It offers valuable data for coaches, analysts, and fans to evaluate and appreciate the diverse playing styles within the league.

**Figure 11** presents the average number of shots on goal (SOG) for various teams grouped into different clusters within the English Football League. This Figure offers valuable insights into teams' ability to generate high-quality scoring opportunities and their proficiency in converting these opportunities into shots on target.



Figure 10. Average shots by cluster in the English Football League.



Figure 11. Average SOG by cluster in the English Football League.

The Figure is structured with cluster numbers (0 through 4) in rows and team names in columns, encompassing teams like Arsenal, Burnley, Chelsea, Everton, Watford, Liverpool, Aston Villa, Bournemouth, Southampton, Norwich City, Crystal Palace, Leicester City, Manchester City, West Ham United, Sheffield United, Manchester United, Tottenham Hotspur, and others.

For example, within cluster 0, we observe that Arsenal averages 26 shots on goal, Burnley averages 21, Chelsea averages 26, and so forth. The data within cluster 1 provides the average SOG for these teams, and the pattern continues for the remaining clusters.

Figure 11 insights are instrumental in evaluating teams' offensive capabilities and potential to score goals in the English Football League. It offers valuable data for coaches, analysts, and fans to understand and compare teams' offensive performances in the league.

**Figure 12** provides a comprehensive breakdown of the count of players grouped by team name and clustered into five distinct groups within the English Football League. The Figure presents teams' names in the leftmost column and clusters (Group 0 through Group 4) as column headers.

Each cell in The Figure contains the count of players belonging to a specific team and cluster. For instance, in Group 0, Arsenal has one player, Aston Villa has one player, Bournemouth has one player, Burnley has one player, and so on. The same pattern continues for all five clusters, providing a clear overview of how players are distributed among the teams and clusters.

**Figure 12** showcases the composition of each team within the clusters and highlights any exclusivity or overlap of players between clusters and teams. This information is crucial for team managers, coaches, and analysts to understand



Figure 12. Overview of the count of players within different clusters, grouped by team names.

player allocations and distributions in different performance groups, facilitating data-driven decisions related to team composition, strategies, and player development within the English Football League

## 4.2. Deep Learning Model Analysis in Sports

Technical and tactical analysis approaches may vary across sports, but the fundamental steps remain consistent: collecting raw data, extracting valuable insights, and conducting a comprehensive analysis. In official competitions, athletes often cannot wear additional equipment, necessitating a focus on extracting game-related information through deep learning techniques. The system transforms seemingly complex and chaotic game data into an easily interpretable structured dataset. Machine learning methods are then applied to empower team data analysts and coaches with profound insights into their teams.

This transformative shift, once deemed a strategic risk, has yielded remarkable results, astounding everyone involved. The deep learning model analysis outcomes are depicted in **Figures 4-11**, illustrating how the intelligent data analysis system offers a fresh perspective on understanding the game. This system challenges and, to some extent, supplants traditional team data analysis methods. It excels in classifying players based on shared characteristics, a task that traditional data analysis struggled to achieve. However, this enhanced access to information entails substantial data processing.

During gameplay, players dedicate more time to running and positioning than shooting and passing. Consequently, if too few features are used for smoothing, excessive smoothing occurs, impairing the system's ability to detect similar player clusters. To address this, our approach employs five clusters to enhance the recognition of shared attributes among players, capturing a comprehensive range of player actions. Missing data points necessitate further position identification, preventing any loss of significance. It's important to note that this analysis employs a limited set of features due to the current relatively small dataset and the general nature of football action recognition. Future work will require expanding data collection and labeling efforts, potentially subdividing actions to enhance the network's accuracy in recognizing player actions.

# **5. Discussion**

The analysis presented in this study showcases the transformative potential of deep learning models in sports, particularly in the context of football. The advent of intelligent data analysis systems driven by machine learning techniques has fundamentally challenged traditional methods of technical and tactical analysis in sports. This evolution has led to a profound shift in understanding and interpreting the game, opening up new avenues for previously unimaginable insight.

The deep learning model employed in this analysis has enabled the extraction of intricate patterns and player clusters from seemingly complex and unstructured game data. These clusters, as illustrated in **Figures 4-11**, reveal the power of machine learning in classifying players based on shared attributes and performance metrics. Such player groupings are not only valuable for understanding team dynamics but also have the potential to inform critical decisions related to team composition, tactics, and strategy.

However, this deep learning approach's success comes with its challenges. As demonstrated, choosing the number of clusters and features is crucial in ensuring meaningful insights. Striking the right balance between capturing nuanced player behaviors and avoiding excessive smoothing is an ongoing challenge that requires careful consideration.

Moreover, the analysis's scope is constrained by the size of the dataset and the current limitations in football action recognition. To further advance this field, expanding data collection efforts and fine-tuning recognition algorithms are essential for enhancing the accuracy and granularity of player action recognition.

Integrating deep learning models into sports analysis represents a significant paradigm shift. It empowers teams and analysts with a data-driven understanding of the game, enabling them to make more informed decisions. As technology advances and datasets grow, we can anticipate even more precise and insightful analyses that redefine our comprehension of sports performance and strategy.

#### 6. Conclusions

The application of big data analytics in football has ushered in a transformative era, enabling precise insights into player behaviors and opponent strategies. By harnessing the power of big data, football teams can craft tailored one-on-one strategies, enhancing the offensive prowess and defining the roles of individual players. Furthermore, it facilitates an in-depth analysis of the offensive and defensive synergies among players, unveiling the tactical styles of opposing teams. This, in turn, allows teams to formulate holistic strategies that align with their strengths and player configurations.

The granular examination of player data, encompassing metrics such as Goals Scored (GS), Minutes Played (MIN), Assists (ASST), SHOTS, and Shots on Goal (SOG), empowers teams to optimize player positions and identify ideal roles. Using big data analytics improves individual and team performance and revolutionizes the efficiency of offensive and defensive tactics.

The Sport AI Model (SAIM) is a powerful tool for visualizing player and team data, facilitating enhanced strategic planning and decision-making. This approach leverages comprehensive statistics from the English Football League, subjected to intelligent analysis systems to extract profound insights. While deep learning techniques have exhibited maturity and remarkable achievements in technical and tactical game analysis, there remains untapped potential in their application to player training and development.

In summary, integrating big data analytics, powered by the SAIM model, can reshape football strategies, elevate player performance, and drive informed decision-making. While its impact on professional game analysis is evident, future research holds promise in unlocking its full potential in player training and development, ushering in a new era of football excellence.

## 7. Future Work

To further advance the capabilities and impact of the Sport AI Model (SAIM) framework in football analytics, several avenues for future work present themselves.

1) Schedule and Automation:

Setting up automated data collection and processing pipelines is imperative to ensure the continuous relevance and accuracy of the statistics. Integrating automated procedures allows the SAIM framework to regularly fetch and process new match data. This keeps the analyses up-to-date and allows real-time player and team performance monitoring, enabling teams and analysts to make more agile decisions.

2) Enhanced Training and Development:

Future research should investigate applying deep learning techniques in player training and development. By leveraging the rich dataset generated by the SAIM framework, it is possible to create personalized training programs for individual players. These programs can focus on refining specific skills and addressing weaknesses, ultimately nurturing a new generation of highly skilled and adaptable football players.

3) Predictive Analytics:

Expanding the SAIM framework to encompass predictive analytics is a promising avenue. The system can forecast various game scenarios and outcomes using advanced machine learning algorithms. This can assist coaches in making data-driven decisions during matches and even aid in scouting new talent based on predictive performance indicators.

4) Fan Engagement:

The insights generated by the SAIM framework can be harnessed to enhance fan engagement. Developing interactive dashboards and applications that provide fans with real-time statistics, player insights, and tactical analyses can create a more immersive and informed viewing experience, fostering greater fan loyalty and interest.

These future work directions will strengthen the SAIM framework's capabilities and contribute to the evolution of football analytics, benefiting teams, players, fans, and the sport.

## **Data Source**

The study's data extraction process involves gathering information from multiple sources, including a training database and online repositories. The primary data sources comprise two repositories: the first is a GitHub repository housing football data related to the English league from the 2020-21 season. This comprehensive collection of football statistics is accessible at

<u>https://github.com/footballcsv/england/tree/master/2020s/2020-21</u>. The second source, available at <u>https://xvalue.ai/stats/en/league/premier\_league</u>, offers de-tailed statistical analyses specifically focused on the Premier League.

## **Conflicts of Interest**

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- Berisha, B., Mëziu, E. and Shabani, I. (2022) Big Data Analytics in Cloud Computing: An Overview. *Journal of Cloud Computing*, 11, Article No. 24. https://doi.org/10.1186/s13677-022-00301-w
- [2] Jadhav, A., Rasool, A. and Gyanchandani, M. (2023) Quantum Machine Learning: Scope for Real-World Problems. *Procedia Computer Science*, 218, 2612-2625. https://doi.org/10.1016/j.procs.2023.01.235
- [3] Saratchandra, M. and Shrestha, A. (2022) The Role of Cloud Computing in Knowledge Management for Small and Medium Enterprises: A Systematic Literature Review. *Journal of Knowledge Management*, 26, 2668-2698. https://doi.org/10.1108/JKM-06-2021-0421
- [4] Al-Jumaili, A.H.A., Muniyandi, R.C., Hasan, M.K., Paw, J.K.S. and Singh, M.J. (2023) Big Data Analytics Using Cloud Computing Based Frameworks for Power Management Systems: Status, Constraints, and Future Recommendations. *Sensors*, 23, Article No. 2952. <u>https://doi.org/10.3390/s23062952</u>
- [5] Li, Y. and Hei, X. (2022) Performance Optimization of Computing Task Scheduling Based on the Hadoop Big Data Platform. *Neural Computing and Applications*. <u>https://doi.org/10.1007/s00521-022-08114-3</u>

- [6] Deshmukh, S.S. (2023) Progress in Machine Learning Techniques for Stock Market Movement Forecast. Proceedings of the International Conference on Applications of Machine Intelligence and Data Analytics (ICAMIDA 2022), Vol. 105, 69. https://doi.org/10.2991/978-94-6463-136-4\_9
- [7] Mentis, A.F.A., Lee, D. and Roussos, P. (2023) Applications of Artificial Intelligence—Machine Learning for Detection of Stress: A Critical Overview. *Molecular Psychiatry*. <u>https://doi.org/10.1038/s41380-023-02047-6</u>
- [8] Ding, Z., Wang, H., Sun, Y. and Qin, H. (2022) Adaptive Prescribed Performance Second-Order Sliding Mode Tracking Control of Autonomous Underwater Vehicle Using Neural Network-Based Disturbance Observer. *Ocean Engineering*, 260, Article ID: 111939. <u>https://doi.org/10.1016/j.oceaneng.2022.111939</u>
- [9] Martin, P.E., Siler, W.L. and Hoffman, D. (1990) Electromyographic Analysis of Bow String Release in Highly Skilled Archers. *Journal of Sports Sciences*, 8, 215-221. https://doi.org/10.1080/02640419008732147
- [10] Dzwil, M. (2023) Predicting Rookie Season Performance Based on National Football League (NFL) Scouting Combine Movement Analysis. Doctoral Dissertation, Worcester Polytechnic Institute, Worcester.
- [11] Nguyen, N.H., An Nguyen, D.T., Ma, B.K. and Hu, J. (2022) The Application of Machine Learning and Deep Learning in Sport: Predicting NBA Players' Performance and Popularity. *Journal of Information and Telecommunication*, 6, 217-235. <u>https://doi.org/10.1080/24751839.2021.1977066</u>
- [12] Yao, P. (2021) Real-Time Analysis of Basketball Sports Data Based on Deep Learning. *Complexity*, 2021, Article ID: 9142697. <u>https://doi.org/10.1155/2021/9142697</u>
- [13] Zhu, H., Zhang, P., Wang, L., Zhang, X. and Jiao, L. (2019) A Multiscale Object Detection Approach for Remote Sensing Images Based on MSE-DenseNet and the Dynamic Anchor Assignment. *Remote Sensing Letters*, **10**, 959-967. <u>https://doi.org/10.1080/2150704X.2019.1633486</u>
- Shan, C., Brea, V.M. and Velipasalar, S. (2020) Special Issue on Smart Cameras for Real-Time Image and Video Processing. *Journal of Real-Time Image Processing*, 17, 1755-1756. <u>https://doi.org/10.1007/s11554-020-01006-6</u>
- [15] Liu, G., Luo, Y., Schulte, O. and Kharrat, T. (2020) Deep Soccer Analytics: Learning an Action-Value Function for Evaluating Soccer Players. *Data Mining and Knowledge Discovery*, **34**, 1531-1559. <u>https://doi.org/10.1007/s10618-020-00705-9</u>
- [16] Rangasamy, K., As'ari, M.A., Rahmad, N.A., Ghazali, N.F. and Ismail, S. (2020) Deep Learning in Sport Video Analysis: A Review. *Telkomnika (Telecommunication Computing Electronics and Control)*, **18**, 1926-1933. <u>https://doi.org/10.12928/telkomnika.v18i4.14730</u>
- [17] Rana, M. and Mittal, V. (2020) Wearable Sensors for Real-Time Kinematics Analysis in Sports: A Review. *IEEE Sensors Journal*, 21, 1187-1207. https://doi.org/10.1109/JSEN.2020.3019016
- [18] Van Rossem, S., Tavernier, W., Colle, D., Pickavet, M. and Demeester, P. (2019) Profile-Based Resource Allocation for Virtualized Network Functions. *IEEE Transactions on Network and Service Management*, **16**, 1374-1388. <u>https://doi.org/10.1109/TNSM.2019.2943779</u>
- [19] Van Rossem, S., Tavernier, W., Colle, D., Pickavet, M. and Demeester, P. (2020) Optimized Sampling Strategies to Model the Performance of Virtualized Network Functions. *Journal of Network and Systems Management*, 28, 1482-1521. https://doi.org/10.1007/s10922-020-09547-8

 Schneider, S., Satheeschandran, N.P., Peuster, M. and Karl, H. (2020) Machine Learning for Dynamic Resource Allocation in Network Function Virtualization. *Proceedings of the* 2020 6th *IEEE Conference on Network Softwarization (NetSoft)*, Ghent, 29 June-3 July 2020, 122-130. <u>https://doi.org/10.1109/NetSoft48620.2020.9165348</u>