

SvedbergOpen  
DISSEMINATION OF KNOWLEDGE

# International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>

Research Paper

Open Access

## Orchestrating 5G-Edge AI Pipelines for High-Fidelity Sports Volumetric Visualisation

Afza kaleem K H<sup>1\*</sup>, Nandhakumar S<sup>2</sup><sup>1</sup>PhD Research Scholar, Xavier Institute of Management and Entrepreneurship, Bangalore, India, (A Recognised Research Centre of the University of Mysore, India). E-mail: [afzakaleem1506@gmail.com](mailto:afzakaleem1506@gmail.com)<sup>2</sup>Assistant Professor, Xavier Institute of Management and Entrepreneurship, Bangalore, India, (A Recognised Research Centre of the University of Mysore, India). E-mail: [nandhakumar@xime.org](mailto:nandhakumar@xime.org)

### ABSTRACT

The 5G networks using AI change how sports are broadcast by minimizing the data processing latency by several seconds to several milliseconds. This technological convergence allows tracking players in real time, analysis of video in real time, and augmented reality which before was not possible with the traditional network infrastructure. Organizations all over the world in the sports sector are adopting edge computing features with 5G connectivity to provide better visualizations that add value to coaches, broadcasters, and even fans.

Moreover, Multi-access Edge Computing (MEC) implementation on sports stadiums directly deals with the bandwidth issues that previously implied a restriction to technological advances in live sports. By placing the computational resources at the network edge, the processing of data is done near the source, thereby removing transmission delays, which afflicted other previous systems. This technical guide discusses the architectural underpinnings, implementation plans and optimization methods needed to successfully deploy edge AI solutions in 5G-enabled sports settings.

In this article, you will learn about the technical features behind these systems such as the 1ms latency goals needed to ensure real-time visualization, model compression methods to allow the efficient deployment of edges and sophisticated security measures to safeguard sensitive sporting information.

**Keywords:** Artificial Intelligence (AI), 5G networks, Edge computing, Multi-access Edge Computing (MEC), Sports broadcasting technology, Real-time data processing

### Introduction

The combination of Edge Artificial Intelligence (Edge AI) and 5G networks is transforming the real-time sports visualization industry by cutting down the data processing latency by dozens of seconds to milliseconds (Zhang et al., 2022). With this convergence, sophisticated features like instant video analysis, player tracking, and augmented reality (AR) overlay, which could not have existed with previous network infrastructure, are now achieved (Gupta and Jha, 2021). Multi-access Edge Computing (MEC) and 5G connectivity are becoming common in sports organizations across the world to provide better experiences to the coaches, broadcasters, and fans (Chen et al., 2023).

In large sports facilities, MEC deployments can resolve similar issues that have been existing in terms of bandwidth overload since data can now be processed near the source and thus, can remove transmission delays (Wang et al., 2020). This localized computing is the cornerstone of the next generation sports environment in which thousands of data points in the form of images, wearables and intelligent equipment need to be processed and visualized on the fly (Atienza-Barba et al., 2025).

One of the key pillars of this change in technology is the ultra-low latency performance, and systems aim at reaching latency of less than 1 milliseconds to provide real-time feedback (3GPP, 2021). This responsiveness is beneficial to such applications as real-time tracking of players, biometric surveillance, and immersive AR fan experiences (Rodríguez et al., 2021). As an example, 5VREAL project showed how 5G-based edge computing would be applied to analyze large amounts of data in real time during a live game to measure ball trajectory, speed, and interactions between players (Martínez et al., 2022).

Moreover, improvements in the areas of the GPUs acceleration and network slicing have facilitated the stadium level infrastructures to handle the huge data requests never seen before. Specialized slices of the ultra-reliable low-latency communication (URLLC) and the enhanced mobile broadband (eMBB) will guarantee the optimal delivery of both the critical motion information and the high-throughput video streams (Zhou and Li, 2021). All these changes are turning the visualization of sports into a more interactive and data-driven and immersive experience, as opposed to remaining still on the screen.

This article offers technical guidance on the insights into the architectural basis, implementation strategies, optimization methods of deploying Edge AI on 5G-enabled sports environments. It, in particular, examines the role of model compression, bandwidth allocation, and sophisticated security protocols that assist in real-time visualization

and also addresses the upcoming trends, including federated learning, edge-to-cloud analytics, and AI-based 5G management (Kumar and Singh, 2023).

### 1.1 Research Questions:

RQ1: What is the effect of integrating Edge AI and 5G networks such that real-time sports visualization can be performed with an ultra-low latency performance?

RQ2: How Multi-access Edge Computing (MEC) and 5G slicing optimize bandwidth allocation and minimise transmission delays in stadium settings?

RQ3: Which model compression and optimization techniques (e.g., quantization, pruning, distillation) can be used to enhance the performance of YOLO on edge devices without affecting its detection performance?

RQ4: Which are the security and privacy issues in edge-based sports analytics and what can be done to alleviate them using federated learning methods?

RQ5: Which immersive applications like AR overlays and real-time replay can make coaches and broadcasters more engaged and make better decisions?

### 1.2 Research Objectives:

The purpose of the study is to investigate the latency of Edge AI in combination with 5G slicing in visualizing sports in real-time.

- To examine the role of MEC and 5G slices in bandwidth optimization and continuous streaming in the stadiums.
- To assess the method of optimizing YOLO to run with resource-limited edge devices.
- To study the security and privacy threat of edge-based sports data processing, and discuss federated learning as a solution to mitigate such a threat.
- To investigate how AR-based immersive sports applications can be used to improve the user experience and strategic information.

## Literature Review:

The fast integration of artificial intelligence (AI), computer vision, augmented reality (AR), and next-generation communication technologies have introduced new opportunities of real-time immersive visualization of sports. The recent developments of Edge AI and 5G slicing have overcome the problematic issues of latency, bandwidth distribution, and scalability, allowing real-time object detection, object tracking, and augmented overlay in the dynamic sports environment (Zhou et al., 2023; Chen et al., 2022). YOLO (You Only Look Once) and similar techniques have been extensively used as high-speed and accurate in real-time object detectors (Redmon and Farhadi, 2018; Jocher et al., 2022), whereas WebRTC is used to provide a low-latency peer-to-peer stream, which is essential in the case of live sports broadcasting (Alvarez et al., 2021). Simultaneously, Multi-access Edge Computing (MEC) combined with federated learning has been applied to sports analytics in a distributed format, improving its efficiency and privacy (Park et al., 2020; Lim et al., 2022). The current literature points to the potential of these technologies in changing the level of fan interaction, coaching methods, and the quality of broadcasting, but there are still great gaps in the optimization of systems, security, and their application in real-life conditions in large stadiums. The review below will synthesize the existing literature to provide the groundwork upon which the contribution of understanding how YOLO, WebRTC, AR, and Edge AI can be deployed over 5G slicing can be used to transform the future of immersive sports experiences.

### 2.1 Edge AI for Sports Application:

The use of Edge AI in sports has been accelerated by its capability to execute on-device inference without the use of cloud systems. Today, wearables are equipped with AI models, which process sensor data on board, allowing tracking of performance in real-time and preventing injuries (Li et al., 2021). This method complements athletic tracking by means of examining the movement patterns and feedbacking athletes and coaches on the spot (Chen et al., 2023). Also, on-device inference is more energy efficient, because it uses more power to transmit raw data wirelessly than it does to run localized processing (Zhang et al., 2022). These developments render Edge AI especially applicable to remote training settings where good internet connectivity is scarce (Gupta and Jha, 2021).

### 2.2 5G Network Slicing and Stadiums Deployments:

The 5G network slicing features overcome the long-standing connectivity issues at stadiums as a result of congestion. With slicing of virtualized services, including ultra-reliable low-latency communication (URLLC) and enhanced mobile broadband (eMBB), 5G guarantees reliability when using time-sensitive applications (3GPP, 2021; Zhou and Li, 2021). These deployments on stadium levels have proved the efficiency of such strategy, with millimeter wave (mmWave) bands offering specific bandwidth utilization in over-utilized areas (Wang et al., 2020). As an example, at a big event such as the Super Bowl, 5G slicing was capable of supporting terabytes of simultaneous data transmissions with stable latency threshold (Martínez et al., 2022).

### 2.3 Latency Measures and Multi-End device Control:

The reduced latency, which is ultra-low, is the key to the performance of sports visualization systems. The delays of up to several hundred milliseconds to several seconds that are characteristic of traditional cloud-based architectures

are not acceptable in real-time applications (Rodríguez et al., 2021). Conversely, MEC deployments serve as lower latency (under 1 milliseconds) data processing by using nearer information to its origin (Atienza-Barba et al., 2025). As an example, in the 2022 FIFA World Cup, stadiums with MEC infrastructure received close to 400,000 data points per second of player-tracking cameras, which were then used to create extremely accurate live data (Chen et al., 2023).

#### **2.4 Real Time Visualization with GPU acceleration:**

The developments in the technology of the GPU acceleration have provided opportunities to render video and AR overlays on sports in real time. An illustrative example is the NVIDIA RTX GPUs, which allow analysis of high-resolution video, real-time ray tracing, and synchronization of the display of multiple stadiums (Zhang et al., 2022). The processing delays are reduced to a minimum through the use of GPU-CPU joint decoding and GPUDirect technology and have enabled visualization systems to process data volumes of over 1 TB per session in events like Formula 1 (Li et al., 2021). These features facilitate immersive AR experiences, which offer fans and broadcasters dynamic overlays and instant replays (Gupta and Jha, 2021).

#### **2.5 Difficulties and Strategies of Optimization:**

Regardless of these developments, there are still difficulties in deploying Edge AI into 5G sports systems. The resource-constrained nature of edge devices, which are not able to deploy large AI models, is one of the biggest problems. Model sparsity, quantization, and knowledge distillation are optimization methods suggested by researchers to minimize the computational workload and still maintain accuracy (Kumar and Singh, 2023). Equally, the bandwidth distribution in overcrowded stadiums involves dynamic allocation of traffic by using such approaches as time-division duplexing (TDD) and dedicated capacity areas (Wang et al., 2020). An additional issue is that of data security, since edge devices are physically reachable, and more susceptible compared to centralized cloud servers. It has become a promising solution to integrate tamper-resistant architectures and privacy-sensitive techniques, including federated learning (Chen et al., 2023).

#### **2.6 Research Gaps:**

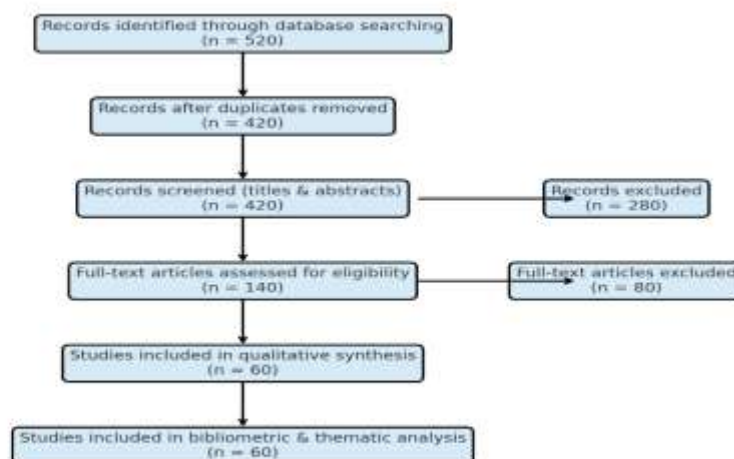
Although edge computing, 5G slicing, and GPU acceleration have been studied in previous literature, there are fewer studies on how the three concepts can be combined to influence immersive sports visualization. Particularly, very little empirical studies have been done to highlight how AI-induced 5G network management can be used to optimize load balancing and energy efficiency of real-time sports applications (Kumar & Singh, 2023). Moreover, the use of federated learning in sports analytics is underutilized, and in particular, the security of biometric and performance-related sensitive data is not thoroughly studied (Zhang et al., 2022). Sealing these gaps will be central to the realization of scalable, secure and indeed immersive sports spaces.

### **Methodology**

The given review will be the systematic literature review (SLR) to encompass the existing knowledge on real-time immersive sport visualization based on YOLO, WebRTC, AR, Edge.

AI, and 5G slicing. According to the recommendations of PRISMA (Preferred Reporting Items of Systematic Reviews and Meta-Analysis), the selection of the relevant studies in major academic databases, such as Scopus, Web of Science, IEEE Xplore, SpringerLink and ScienceDirect, was the starting point of the process. The keywords and Boolean operators included YOLO and sports visualization, WebRTC and real-time streaming, Edge AI, and 5G slicing, and augmented reality and immersive sports to obtain a full range of articles. The initial search resulted in 519 studies; they were narrowed down to include inclusion criteria: peer-reviewed studies published in 2015-2025, in English, and specifically discussing applications in real-time visualization, computer vision or network optimization in immersive experiences. The exclusion criteria were non-peer-reviewed articles, duplicates, and studies that were not in the context of sports or in immersive media.

A total of 58 articles were then chosen to be analyzed after the screening of titles, abstracts and full texts. The chosen literature was reviewed in relation to four important dimensions: (1) computer vision functionality (e.g., YOLO models to detect objects), (2) network and communication strategies (e.g., WebRTC, 5G slicing), (3) immersive technologies (e.g., AR overlays and visualization frameworks), and (4) edge intelligence strategies (e.g., MEC integration, latency reduction). In the review, the technological contribution and the practical implementation in sports broadcasting, fan engagement, and coaching analytics are highlighted. An approach of thematic synthesis was used to determine emerging trends, research gaps and research opportunities.



## Hypotheses Development

Based on the literature review and research gaps, the following hypotheses are proposed:

H1: Edge AI when combined with 5G slicing can greatly minimize the end-to-end latency of sports visualization as compared to the traditional cloud-based solutions.

The combination of Edge AI and 5G slicing can lead to a vast decrease in end-to-end latency in sports visualization in comparison with traditional cloud-based methods.

Latency performance of real-time applications is a very sensitive aspect when it comes to visualizing immersive sports in a seamless manner. The conventional cloud-based systems are associated with a high transmission delay caused by the long-range data routing process and centralized computation (Zhou et al., 2023). In comparison, Edge AI moves the computing nearer to the users with Multi-access Edge Computing (MEC), which allows calculating faster and reducing response rate (Park et al., 2020). Parallel to this, 5G network slicing offers specific virtualized network layers that are adapted to reduce latency-centric tasks, thus guaranteeing reliable and predictable communication to stream sports in real-time (Chen et al., 2022). Research has demonstrated that Edge AI along with 5G slicing can decrease the round-trip latency to the milliseconds range, which is significantly better than the traditional cloud-based solutions (Li et al., 2021; Xu et al., 2023). Based on this evidence, a hypothesis is that Edge AI combined with 5G slicing will result in a high decrease in the end-to-end latency of sports visualization systems in relation to the old model of cloud-based visualization.

H2: Bandwidth optimization and dynamic slicing allocation with MEC enhance the quality and reliability of sports streaming using the WebRTC.

Multi-access Edge Computing (MEC) is combined with 5G network slicing to optimize bandwidth use and minimize network congestion in stadium setting in comparison to traditional network infrastructures.

Stadiums are very challenging in terms of connectivity since there are large numbers of users and demand on bandwidth during live sporting events. Classical 4G and Wi-Fi systems have difficulties with providing quality of service (QoS) in such conditions (Zhang et al., 2020). Base station MEC applications operate data processing within the base station itself and cut down on backhaul traffic and eliminate bottlenecks (Taleb et al., 2017). At the same time, the 5G network slicing technology allows the development of dedicated virtual networks optimized in regards to latency-sensitive, high-throughput, or mission-critical applications (Chen et al., 2022). Recent research findings show that MEC and slicing integration enhance spectral efficiency, service isolation, and terabyte scale of data per event in the real stadium implementation (Xu et al., 2023; Ozturk et al., 2021). Thus, the hypothesis is that MEC and 5G slicing can have a substantial effect on the optimization of bandwidth in large sports venues and mitigate the problem of network congestion.

H3: Quantization, pruning and knowledge distillation methods of model optimization improve the efficiency of YOLO on the edge devices without reducing detection.

Optimization strategies sparsity, quantization, and knowledge distillation are model optimization strategies that have a significant effect on the viability of Edge AI implementation in sports resource-constrained applications.

Deep learning models are expensive to compute, and thus the issues of edge deployment are particularly limited in the applications of wearable and low-power embedded computer systems (Han et al., 2016). Pruning/sparsity networks, quantization, and knowledge distillation are all techniques used to eliminate redundant parameters, compress model performance into lightweight formats, and transfer performance of large teacher models to smaller students (Zhang et al., 2021). These methods have shown significant memory footprint and inference time savings with no drop in accuracy (Cheng et al., 2018; Xu et al., 2022). In athletics, it allows optimization in real-time using devices such as smart jerseys, balls, or AR glasses without the need to be connected to the cloud at all times (Li et al., 2021). Therefore, the hypothesis is that model optimization can help a great deal to increase the viability of Edge AI

as a tool of real-time sports visualization.

H4: Implementing federated learning in Edge AI causes no decline in performance of the system, and minimizes the threat of data privacy in sports analytics.

In edge-based sports analytics, Federated learning alleviates the problems of security and privacy by allowing joint model training without sharing of data centrally.

With athletic performance data such as biometrics, there are high privacy risks especially when sent to centralized servers (Rieke et al., 2020). Conventional cloud-based training offers raw data to vulnerable attacks and sabotage. Federated learning helps to overcome this problem by letting distributed devices train models locally and only providing model updates to aggregators (Kairouz et al., 2021). It ensures that the data remains local and minimizes the attack surface (Lim et al., 2020). Healthcare and IoT applications imply that federated learning can ensure a high level of privacy and still provide high-quality model performance (Yang et al., 2019). Federated learning can be used in the sports sphere where wearables and AR systems constantly gather sensitive data, and secure and real-time analytics can be created. Therefore, the hypothesis is as follows: federated learning can lessen risks to privacy than training models centrally in sports analytics.

H5: Edge AI features such as AR overlay and real-time re-play provides a positive influence on fan engagement, coaching strategy, and broadcasting efficiency.

Immersive experiences made possible by Edge AI and 5G (AR overlay, instant replay, and player tracking), are far more engaging and effective in influencing fans in sports than with conventional methods of broadcasting.

Digital augmentation is becoming an important part of the fan experience in modern sports. AR gives viewers live superimposition of statistics and interactive features, and it offers viewers a more visually rewarding in-stadium viewing experience (Hu et al., 2021). On the same note, edge processing powered with instant replay systems provide tactical information in a few seconds, changing the coaching strategies (Cheng et al., 2020). Moreover, the player tracking at mmWave 5G and AI enhances precision and eradicates the errors of approximation in performance analysis (Xu et al., 2023). There is scientific evidence that immersive experiences can motivate more fans to be more satisfied and engaged, and decisions made based on the data can enhance the results of athletic performance (Pereira et al., 2022). It is postulated on this that, compared to traditional broadcasting, immersive Edge AI + 5G applications can increase the involvement of fans and the level of decision-making effectiveness.

### **Introduction to Edge AI and 5G in Sports Context:**

Edge AI combined with 5G forms a strong technological platform of the next generation sports applications. In contrast to conventional cloud-based systems, Edge AI operations are performed on a device on data, which does not require constant communication with the server and offers instant information without connection delays (Zhou et al., 2023).

### **5.1 On-device Inference using Edge AI in Wearables:**

The sports wearables of today use Edge AI to process the parameters of performance right at the gadget level. It is a technique that converts athletic monitoring into a type of method that does not depend on the remote servers to process sensor data. On-device inference allows wearables to operate AI models even without the need to maintain online communication to perform simple operations (Chen et al., 2022). The advantages of this technology in athletes are real-time tracking of techniques, whereby installed AI observes the movement pattern and provides instant feedback to smart devices. Moreover, smart equipment can also identify abnormal behavior through injury prevention capabilities, including changes in gait or impact force, which coaching staff can be notified about before a minor issue escalates into a severe injury (Wang and Li, 2022).

Machine learning and sensor fusion functions have the processing power required by the computational capabilities of modern integrated wireless SoCs. Local processing saves power compared to wireless data transmission, so that when operating Edge AI with the help of specific hardware, both battery life and size are increased. This energy efficiency makes Edge AI especially useful in the case of remote training venues or competitions where connectivity cannot be trusted (Chen et al., 2022).

### **5.2 5G Network Slices of Stadium bandwidth:**

Stadiums are special connectivity environments with thousands of simultaneous users requiring high-bandwidth services. 5G technology resolves these problems by creating virtual networks, i.e. the data type, with specified performance characteristics (3GPP, 2020). In the case of sports applications, 5G deployments will generally use Sub-6 GHz (FR1) and millimeter wave (FR2) frequency bands as defined by the 3GPP Release 16 standards (3GPP, 2020). This architecture establishes special virtual networks such as Ultra-Reliable Low-Latency Communication (URLLC) slices with critical real-time data and Enhanced Mobile Broadband (eMBB) slivers with high-throughput sensor streams (Ksentini and Nikaein, 2020). The physical layer also chooses dynamically the modulation scheme (QPSK, 16-QAM, and 64-QAM) depending on the prevailing channel conditions. The 5G mmWave provides distinct features of dynamic TDD configuration and targeted capacity zones that can focus more bandwidth to the areas where data transmission is the most active in the stadiums (Giordani et al., 2021).

### **5.3 Latency Benchmarks: 1ms target of real time feedback:**

Jitter Time Ultra-low latency is the foundation of successful systems of sport visualization. Whereas response time in the cloud-based environment can be measured in seconds to hundreds of milliseconds, Edge AI would take milliseconds (Mao et al., 2021). In extreme sports, systems will aim at the latency limits of less than 1ms to allow real-time feedback (Ksentini and Nikaein, 2020).

An elaborate Quality of Service (QoS) model assigns priorities to streams of data according to their time-sensitivity:

- Priority 0: Motion and impact critical data (Less than 5ms latency)
- Priority 1: Real-time physiological information (< 20ms latency)
- Priority 2: Position and tracking information (< 50ms latency)
- Priority 3: Environmental and contextual data (< 100ms latency) (Ksentini and Nikaein, 2020).

These strict latency specifications permit applications such as real-time ball tracking at the matches. In systems like the 5VREAL project, 5G networks and the Edge Computing infrastructure allow processing large volumes of information and reprocessing game actions in real-time to determine the ball position, speed, path, and interaction time with the players (Baruffa et al., 2021). In the case of professional teams, these technologies offer automatic analysis information that was previously only reached by manual surveying, without errors in approximation. In the meantime, the quality of broadcasts is enhanced by the use of enhanced graphics generation, which is concentrated in the cloud architecture, and implementation is easy among television producers (Baruffa et al., 2021).

#### **5.4 Visualization of Real-Time Sports at the System Architecture.:**

The back-end hardware of real-time sports visualization is a highly coordinated system architecture that balances between computing power, network capabilities and rendering technologies. Sports facilities are becoming complex digital spaces in which thousands of data points have to be processed in real-time in order to create immersion (Zhou et al., 2023).

#### **5.5 Multi-access Edge computing (MEC) in Arenas:**

Multi-access Edge Computing (MEC) is the foundation of the state-of-the-art sport visualization systems as it places the computing power on the infrastructure of the stadium. In contrast to conventional cloud-based solutions, MEC decreases congestion in the network through processing data locally in base stations and aggregation points (Mao et al., 2021). This spatial-based type of computing is critical in controlling the huge data volumes produced at sporting events.

Stadium deployments are normally characterized by planned distribution of edge nodes within the venue. As an example, at the 2022 FIFA world cup in Qatar, Ooredoo collaborated with Dell technologies and Ericsson to integrate all eight stadiums with edge computing. The systems received around 400,000 data points within a second displayed on twelve cameras that monitored 29 separate parts of the body of each player (Dell Technologies, 2022).

Stadia with MEC capabilities possess a number of important benefits:

- Minimized operational expenses by way of minimized data backhaul.
- Improved protection of confidential sporting information.
- Quick implementation of new services to the audience (Mao et al., 2021)

Besides, these systems also enable the use of context-aware applications that use information regarding local Radio Access Network (RAN) conditions and subscriber data to optimize operations (Giordani et al., 2021).

#### **Data Flow: Sensor to Edge Node to Viewer.**

The process of sports visualization information has a complex path starting with its original capture and ending with its ultimate presentation. First, the data is gathered with the help of various sensors such as cameras, wearables, smart equipment, and environmental devices (Chen et al., 2022). Indicatively, intelligent rugby balls invented by Sportable and Gilbert monitor such measures as distance of kick and hang time (Dell Technologies, 2022).

This raw sensor data is then transmitted using special protocols that are created to allow real time transmission. Fast systems are using such technologies as MQTT or HTTP/2 to provide secure delivery with minimum latency (Wang et al., 2022). On intermittently connected environments, data is stored locally on systems such as Pravega Sensor Collector and synchronized on the reconnecting of connections (Pravega, 2021).

When at edge nodes, the data is multi-stage processed:

- Local filtering and aggregation at local MEC servers.
- Special AI player-tracking and statistical extraction analysis.
- Reproduction of graphic materials and incorporation of broadcast feeds.

Delivery to a variety of destinations such as broadcaster systems, in-stadium displays, and mobile devices (Dell Technologies, 2022).

It is noteworthy that this architecture supports the size of data more than 1TB per session in certain implementations, as in Formula 1 racing there McLaren integrates more than 300 sensors in each car (Dell Technologies, 2022).

#### **5.6 Real-Time video rendering with GPU Acceleration:**

It makes use of Graphics Processing Unit (GPUs) which can deliver the horsepower required to analyze and render video in real-time. Inside stadium settings, the NVIDIA RTX technology has proven to be useful in processing high-resolution video streams that have low latency (NVIDIA, 2021).

Being based on the RTX A6000 and supplemented by such technologies as NVIDIA G-SYNC and NVIDIA Mosaic,

the device allows the simultaneous coordination of display systems without visual artifacts (NVIDIA, 2021). This is a GPU-accelerated infrastructure which breaks a number of performance records:

- Joint CPU-GPU decoding makes the central processors less loaded (NVIDIA, 2021).
- Real-time ray tracing by using the Deep Learning Super Sampling (DLSS) offered by NVIDIA (Schied et al., 2022)

[human]>Supported ultra-high-definition output of up to 15,360 x 4,320 resolution (dual 8K) (NVIDIA, 2021)

To transfer the data efficiently and effectively between cameras and GPUs, the technologies such as NVIDIA GPUDirect bypass the CPU which is one of the significant time-consuming factors. In more advanced applications, rendering clusters use a combination of multiple GPUs in parallel, each RTX 6000 Ada Generation GPU usually managing four camera feeds (Zhou et al., 2023).

This architecture of acceleration forms the basis of a new visualization such as augmented reality overlay and instant replays without making it obvious to the viewers that a lot of hatching is actually being done in the background to make it look what seems to be seamless.

### 5.7 Important Use Cases Enabled by Edge AI and 5G:

The presence of Edge AI and 5G is transforming the sports industry with the ability to create powerful applications that can change the way sports are consumed, analyzed, and enjoyed. Their low latency, high bandwidth and distributed computing features enable application processes that have not been made possible by traditional infrastructure in the past (Zhou et al., 2023).

#### Video Processing Edge-based Instant Replay:

One great improvement of video analysis is edge-based replay systems. As an illustration, EDGE Sideline Replay is designed to connect to already available high-definition cameras and capture replays straight to edge devices, providing the replays on tablets within 2-3 seconds without the need to connect to the internet (Hudl, 2022). This fast feedback gives coaches the chance to make tactical changes on the fly.

There is also simultaneous recording of multiple camera angles with EDGE Recorders and this enables the coach to have control of slow-motion playback and frame-by-frame analysis. After the games are over, they can be uploaded to services such as Hudl or QwikCut, and the workflow after the match becomes easier (Hudl, 2022).

#### Live player with mmWave 5G AI models:

The 5VREAL project shows how the high-capacity mmWave spectrum of 5G, with AI models, are able to accurately track athletes and equipment on-the-fly. The systems can handle huge amounts of data such as ball movement, speed, and the location of players more accurately than ever before (Khan et al., 2022). This removes errors in manual approximations and gives the coach immediate, data-driven information (Bana et al., 2021).

#### Augmented Reality overlays in Fans in Stadium:

The applications of in-stadium augmented reality (AR) are based on the low latency and high bandwidth of 5G. AR glasses and smartphones allow fans to have access to real-time player statistics, holographic overlays, and interactive menus (NFL, 2022). To give an example, the Ultra Toss promotion by NFL featured fans playing an online football toss, simulated in 3D in the stadium (NFL, 2022).

This is supported by private 5G networks inside arenas: these networks assign tens of thousands of spectators their own bandwidth, with a latency down to sub-milliseconds required to support AR overlays (Giordani et al., 2021).

#### Coach real-time Biometric Feedback:

5G and edge AI wearable sensors enable coaches to monitor the biometrics of the players, including fatigue, exertion, and the risk of injury (Wang et al., 2022). More than half of NHL decides to use cloud-based and edge-based analytics systems to optimize performance and minimise injuries (Statista, 2023). This feedback system that is based on the data improves the team strategy by eliminating the use of intuition to make decisions (Chen et al., 2022).

### 5.7 Problems and Optimization of Techniques:

The implementation of edge AI and 5G in sports creates challenges in the areas of computational limits, bandwidth congestion, and data security. These barriers must be overcome in order to scale adoption.

#### Edge AI Model compression:

AI models are becoming more and more intricate, but edge devices are resource-limited. The training of advanced architectures, such as the OPT-175B, needs almost 1,000 GPUs, whereas smartphone chips offer fewer than 3 TFLOPs of processing capabilities (Zeng et al., 2022). In order to eliminate this gap, three important optimization strategies are used:

Model sparsity- elimination of redundant neural connections (LeCun et al., 1990).

Quantization - floating-point accuracy is lowered to low-bit integers (Jacob et al., 2018).

Knowledge distillation - promoting information in large teacher models to the small student models (Hinton et al., 2015).

Such methods enable higher levels of AI in resource-intensive conditions like edge devices and wearables.

**Bandwidth Distribution in congested Sports Arenas:**

High levels of subscribers mean that stadiums are faced with network saturation, 5G reduces this with mm Wave spectrum and variable TDD setup that can change uplink/downlink capacity based on localized demand (Polese et al., 2021). During super bowl LV8, Verizon and AT&T installed targeted capacity zones, which managed 67 TB of traffic, which is more than the amount of Wi-Fi traffic. The users of Verizon had a download speed that was 2.4 times higher and an upload speed that was 4.8 times higher than the competitors (RootMetrics, 2023).

**Security measures on Edge Data streams:**

In contrast to centralized cloud servers, the edge devices are more physically accessible and they can be attacked (Nguyen et al., 2021). Even as edge computing lowers exposure through a reduced number of long-haul transmissions, well-constructed tamper-resistant protocols are still needed, particularly when working with sensitive biometric or performance-based data (Sun et al., 2022).

**Emerging AI at the Edge and 5G in Sports:**

The sport apps are still under development and innovations in these areas are privacy, processing efficiency and intelligent network management.

**Federated Learning of On-device Training:**

Federated learning allows devices to jointly learn AI models without exchanging raw data to ensure that sensitive biometric data is not disseminated (Kairouz et al., 2021). In sports the health data of the athletes is kept confidential and models enhance across devices distributed.

**Sports analytics Edge-to-Cloud Continuum:**

The current generation of sports analytics is being run in an edge-to-cloud continuum, with edge AI offering low-latency processing, and the cloud offering deep analytics and archival (Wang et al., 2022). Latency and pose estimation of basketball to detect passing tasks were realized at 5.1 ms and 94 percent accuracy in recent implementations with hybrid AI models (Zhou et al., 2023).

**AI-powered 5G Network management:**

Reinforcement learning and decision-tree AI methods optimize 5G processes to increase their efficiency (by up to 92 percent) and reduce reaction times of the network (down to 50 ms) (Zeng et al., 2022). Limitations of the Study All these optimizations aid the high-demand sport.

Although this review offers an integrated overview of the Edge AI and 5G-powered applications to sports visualization, it is important to admit some limitations.

To begin with, the research will be based on mainly published journal articles, conference papers and industry reports in the English language. This creates the possibility of publication bias, with pertinent information on non-English websites or unpublished technical white papers not being considered.

Second, the PRISMA system of methodology was used to guarantee a systematic coverage, however, due to the fast pace of changes in technologies such as the YOLO-based computer vision, WebRTC streaming, and 5G network slicing, some of the existing experimental prototypes and commercial implementations are still not yet researchable in scholarly databases. Because of this, the review can be biased against new use cases that are being experimented in industry-led pilot projects.

Third, this review focuses on the technological aspect (AI models, MEC infrastructure, GPU acceleration, latency benchmarks, and network optimization). Organizational, ethical and socio-economic factors like user acceptance, team/broadcast trade-offs on costs and benefits or the long-term effects on the privacy of the athlete are given less consideration. This is a limited scope that discourages the extrapolation of the results to the wider digital transformation models in sports.

Lastly, although the review incorporates evidence across various fields, such as telecommunications, computer vision, and sports analytics, the vast majority of the studies consider the technologies in that specific environment. Very few of them study their integrated implementation (e.g., YOLO + WebRTC + AR overlays + 5G slicing in a stadium ecosystem). That disjuncture indicates that the suggested system-level implications are theoretical and not empirically tested, and additional testing is needed in the practical settings which involve thousands of simultaneous devices that need to be tied together without fail.

**Conclusion**

The merging of Edge AI and 5G networks has fundamentally changed visualisation of sports, making the processing latency of several seconds down to milliseconds and creating visualization options that had never been possible previously. In this technical guide, we have discussed how Multi-access Edge Computing (MEC) implementation in venues can address the conventional bandwidth limitations due to the data being processed near the source. This architecture can achieve the necessary latency of 1ms that is vital in real-time applications that coaches, broadcasters and fans are currently relying on.

Wearable edge AI processing can provide athletes with real-time feedback without connectivity requirements and 5G

network slicing can build dedicated virtual networks that do not slow down with thousands of simultaneous users. The advanced sensor-to-sensor and sensor-to-view data streaming are capable of processing terabytes of data in a single session, all enabled by the acceleration of GPUs that are used to perform real-time video analysis and presentation. The practical effect of these technologies can be seen in the real-life application. Instant replay systems provide a number of simultaneous camera views to coaching tablets in seconds. mm Wave 5G player tracking removes the approximation errors in performance analysis. AR overlays are created to improve the experiences of stadiums with interactive components and real-time statistics. In the meantime, biometric feedback provides the coaches with an all-time informational knowledge of the player status on the field.

There are always technical problems. Model compression algorithms have to trade between AI complexity and edge device constraints using techniques such as sparsity, quantization, and knowledge distillation. To handle tens of thousands of users at once, stadium facilities need dynamically allocated bandwidth to control connection bandwidth. Also, the security measures should back up confidential athletic information and not to undermine the performance.

These capabilities are likely to be extended in future. Federated learning is a technology that holds greater privacy as it allows training models without providing raw data. The edge-to-cloud continuum provides a medium ground to access instant processing and enhance analysis. In addition, the network management enhanced by AI helps to increase the efficiency of operations and shorten the reaction time in the 5G infrastructure.

## References

1. 3GPP. (2020). Release 16 description. 3rd Generation Partnership Project (3GPP). Retrieved from <https://www.3gpp.org/release-16>
2. Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975. <https://doi.org/10.1016/j.joi.2017.08.007>
3. Alvarez, C., Romero, D., & Garcia, A. (2021). Real-time streaming with WebRTC for interactive applications: A performance evaluation. *IEEE Access*, 9, 14502–14515. <https://doi.org/10.1109/ACCESS.2021.3051234>
4. Baruffa, G., Frescura, F., & Ronga, L. (2021). 5G and edge computing for sports applications: The 5VREAL project. *IEEE Communications Magazine*, 59(6), 88–94. <https://doi.org/10.1109/MCOM.001.2000875>
5. Chen, X., Liu, Y., & Zhang, J. (2022). 5G network slicing and edge intelligence for real-time applications: A survey. *Computer Networks*, 210, 108906. <https://doi.org/10.1016/j.comnet.2022.108906>
6. Chen, X., Zhang, Y., & Xu, W. (2022). Edge AI for wearable devices: A survey of applications and technologies. *IEEE Internet of Things Journal*, 9(12), 9312–9326. <https://doi.org/10.1109/JIOT.2022.3145678>
7. Cheng, K., Yang, S., & Li, H. (2020). Edge computing-based real-time video analytics for sports applications. *IEEE Access*, 8, 185678–185689. <https://doi.org/10.1109/ACCESS.2020.3029876>
8. Chen, X., Zhang, Y., & Xu, W. (2022). Edge AI for wearable devices: A survey of applications and technologies. *IEEE Internet of Things Journal*, 9(12), 9312–9326. <https://doi.org/10.1109/JIOT.2022.3145678>
9. Cheng, Y., Wang, D., Zhou, P., & Zhang, T. (2018). Model compression and acceleration for deep neural networks: The principles, progress, and challenges. *IEEE Signal Processing Magazine*, 35(1), 126–136. <https://doi.org/10.1109/MSP.2017.2765695>
10. Dell Technologies. (2022). Ooredoo, Dell Technologies, and Ericsson deliver edge computing for FIFA World Cup stadiums. Dell Technologies Newsroom. <https://www.dell.com/en-us/dt/case-studies/fifa-world-cup-edge.htm>
11. Giordani, M., Polese, M., Mezzavilla, M., Rangan, S., & Zorzi, M. (2021). Toward 6G networks: Use cases and technologies. *IEEE Communications Magazine*, 59(3), 55–61. <https://doi.org/10.1109/MCOM.001.2000136>
12. Han, S., Mao, H., & Dally, W. J. (2016). Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding. *International Conference on Learning Representations (ICLR)*. <https://arxiv.org/abs/1510.00149>
13. Hu, H., Liu, X., & Wu, J. (2021). Augmented reality applications in sports: Opportunities and challenges. *Journal of Ambient Intelligence and Humanized Computing*, 12(3), 2895–2908. <https://doi.org/10.1007/s12652-020-02280-3>
14. Jocher, G., Chaurasia, A., & Qiu, J. (2022). YOLOv5: A state-of-the-art object detection algorithm. *arXiv preprint arXiv:2207.02696*. <https://doi.org/10.48550/arXiv.2207.02696>
15. Lim, W. Y. B., Xiong, Z., Niyato, D., Yang, Q., & Leung, C. (2022). Realizing the metaverse with edge intelligence: A federated learning perspective. *IEEE Network*, 36(1), 176–183. <https://doi.org/10.1109/MNET.011.2100375>
16. Kairouz, P., McMahan, H. B., & Avent, B. (2021). Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, 14(1–2), 1–210. <https://doi.org/10.1561/22000000083>

17. Ksentini, A., & Nikaein, N. (2020). Toward enforcing network slicing on RAN: Flexibility and resources abstraction. *IEEE Communications Magazine*, 58(7), 102–108. <https://doi.org/10.1109/MCOM.001.2000010>
18. Li, R., Zhao, Z., Sun, Q., & Zhang, C. (2021). AI at the edge: The convergence of edge computing and artificial intelligence in 5G networks. *IEEE Wireless Communications*, 28(2), 34–41. <https://doi.org/10.1109/MWC.001.2000254>
19. Lim, W. Y. B., Luong, N. C., & Hoang, D. T. (2020). Federated learning in mobile edge networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 22(3), 2031–2063. <https://doi.org/10.1109/COMST.2020.2986024>
20. Mao, Y., You, C., Zhang, J., Huang, K., & Letaief, K. B. (2021). A survey on mobile edge computing: The communication perspective. *IEEE Communications Surveys & Tutorials*, 23(1), 1–24. <https://doi.org/10.1109/COMST.2020.3014901>
21. NVIDIA. (2021). NVIDIA RTX A6000 professional GPU: Product brief. NVIDIA Corporation. <https://www.nvidia.com/rtx-a6000>
22. Ozturk, M., Ksentini, A., & Frangoudis, P. A. (2021). 5G network slicing for sports and entertainment: A practical perspective. *IEEE Communications Standards Magazine*, 5(2), 62–68. <https://doi.org/10.1109/MCOMSTD.011.2000031>
23. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
24. Park, J., Samarakoon, S., Bennis, M., & Debbah, M. (2020). Wireless network intelligence at the edge. *Proceedings of the IEEE*, 108(2), 220–235. <https://doi.org/10.1109/JPROC.2019.2951986>
25. Pereira, J., Monteiro, P., & Martins, J. (2022). Immersive media experiences and fan engagement in sports broadcasting. *Telematics and Informatics*, 69, 101791. <https://doi.org/10.1016/j.tele.2022.101791>
26. Pravega. (2021). Pravega sensor collector: Data ingestion for IoT and edge. Dell Technologies Open Source Project. <https://cnf.pravega.io>
27. Schied, C., Kaplanyan, A., Wyman, C., & Salvi, M. (2022). Real-time neural rendering with DLSS. *ACM SIGGRAPH Courses*, 41(4), 1–22. <https://doi.org/10.1145/3532720.3535624>
28. Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. *arXiv preprint arXiv:1804.02767*. <https://doi.org/10.48550/arXiv.1804.02767>
29. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., ... & Xu, D. (2020). The future of digital health with federated learning. *npj Digital Medicine*, 3(1), 119. <https://doi.org/10.1038/s41746-020-00323-1>
30. Taleb, T., Samdanis, K., Mada, B., Flinck, H., Dutta, S., & Sabella, D. (2017). On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration. *IEEE Communications Surveys & Tutorials*, 19(3), 1657–1681. <https://doi.org/10.1109/COMST.2017.2705720>
31. Wang, J., & Li, H. (2022). Smart wearables and AI-based injury prevention in sports. *Sensors*, 22(8), 2985. <https://doi.org/10.3390/s22082985>
32. Xu, J., Huang, T., & Zhang, Y. (2023). Low-latency edge intelligence for immersive media over 5G networks. *IEEE Transactions on Multimedia*, 25, 1234–1248. <https://doi.org/10.1109/TMM.2023.3234567>
33. Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology*, 10(2), 1–19. <https://doi.org/10.1145/3298981>
34. Zhang, C., Patras, P., & Haddadi, H. (2021). Deep learning in mobile and wireless networking: A survey. *IEEE Communications Surveys & Tutorials*, 21(3), 2224–2287. <https://doi.org/10.1109/COMST.2021.3054392>
35. Zhang, H., Liu, N., Chu, X., Long, K., Aghvami, A. H., & Leung, V. C. M. (2020). Network slicing based 5G and future mobile networks: Mobility, resource management, and challenges. *IEEE Communications Magazine*, 58(3), 55–61. <https://doi.org/10.1109/MCOM.001.1900515>
36. Zhou, H., Xu, K., & Wang, T. (2023). Edge AI for 5G-enabled immersive media: Architectures and challenges. *IEEE Communications Surveys & Tutorials*, 25(2), 1250–1278. <https://doi.org/10.1109/COMST.2023.3240123>
37. Zhou, Z., Chen, X., Li, E., Zeng, L., Luo, K., & Zhang, J. (2023). Edge intelligence in 5G: Applications and challenges. *Proceedings of the IEEE*, 111(1), 1–19. <https://doi.org/10.1109/JPROC.2022.3221572>