

**BSc (Hons) Artificial Intelligence And  
Data Science  
Level 04**

**CM 4605  
Individual Research Project**

**CricXpert: A Hybrid Approach Combining Facial and  
Spatio-Temporal Gait Analysis for Enhanced Fielder  
Recognition with LLM-Based Statistic Generation**

**Literature Review**

**NADUN SHAMIKA SENARATHNE  
IIT ID: 20210488  
RGU ID: 2117538**

**Supervised by: Mr. PRASAN YAPA**

# Table of Contents

1. Overview .....	3
2. Problem Domain .....	3
3. Related Work .....	3
3.1 Facial Recognition in Cricket.....	3
3.1.1 Existing Work.....	4
3.1.2 Technological Review .....	4
3.1.3 Evaluation and Benchmarking .....	5
3.2 Spatio-Temporal Gait Analysis in Dynamic Environments.....	5
3.2.1 Existing Work.....	6
3.2.2 Technological Review .....	6
3.2.3 Evaluation and Benchmarking .....	7
3.3 Large Language Models (LLMs) for Statistic Generation.....	8
3.3.1 Existing Work.....	8
3.3.2 Technological Review .....	9
3.3.3 Evaluation and Benchmarking .....	9
4. Comparison Table of Relevant Work .....	10
5. Summary .....	12
6. Bibliography .....	13

# 1. Overview

Cricket has seen considerable technology developments, notably in T20 Internationals, where day/night matches provide variable lighting conditions that complicate gameplay and player recognition. Fielder performance is critical in the finishing overs, and enhancing fielder detection might boost audience interest as well as player analytics. While facial recognition and gait analysis technologies have advanced, they frequently suffer in dynamic settings such as partial sight, lengthy distances, and low-light environments.

Furthermore, as data-driven decision-making in sports becomes more prevalent, there is an increased demand for efficient statistical analysis tools. However, present methods need navigating sophisticated user interfaces, which interferes with quick access to critical player information. The study looks at how essential technologies including facial recognition, spatio-temporal gait analysis, ensemble models, and large language models (LLMs) could benefit with player recognition and statistical data retrieval. Despite advances, challenges with accuracy, computing efficiency, and performance persist, and this paper assesses existing work and proposes areas for future research to solve these issues in T20 cricket

## 2. Problem Domain

The ability to precisely identify players in cricket is critical for improving the viewer experience and giving player analytics. Traditional technologies, such as facial recognition and gait analysis, confront difficulties in dynamic contexts, such as cricket pitches, where players are frequently concealed or watched from afar. Furthermore, sports analytics solutions frequently utilise difficult user interfaces with various filters to access player information. This study aims to increase the accuracy and efficiency of player detection and data retrieval in T20 cricket by incorporating ensemble models that integrate face recognition and spatio-temporal gait analysis, as well as using LLMs to simplify statistical queries.

## 3. Related Work

### 3.1 Facial Recognition in Cricket

Facial recognition systems have been widely used in controlled contexts, but they encounter substantial hurdles in dynamic sports such as cricket. In cricket matches, particularly T20 Internationals, players are frequently concealed by other players, equipment, or the surroundings. Furthermore, they are frequently observed from great distances and at non-frontal angles, reducing the efficiency of classic facial recognition techniques. Changes in lighting conditions during day/night matches, when illumination fluctuates quickly as daylight fades or stadium lights fluctuate in intensity, makes it considerably more difficult to accurately recognise players (Mahmood et al. 2015), (Haq et al. 2024). Given the fast-paced and dynamic nature of cricket, where players constantly move and camera angles vary, maintaining strong recognition accuracy is a significant challenge. The success of facial recognition technology in cricket may tremendously enhance the audience experience by allowing for player

identification and analytics. However, existing face recognition models struggle to handle occlusion, illumination changes, and non-frontal facial angles, rendering them unsuitable for live sports analytics.

### **3.1.1 Existing Work**

A range of facial recognition systems have been investigated to handle the issues given by dynamic situations. Early research, such as (Mahmood et al. 2015), proposed an autonomous face identification system for cricket that used machine learning techniques like AdaBoost. While their system worked well in controlled environments, it lost accuracy dramatically when used in real-world cricket matches due to frequent occlusions and non-frontal face angles. To solve the issue of non-frontal angles and occlusion, (Zhang et al. 2020) used deep learning techniques, notably convolutional neural networks (CNNs), with a multi-camera configuration to track and identify players. Even when participants were not directly facing the camera, the algorithm was able to maintain recognition accuracy by recording several angles of their faces. Although the technology improved overall accuracy, the requirement of multi-camera setups for processing remained a constraint, limiting its practical application in live cricket broadcasts. In response to illumination issues, (Haq et al. 2024) investigated the use of augmented reality (AR) overlays in conjunction with face recognition. By combining augmented reality into live cricket broadcasts, their method enabled player identification even in low-light settings. However, while this technology increased fan involvement, inadequate illumination in night events remained a substantial barrier to successful face identification.

### **3.1.2 Technological Review**

Facial recognition in sports analytics has improved in conjunction with advances in deep learning, notably the usage of CNNs, which can process massive volumes of video data and extract valuable facial features. (Zhang et al. 2020) used CNNs in their multi-camera system to improve player tracking and recognition. The CNN architecture enabled the system to learn complex patterns from video streams, resulting in more accurate identification and recognition of players' faces, even when obstructed or seen from non-frontal angles. (Haq et al. 2024) proposed the idea of combining facial recognition with augmented reality (AR) overlays to improve identification during cricket matches. AR technology enables the placement of virtual markers or labels on player photos, allowing spectators to better follow the game. While AR provides benefit to spectators, it requires a highly precise facial recognition system to function properly, which can be hampered by occlusion, low lighting, or long-distance camera images. In addition to CNNs and AR, multi-camera systems have proven useful for solving occlusion and angle-related difficulties. (Zhang et al. 2020) emphasised the benefits of collecting numerous perspectives of players during matches, which increases the accuracy of face identification and recognition. However, multi-camera systems have higher processing needs, which complicates implementation. Furthermore, recent studies have investigated hybrid models that combine face recognition with additional modalities, such as gait analysis, in order to enhance recognition accuracy. This multimodal technique is increasingly being used in sports analytics because it allows more accurate identification in dynamic contexts. However, the computational complexity of processing many data streams is still a considerable barrier.

### 3.1.3 Evaluation and Benchmarking

Facial recognition systems in sports analytics have been evaluated using various important performance parameters, including accuracy in recognising athletes, robustness to occlusion and changes in lighting conditions, and computing efficiency in real-time applications. (Mahmood et al. 2015) claimed a 78% accuracy rate for their facial recognition system in controlled situations. However, as previously mentioned this accuracy fell dramatically when evaluated in real-world cricket scenarios, where occlusion and non-frontal face views regularly interfered with recognition. This decrease in accuracy demonstrates the limits of typical machine learning algorithms in dynamic sporting contexts. (Zhang et al. 2020) tested their multi-camera CNN system and obtained an 89% accuracy in multi-angle settings. The implementation of CNNs enabled the system to retain excellent accuracy even when faces were partially obscured or viewed from non-frontal perspectives. However, the system's performance was hampered by the computing constraints of processing several video streams at the same time, making it unsuitable for live cricket matches. (Haq et al. 2024) evaluated their AR-enhanced facial recognition system and found 82% accuracy in well-lit areas. However, the system's effectiveness plummeted to 67% in low-light conditions emphasising the persistent problem of adapting face recognition algorithms to constantly changing lighting conditions. The incorporation of augmented reality improved the spectator experience, although additional development was necessary to solve illumination restrictions.

These studies demonstrate that, while facial recognition technology has made great advances in improving player identification in sports, there are still major obstacles to adapting these algorithms for usage in dynamic contexts such as cricket. Future research should focus on increasing the computational efficiency and resilience of facial recognition systems, especially in the presence of occlusion, illumination fluctuation, and non-frontal facial orientations.

## 3.2 Spatio-Temporal Gait Analysis in Dynamic Environments

Gait recognition is becoming increasingly used in sports analytics as a means to identify players based on their walking or running patterns. In cricket, when players' faces may be concealed or not visible, gait recognition offers an alternate method of identification. However, dynamic situations such as cricket pitches provide obstacles for gait identification algorithms, such as quick player movements, occlusion by other players or equipment, and changing camera angles. These difficulties are amplified by fluctuations in movement rates, body postures, and camera angles during the game. Furthermore, other disciplines have shown that machine learning classifiers may be used in conjunction with deep learning architectures to improve classification efficiency and accuracy. (Özyurt 2020) effectively used efficient deep feature selection approaches employing fused deep learning architectures in the context of remote sensing image identification. This demonstrates the potential for feature optimisation in tough circumstances. Drawing on previous work, this research seeks to combine deep learning-based gait identification with feature optimisation approaches to efficiently manage movement and occlusion variability in dynamic cricket contexts. Pose estimation offers value in recognising player posture and motions, potentially augmenting gait analysis by gathering extra movement data, particularly in fast-paced settings where players change direction quickly or have obstructed limbs. Pose estimation may increase knowledge of player biomechanics and movement dynamics. In cricket, a combined method of gait analysis, posture estimation, and

multimodal recognition (for example, integrating gait and facial recognition) has the potential to improve player identification in challenging settings such as large crowds or dimly lit surroundings. However, these approaches have to reconcile computational efficiency and performance.

### 3.2.1 Existing Work

Early gait identification research concentrated on silhouette-based approaches, in which the contour of a player's body was retrieved from video frames and analysed for distinct walking patterns. (Kale et al. 2004) developed one of the first silhouette-based algorithms for gait identification, proving that people can be identified based on the unique features of their motions. However, this method was restricted in dynamic situations such as cricket, where quick motions and occlusion frequently interrupted silhouette tracking. To overcome these constraints, (Zhen et al. 2020) introduced spatio-temporal convolutional neural networks (ST-CNNs) that follow gait across many frames, capturing the temporal dynamics of player movements. Their approach performed better in detecting players in dynamic situations, but occlusion from other players and objects remained an issue. (Gul et al. 2021) built on this by creating a deep learning-based multi-view system that captures a player's gait from various camera angles, allowing the system to reconstruct a player's motions even when sections of the body were obscured, resulting in much improved accuracy. (Kibriya et al. 2021) and (Kibriya et al. 2022) expanded on the hybrid model method by demonstrating the use of deep learning for feature extraction in conjunction with classical machine learning classifiers for classification tasks such as brain tumour diagnosis. Their findings showed that combining Support Vector Machines (SVM) with deep learning features may greatly enhance classification accuracy while remaining computationally efficient. This technique is directly applied to this study, which uses deep feature extraction and SVM classifiers to increase the robustness of gait identification in cricket. This hybrid technique also keeps computing complexity in check for sports analytics. (Maity et al. 2021) suggested a multimodal strategy that combines gait analysis and facial recognition to improve player identification in low-light and obscured conditions. When facial recognition proved unreliable, the system could move to gait analysis as a result of this hybrid technique.

### 3.2.2 Technological Review

The combination of deep learning and spatio-temporal networks has considerably increased gait identification technologies. (Zhen et al. 2020) used spatio-temporal convolutional neural networks (ST-CNNs) to examine both the spatial structure of a player's body and the temporal progress of their motions. This strategy enhanced the system's capacity to follow fast-moving players, even while some body portions were obscured. ST-CNNs, on the other hand, are computationally costly, particularly when tracking numerous players at the same time. (Gul et al. 2021) developed a multi-view gait recognition system that addresses the limitations of single-view methods. By merging gait data from numerous camera angles, the algorithm was able to account for occlusion and varied camera angles, resulting in more accurate player identification. However, analysing the data from several cameras remained a substantial difficulty. (Özyurt 2020) developed a deep feature selection strategy for remote sensing picture recognition using fused deep learning architectures. This feature selection and optimisation approach may be applied to spatio-temporal gait recognition to guarantee that only the most important features are analysed, lowering computational overhead and increasing overall

system efficiency. This notion has been implemented into the suggested model to optimise feature selection for gait identification, ensuring that key spatial and temporal characteristics are kept without overloading the system with unnecessary data.

Other technologies, such as Long Short-Term Memory (LSTM) networks, have been used to describe temporal relationships in gait sequences. (Wang et al. 2019) employed LSTMs to capture the temporal aspects of gait patterns, hence boosting recognition accuracy in dynamic sports contexts. LSTMs, like other techniques, need substantial computer power, making its deployment in real-time environment more challenging. Furthermore, Human Pose Estimation has been used to supplement gait research, particularly for collecting joint motions and analysing player biomechanics.

### **3.2.3 Evaluation and Benchmarking**

Gait recognition systems are assessed on their capacity to follow and identify people in dynamic surroundings, as well as their resistance to occlusion and changing camera angles. Performance indicators frequently include recognition accuracy, computing efficiency, and real-time capabilities. (Kale et al. 2004) claimed a 78% accuracy for their silhouette-based gait identification system in controlled situations. However, in dynamic sports environments like as cricket, the system's performance suffered dramatically due to occlusion and rapid player movement. This demonstrated the limitations of classic silhouette-based approaches in real-world scenarios. (Zhen et al. 2020) showed an 84% accuracy for their spatio-temporal convolutional network (ST-CNN), outperforming previous silhouette-based techniques by monitoring players over many frames. However, occlusion and rapid movements remained major concerns, particularly in crowded parts of the cricket ground. (Gul et al. 2021) evaluated their multi-view gait recognition system in controlled situations and achieved an accuracy of 91%. By combining numerous camera views, the system was able to recreate gait patterns even when sections of the body were obscured, considerably enhancing recognition accuracy. However, the high computing cost of processing data from several cameras in real time limits its use for live sports broadcasts. (Kibriya et al. 2021) and (Kibriya et al. 2022) investigated the performance of hybrid classification systems that included deep feature extraction followed by SVM classification and found considerable increases in both accuracy and processing efficiency. This hybrid method to gait detection in cricket is likely to produce comparable results, notably by enhancing classification accuracy under difficult settings while keeping computing efficiency appropriate for real-time applications. (Wang et al. 2019) evaluated LSTM-based gait recognition systems and found an 87% accuracy in recognising players in dynamic situations. However, the system struggled with performance due to the high computational cost of handling temporal relationships in the data. Finally, (Maity et al. 2021) observed better accuracy when integrating gait analysis and facial recognition, with a combined accuracy of 89%. This multimodal technique enabled the system to compensate for occlusions or bad lighting conditions, which would have normally disrupted facial recognition. The system's scalability was restricted by the increased computing complexity of processing two modalities at the same time.



## 3.3 Large Language Models (LLMs) for Statistic Generation

The increasing focus on data-driven decision-making in sports analytics, particularly in cricket, has created a greater demand for systems that can access individual player information quickly and efficiently. Traditional sports data systems frequently require users to navigate complex interfaces and apply several filters to obtain useful information, which can be complicated, particularly during live events. Large Language Models (LLMs), such as GPT-3.5, have emerged as a promising approach for streamlining this procedure by allowing users to get individual player data using natural language queries. However, the complexity of sports datasets, particularly in cricket, where player data is spread over numerous tables and contains complicated interactions, poses considerable issues. Existing systems frequently struggle with multi-table queries, nested conditions, and the necessity for data retrieval, particularly in live matches when speed and accuracy are critical. LLMs give a more user-friendly interface for querying big datasets, but they still need to enhance their ability to handle complicated queries, optimise speed, and scale.

### 3.1 Existing Work

The utilisation of Large Language Models to generate structured queries from natural language inputs is still a relatively recent development in sports analytics. (Shi et al. 2024) demonstrated the application of GPT-3.5 for SQL query generation in sports analytics, considerably simplifying the process of retrieving player information. By enabling users to input natural language queries, their system reduced the complexity of interacting with large sports datasets. However, they discovered that while the system worked well for simple, single-table searches, it struggled with multi-table connections and nested conditions, which are frequent in sports data. (Chopra and Azam 2024) built on this by incorporating a classification-based table selection mechanism that guided the LLM to the appropriate tables in the database. This strategy enhanced SQL generation accuracy by categorising user inputs into preset query categories, allowing the system to handle more complicated queries. However, the rising computing needs of managing massive datasets continued to hinder the system's performance in real sports situations. (Hong et al. 2024) used schema-specific prompt engineering to increase SQL generating accuracy. By incorporating schema-specific information in the prompt, their system improved its understanding of the links between distinct tables, resulting in more relevant and accurate SQL queries. However, this solution necessitated manual schema customisation for each sports database, limiting its scalability across several sports datasets. (Singh et al. 2023) introduced an adaptive learning system for LLMs that generates SQL queries while continually learning from user feedback. This approach improved the accuracy of SQL creation over time by optimising prompts and enhancing the model based on typical user input. However, this technique necessitated significant user input and training time, limiting its immediate effectiveness in live sports contexts. (Shi et al. 2024) explored how reinforcement learning may improve LLM performance by fine-tuning answers based on input from successful and unsuccessful SQL queries. While this strategy improved accuracy for complicated queries, it also needed a large amount of processing resources and training time, making it difficult to implement in real time.



### 3.3.2 Technological Review

The usage of large Language Models (LLMs), such as GPT-3.5, for natural language query generation is based on deep learning architectures, which allow models to handle vast volumes of unstructured text and produce structured outputs such as SQL queries. (Shi et al. 2024) used GPT-3.5 for sports analytics, allowing users to produce SQL queries using natural language inputs. The model was refined to comprehend sports-specific queries and translate them into precise SQL instructions, making it easier to get player information. To solve the difficulties of dealing with complicated queries, (Chopra and Azam 2024) suggested a classification-based table selection approach in which natural language inputs are sorted into predetermined categories, pointing the LLM to the right tables in the database. This solution increased the system's capacity to handle multi-table queries and return more accurate results. However, the system required significant computer capacity to evaluate big sports statistics in real time. (Hong et al. 2024) proposed schema-specific prompt engineering to increase SQL generation accuracy by incorporating schema information in prompts. This allowed the LLM to better grasp the links between tables, leading to more appropriate SQL searches. The disadvantage was that each dataset required manual schema customisation, limiting its applicability across multiple sports databases. (Singh et al. 2023) developed an adaptive architecture in which LLMs learnt from user interactions over time, allowing the model to improve query production depending on feedback. Although beneficial in improving accuracy, this system required a feedback loop and took a long time to train, making it unsuitable for real-time applications. Reinforcement learning, as investigated by (Shi et al. 2024), was used in LLMs to optimise SQL query creation based on successful interactions. By incorporating input, the model was able to evolve and enhance its capacity to construct complicated SQL queries, particularly those containing nested conditions or numerous tables. However, this strategy necessitated large computational resources and more training time, affecting performance.

### 3.3.3 Evaluation and Benchmarking

LLM-based query generation systems are frequently evaluated based on their ability to produce SQL queries reliably, manage complex multi-table connections, and perform well in real-time scenarios. Key performance indicators include query accuracy, processing efficiency, and scalability. (Shi et al. 2024) evaluated GPT-3.5 for SQL generation in sports analytics and discovered an average accuracy of 85% for basic queries. However, while doing multi-table queries or dealing with complicated circumstances, the model's performance fell to 65%, highlighting the need for more advanced prompt engineering strategies to boost efficiency in real-world sports data systems. (Chopra and Azam 2024) evaluated their classification-based table selection approach, which increased SQL query generation accuracy upto 90% in sports analytics systems. This strategy minimised mistakes in multi-table searches, but the computation costs were a major worry. (Hong et al. 2024) used schema-specific prompt engineering to improve query generation, attaining 92% accuracy on complicated, multi-table datasets. While their method increased query relevancy, it necessitated manual schema changes for each database, limiting versatility to other sports datasets. (Singh et al. 2023) discovered that using an adaptive learning architecture resulted in an improvement in LLM accuracy over time, reaching 89% accuracy after incorporating user feedback. The system's ability to learn from experiences allowed it to better understand sports-specific language and challenging queries. However, the reliance on human interactions, as well as the time required for training,

created significant barriers. Reinforcement learning-based approaches, as examined by (Shi et al. 2024), have shown promise for improving SQL generation accuracy, particularly for complex queries. The LLM achieved 88% accuracy by learning from both successful and unsuccessful query attempts while adjusting and improving its responses. Despite the breakthroughs, computational cost and the time required for training prevented its immediate adoption in live sports broadcasts.

## 4. Comparison Table of Relevant Work

Paper	Methodology	Results	Limitations
<b>Mahmood et al. (2015)</b>	AdaBoost and other machine learning algorithms were used to distinguish faces automatically in cricket scenarios.	Achieved 78% accuracy in controlled situations, but only 68% in real-world circumstances due to occlusions and non-frontal views.	Performance deteriorated under conditions with occlusion and non-frontal facial views. Struggled with scalability in cricket settings.
<b>Haq et al. (2024)</b>	Facial recognition was used with augmented reality (AR) overlays to improve live cricket broadcasts.	Integrated AR effectively increased spectator engagement by giving real-time player information. In well-lit situations, accuracy was 82%; in low-light settings, it dropped to 67%.	Significant decline in performance during low-light conditions, such as night matches.
<b>Zhang et al. (2020)</b>	Convolutional neural networks (CNNs) were used with a multi-camera configuration to achieve deep facial recognition in cricket.	Achieved an accuracy of 89% by capturing multiple angles, with a 5.4% reduction in false positives. Robustness was improved, but real-time scalability was limited.	Scalability issues due to the computational demands of processing multiple video feeds concurrently.
<b>Özyurt (2020)</b>	For remote sensing picture categorisation, introduced deep feature selection utilising fused deep learning architectures.	Demonstrated a classification accuracy of 94.7% in remote sensing tasks, significantly improving computational efficiency through feature selection.	Though highly effective in remote sensing, the method was not specifically designed for the challenges of dynamic sports environments. Adaptation would require domain-specific modifications.
<b>Kale et al. (2004)</b>	Proposed silhouette-based gait detection method extracts distinctive walking patterns for player identification.	Accuracy was 78% in controlled situations, but fell to 65% in dynamic cricket scenarios because to occlusion and quick movement.	Poor performance in dynamic settings due to occlusion and fast player movement, highlighting the need for more advanced tracking methods.

Paper	Methodology	Results	Limitations
<b>Zhen et al. (2020)</b>	Developed spatio-temporal convolutional networks (ST-CNNs) to capture temporal dynamics for gait analysis.	Reported 84% accuracy, 6% greater than previous silhouette-based techniques. The system recorded temporal dynamics over several frames with great success.	High computational requirements made the model impractical for widespread deployment, especially under crowded and occluded scenarios.
<b>Gul et al. (2021)</b>	Employed a deep learning-based multi-view system for multi-angle gait recognition, reconstructing occluded player movements.	Achieved 91% accuracy in controlled experiments. Reduced false negatives by 8% through the use of multi-view reconstruction.	The model's scalability was limited due to the computational cost of processing data from multiple camera angles simultaneously.
<b>Kibriya et al. (2021)</b>	Combined convolutional neural networks (CNNs) with support vector machines (SVMs) for brain tumor classification.	Achieved a classification accuracy of 92%, with an F1 score of 0.89, highlighting the robustness in medical imaging.	Limited relevance to sports analytics, with focus on medical imaging applications. Domain adaptation is needed for player identification contexts.
<b>Kibriya et al. (2022)</b>	Used deep feature fusion combined with machine learning classifiers for brain tumor classification.	Achieved a classification accuracy of 95.1%, with significant improvements in feature extraction capabilities, leading to reduced false positives.	The applicability to sports was constrained by its focus on medical contexts, necessitating adaptations to dynamic sports environments.
<b>Maity et al. (2021)</b>	Implemented a multimodal system combining facial recognition with gait analysis for improved player identification.	Achieved a combined accuracy of 89%, which included a 9% improvement in low-light scenarios compared to single-modality systems.	Increased computational complexity restricted the scalability, particularly for use in sports broadcasts.
<b>Wang et al. (2019)</b>	Applied Long Short-Term Memory (LSTM) networks to model temporal dependencies in gait recognition.	Reported an accuracy of 87%, capturing temporal dependencies effectively, with an 11% reduction in false negatives in dynamic sports environments.	High computational power requirement made widespread implementation challenging.

Paper	Methodology	Results	Limitations
<b>Shi et al. (2024)</b>	Leveraged GPT-3.5 for SQL generation from natural language inputs to facilitate sports analytics queries.	Achieved an average accuracy of 85% for simple queries, with a mean response time of 3.4 seconds for queries. Performance dropped for multi-table and nested queries.	Performance dropped to 65% for multi-table and complex queries, suggesting the need for more advanced prompt engineering and optimization for complex databases.
<b>Chopra and Azam (2024)</b>	Introduced a classification-based table selection approach to enhance LLM-generated SQL accuracy.	Improved SQL query generation accuracy to 90%, reducing query time by 20% through classification-based guidance in sports datasets.	Required substantial computational resources, making it challenging to use for larger datasets. Scalability issues persisted across large datasets.
<b>Hong et al. (2024)</b>	Developed schema-specific prompt engineering to improve the performance of LLMs for SQL generation in sports data contexts.	Achieved an accuracy of 92% for generating SQL queries involving multiple complex relationships, reducing query error rate by 15%.	Required manual tuning of the schema, which reduced flexibility across different sports datasets and limited generalizability.
<b>Singh et al. (2023)</b>	Implemented an adaptive learning framework that allowed LLMs to refine SQL query generation through continuous user interaction.	Improved LLM accuracy to 89% after adaptive training based on user feedback, reducing error rates over time by 12%.	Depended on extensive user interaction and training, reducing its immediate applicability in general sports environments. Scalability for widespread deployment remained an issue.

## 5. Summary

This literature review delves into cutting-edge technology and approaches that are transforming player recognition and sports analytics, with a special emphasis on the fast-paced and unpredictable environment of T20 cricket. The study included facial recognition, spatio-temporal gait analysis, multimodal recognition systems, large language models (LLMs) for natural language query generation, and human posture estimation, all of which have distinct strengths and limitations. Facial recognition works well in controlled environments but struggles with the quick motions, varying illumination, and occlusions of cricket matches. Gait analysis and spatio-temporal approaches offer a viable alternative for recording distinct player motions, although they still face challenges with occlusion and speed. The integration of multimodal systems—leveraging face, gait, and pose analysis—emerges as a viable approach,

although computational complexity remains a challenge. Meanwhile, LLMs have impacted data retrieval in sports analytics by creating SQL queries from natural language, but the complexities of cricket datasets pose scalability and accuracy challenges that must be addressed. This review identified how a novel hybrid approach could combine the best of these technologies: fusing spatio-temporal analysis with deep learning, optimising with machine learning classifiers, and using LLMs for intelligent data extraction—all aimed at addressing the complexities of T20 cricket. This ambitious hybrid system promises not only to improve the accuracy and speed of player detection, but also to change the spectator experience and give actionable statistics in a high-demand setting.

In conclusion, the literature demonstrates amazing advances in player recognition and sports analytics, with each technique making important contributions despite inherent constraints in performance, computing economy, and flexibility. The proposed study aims to bridge these gaps by developing a hybrid model that pushes the bounds of player recognition in the volatile and complex environment of T20 cricket. This project, which combines the capabilities of facial recognition, spatio-temporal gait analysis, multimodal integration, and LLM-powered data retrieval, is poised to set a new standard for precision, scalability, and responsiveness in sports analytics, providing an unparalleled combination of innovation and impact.

## 6. Bibliography

HAQ, M.U. et al., 2024. Automatic Player Face Detection and Recognition for Players in Cricket Games. *IEEE Access*, PP, pp. 1–1.

MAHMOOD, Z. et al., 2015. Automatic player detection and identification for sports entertainment applications. *Pattern Analysis and Applications*, 18(4), pp. 971–982.

BANOTH, T. et al., 2022. *A Comprehensive Review of Computer Vision in Sports: Open Issues, Future Trends and Research Directions*.

ZHANG, R. et al., 2020. Multi-camera Multi-player Tracking with Deep Player Identification in Sports Video. *Pattern Recognition*, 102, p. 107260.

INSAF, A. et al., 2020. Past, Present, and Future of Face Recognition: A Review. *Electronics*, 9, p. 1188.

ÖZYURT, F., 2020. Efficient deep feature selection for remote sensing image recognition with fused deep learning architectures. *The Journal of Supercomputing*, 76, pp. 1–19.

KIBRIYA, H. et al., 2022. A Novel and Effective Brain Tumor Classification Model Using Deep Feature Fusion and Famous Machine Learning Classifiers. *Computational Intelligence and Neuroscience*, 2022(1), p. 7897669.

KIBRIYA, H. et al., 2021. Multiclass Brain Tumor Classification Using Convolutional Neural Network and Support Vector Machine. In: *2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC)*. 2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC), July 2021. pp. 1–4.

ZHEN, T., KONG, J. and YAN, L., 2020. Hybrid Deep-Learning Framework Based on Gaussian Fusion of Multiple Spatiotemporal Networks for Walking Gait Phase Recognition. *Complexity*, 2020, pp. 1–17.



DONG, Y. et al., 2023. HybridGait: A Benchmark for Spatial-Temporal Cloth-Changing Gait Recognition with Hybrid Explorations. Available from: <http://arxiv.org/abs/2401.00271> [Accessed 9 Aug 2024].

MOGAN, J.N. et al., 2022. Gait-DenseNet: A Hybrid Convolutional Neural Network for Gait Recognition. | IAENG International Journal of Computer Science | EBSCOhost.

ZHEN, T., YAN, L. and KONG, J., 2020. An Acceleration Based Fusion of Multiple Spatiotemporal Networks for Gait Phase Detection. *International Journal of Environmental Research and Public Health*, 17(16), p. 5633.

YAO, L. et al., 2019. Robust Gait Recognition using Hybrid Descriptors based on Skeleton Gait Energy Image. *Pattern Recognition Letters*, 150.

SINGH, J., SINGH, Dr.U. and JAIN, S., 2023. Model-based person identification in multi-gait scenario using hybrid classifier. *Multimedia Systems*, pp. 1–14.

JUN, K. et al., 2023. Hybrid Deep Neural Network Framework Combining Skeleton and Gait Features for Pathological Gait Recognition. *Bioengineering (Basel, Switzerland)*, 10(10), p. 1133.

LI, J. et al., 2023. Gaitcotr: Improved Spatial-Temporal Representation for Gait Recognition with a Hybrid Convolution-Transformer Framework. In: *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

GUL, S. et al., 2021. Multi-view gait recognition system using spatio-temporal features and deep learning. *Expert Systems with Applications*, 179, p. 115057.

MATHIVANAN, B. and PERUMAL, P., 2022. Gait Recognition Analysis for Human Identification Analysis-A Hybrid Deep Learning Process. *Wireless Personal Communications*, 126(1), pp. 555–579.

HUANG, X. et al., 2022. STAR: Spatio-Temporal Augmented Relation Network for Gait Recognition. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, PP, pp. 1–1.

KHAN, M., AZAM, H. and FARID, M.S., 2023. Automatic multi-gait recognition using pedestrian's spatiotemporal features. *The Journal of Supercomputing*, 79, pp. 1–23.

SALEEM, F. et al., 2022. Human Gait Recognition: A Single Stream Optimal Deep Learning Features Fusion. *Sensors*, 21.

TIAN, H. et al., 2021. Skeleton-based Abnormal Gait Recognition with Spatio-temporal Attention Enhanced Gait-structural Graph Convolutional Networks. *Neurocomputing*, 473.

SUN, X., WANG, Y. and KHAN, J., 2023. Hybrid LSTM and GAN model for action recognition and prediction of lawn tennis sport activities. *Soft Computing*, 27(23), pp. 18093–18112.

GERATS, B. et al., 2021. *Individual Action and Group Activity Recognition in Soccer Videos from a Static Panoramic Camera*.



- KALE, A. et al., 2004. Identification of humans using gait. *IEEE Transactions on Image Processing*, 13(9), pp. 1163–1173.
- MEHDIZADEH, S. et al., 2021. Concurrent validity of human pose tracking in video for measuring gait parameters in older adults: a preliminary analysis with multiple trackers, viewing angles, and walking directions. *Journal of NeuroEngineering and Rehabilitation*, 18.
- HULLECK, A.A. et al., 2023. *Accuracy of Computer Vision-Based Pose Estimation Algorithms in Predicting Joint Kinematics During Gait*.
- STENUM, J., ROSSI, C. and ROEMMICH, R., 2020. *Two-dimensional video-based analysis of human gait using pose estimation*.
- SABIR, A., AL-JAWAD, N. and JASSIM, S., 2013. Gait recognition using spatio-temporal silhouette-based features. In: *Mobile Multimedia/Image Processing, Security, and Applications 2013*. Mobile Multimedia/Image Processing, Security, and Applications 2013, 28 May 2013. SPIE. pp. 194–203.
- DONG, Y. and NOH, H., 2024. *Ubiquitous Gait Analysis through Footstep-Induced Floor Vibrations*.
- SIMONI, L. et al., 2021. Quantitative and Qualitative Running Gait Analysis through an Innovative Video-Based Approach. *Sensors (Basel, Switzerland)*, 21.
- VITECKOVA, S. et al., 2020. Gait symmetry methods: Comparison of waveform-based Methods and recommendation for use. *Biomedical Signal Processing and Control*, 55, p. 101643.
- NAFEA, O. et al., 2021. Sensor-Based Human Activity Recognition with Spatio-Temporal Deep Learning. *Sensors*, 21, p. 2141.
- MAITY, S., ABDEL-MOTTALEB, M. and ASFOUR, S.S., 2021. Multimodal Low Resolution Face and Frontal Gait Recognition from Surveillance Video. *Electronics*, 10(9), p. 1013.
- MANSSOR, S.A.F., SUN, S. and ELHASSAN, M.A.M., 2021. Real-Time Human Recognition at Night via Integrated Face and Gait Recognition Technologies. *Sensors (Basel, Switzerland)*, 21(13), p. 4323.
- PRAKASH, A. et al., 2023. Multimodal Adaptive Fusion of Face and Gait Features using Keyless attention based Deep Neural Networks for Human Identification.
- CAI, M., WANG, M. and ZHANG, S., 2023. Gait Recognition by Jointing Transformer and CNN. In: W. JIA et al., eds. *Biometric Recognition*. Singapore: Springer Nature. pp. 312–321.
- FAN, C. et al., 2023. *Exploring Deep Models for Practical Gait Recognition*.
- CATRUNA, A., COSMA, A. and RADOI, E., 2024. GaitPT: Skeletons Are All You Need For Gait Recognition.
- LIAO, R. et al., 2020. A model-based gait recognition method with body pose and human prior knowledge. *Pattern Recognition*, 98, p. 107069.

RANI, V. and KUMAR, M., 2023. Human gait recognition: A systematic review. *Multimedia Tools and Applications*, 82(24), pp. 37003–37037.

(Lee and Grimson 2002) LEE, L. and GRIMSON, W.E.L., 2002. Gait analysis for recognition and classification. In: *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, May 2002. pp. 155–162.

ARSEEV, S., KONUSHIN, A. and LIUTOV, V., 2018. Human Recognition by Appearance and Gait. *Programming and Computer Software*, 44(4), pp. 258–265.

CHOPRA, A. and AZAM, R., 2024. Enhancing Natural Language Query to SQL Query Generation Through Classification-Based Table Selection. In: L. ILIADIS et al., eds. *Engineering Applications of Neural Networks*. Cham: Springer Nature Switzerland. pp. 152–165.

ZHANG, Q. et al., 2024. Structure Guided Large Language Model for SQL Generation.

LEE, D. et al., 2024. MCS-SQL: Leveraging Multiple Prompts and Multiple-Choice Selection For Text-to-SQL Generation.

HONG, Z. et al., 2024. Knowledge-to-SQL: Enhancing SQL Generation with Data Expert LLM.

GUO, C. et al., 2024. Prompting GPT-3.5 for Text-to-SQL with De-semanticization and Skeleton Retrieval. In: F. LIU et al., eds. *PRICAI 2023: Trends in Artificial Intelligence*. Singapore: Springer Nature. pp. 262–274.

YI, J., CHEN, G. and SHEN, Z., 2024. RH-SQL: Refined Schema and Hardness Prompt for Text-to-SQL.

NAN, L. et al., 2023. Enhancing Few-shot Text-to-SQL Capabilities of Large Language Models: A Study on Prompt Design Strategies.

HONG, Z. et al., 2024. Next-Generation Database Interfaces: A Survey of LLM-based Text-to-SQL.

LI, H. et al., 2024. CodeS: Towards Building Open-source Language Models for Text-to-SQL.

TALAEI, S. et al., 2024. CHESS: Contextual Harnessing for Efficient SQL Synthesis.

GUO, C. et al., 2023. Retrieval-augmented GPT-3.5-based Text-to-SQL Framework with Sample-aware Prompting and Dynamic Revision Chain.

RAJKUMAR, N., LI, R. and BAHDANAU, D., 2022. Evaluating the Text-to-SQL Capabilities of Large Language Models.

SHI, L., TANG, Z. and YANG, Z., 2024. A Survey on Employing Large Language Models for Text-to-SQL Tasks.