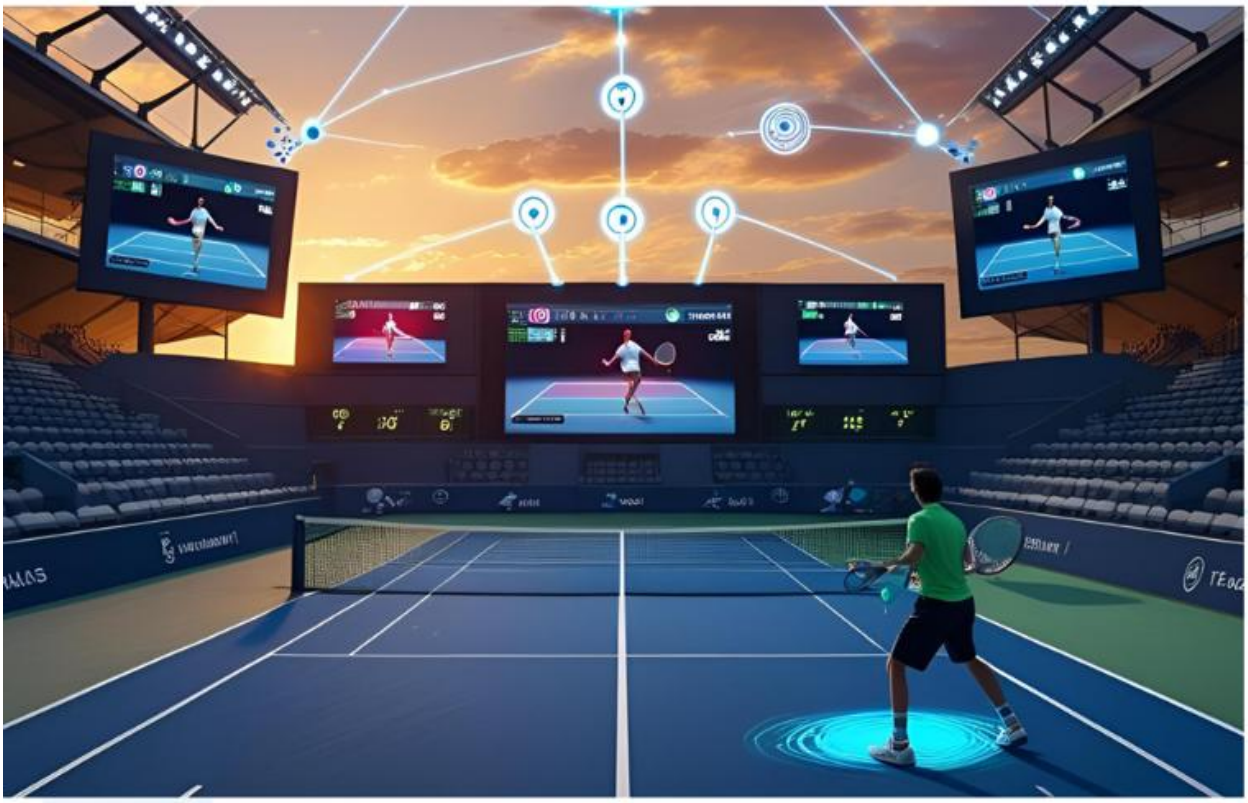
Softwarica College of IT & E-commerce | in collaboration with Coventry University



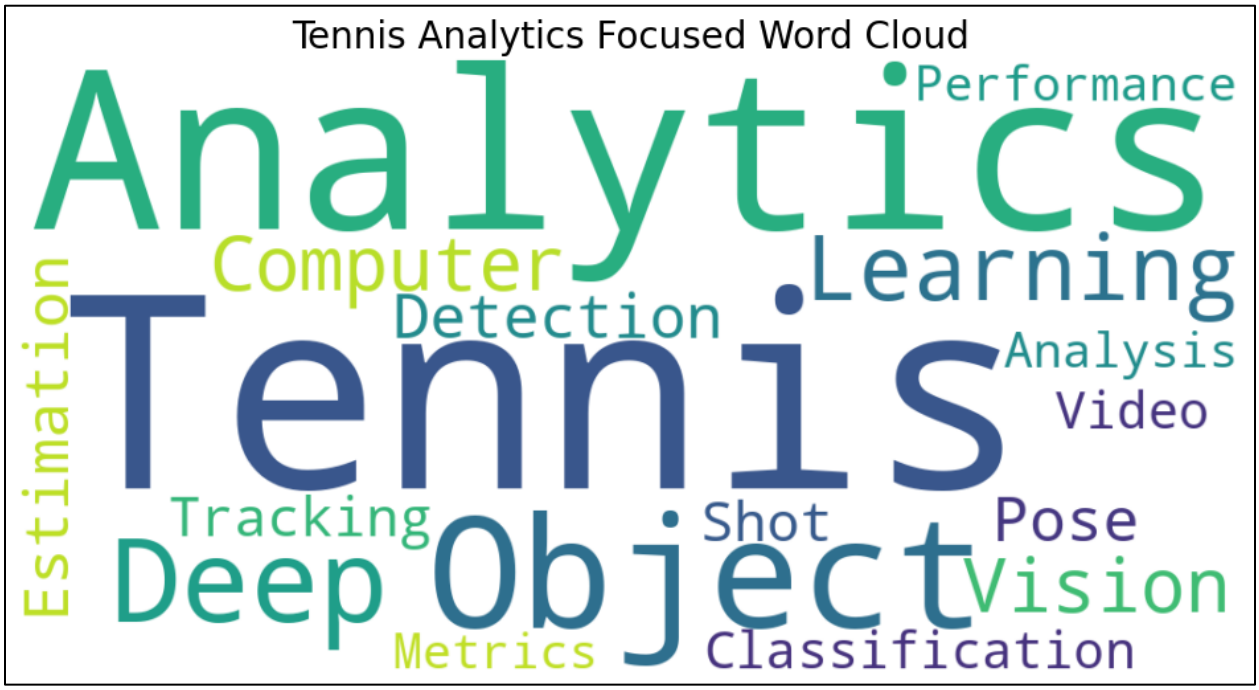
A Deep Learning-Based Framework for Real-Time Tennis Match Analysis Using Object Detection, Key point Extraction, and Object Tracking to Evaluate Player Performance and Shot Patterns from Video Data

Submitted by:
Shubham Khanal

Submitted to:
Manoj Shrestha

June 27, 2025

Keywords


Tennis Analytics Focused Word Cloud
Analytics
Performance
Computer
Learning
Detection
Tennis
Analysis
Video
Estimation
Tracking
Shot
Pose
Deep
Object
Vision
Metrics
Classification

Contents

Introduction

Tennis, a globally *popular racket sport*, demands high levels of technical precision, physical agility, and strategic acumen. Matches are typically decided by a blend of player positioning, shot selection, and real-time tactical decisions, making performance evaluation a crucial component of athlete development. Traditionally, performance analysis in *tennis has relied heavily on manual notation systems and proprietary tracking* technologies such as Hawk-Eye, primarily accessible to elite professional tournaments due to their high cost and infrastructural requirements. Consequently, amateur players, grassroots programs, and independent coaches remain largely excluded from data-driven training tools.

The absence of accessible, real-time analytical solutions has created a significant disparity in the availability of performance metrics for non-professional players. Without affordable and scalable systems, objective assessments of shot quality, player movement, and tactical patterns are rarely achievable at community and academy levels. As a result, decision-making often remains reliant on subjective visual observation, limiting opportunities for targeted skill development and evidence-based coaching.

Moreover, psychological factors such as player motivation, performance anxiety, and confidence levels are significantly influenced by the availability of objective, data-driven feedback. Many amateur and grassroots-level players hesitate to participate in competitive or open sessions due to a lack of personal performance insights. The proposed system aims to address this by providing automated video analysis tools that allow players to track their own mistakes, progress, and performance metrics, ultimately improving their motivation and reducing anxiety around subjective coaching evaluations.

To address this problem, a modular video analysis framework based on deep learning techniques has been proposed. By integrating object detection, pose estimation, and multi-object tracking models, the system aims to automate the extraction of performance indicators such as player positioning, shot selection, and rally dynamics from standard high-definition video footage. The intended framework is designed to deliver affordable, near real-time feedback through open-source tools, democratizing performance analytics and contributing to more inclusive athlete development ecosystems.

Aim

To develop and evaluate a deep learning-based video analysis framework capable of providing near real-time performance indicators for tennis matches using standard high-definition recordings, while considering both technical and psychological impacts on amateur players.

Objectives

- To critically review current deep learning methods for sports video analysis.
- To fine-tune a Convolutional Neural Network model for detecting players and ball instances in tennis videos.
- To configure a pose estimation network for capturing biomechanical joint coordinates of players.
- To design a multi-object tracking pipeline for tracking two players and a ball during singles matches.
- To assess framework performance in terms of precision detection, keypoint estimation accuracy, tracking stability, classification accuracy, and processing speed.

Justification

The proposed system is justified on both technical, psychological, and socio-economic grounds, particularly within the context of Nepal's developing sports infrastructure. From a technical perspective, there exists a noticeable absence of real-time, AI-driven video analysis tools within Nepal's amateur and grassroots tennis community. Unlike elite international tournaments where systems like *Hawk-Eye* are routinely utilized, local tennis ecosystems *lack affordable analytical frameworks, leaving players reliant on subjective*, time-limited feedback from coaches. This technological gap becomes more pronounced for amateur players who, due to time constraints or limited resources, cannot always access personalized in-person coaching. Personal accounts and community observations confirm that numerous players often participate in unsupervised sessions, lacking constructive, objective assessments of their performance.

The implementation of AI-based analytics tools in amateur tennis has potential psychological implications that warrant significant attention. *Contemporary research in sports psychology indicates that objective, immediate performance feedback enhances player confidence, reduces performance anxiety, and minimizes cognitive overload associated with subjective self-assessment.* Accurate data-driven insights enable athletes to focus on quantifiable improvements, shifting attention from self-doubt to actionable metrics. In high-performance environments such as professional football, data analytics have played pivotal roles in team rebuilding and player development. For instance, *clubs like Brentford FC and FC Midtjylland* successfully utilized data-centric approaches to outperform financially superior rivals, validating the psychological and strategic value of objective feedback in athlete development. Applying similar principles to amateur tennis in Nepal is both innovative and necessary.

This project can additionally be conceptualized as an educational tool, designed not to replace traditional coaching but to complement it. AI-driven systems offer valuable opportunities for

athletes to independently analyze their own match footage, fostering a culture of self-assessment, data literacy, and mental discipline. By encouraging players to interpret performance metrics and positional data, the system supports the development of tactical awareness, reflective practice, and a growth-oriented mindset qualities essential for long-term athletic progression.

The introduction of this framework also addresses inequalities inherent in coaching resource distribution at the grassroots level. *Community sports programs frequently suffer from disproportionate coach-to-player ratios, resulting in limited individual feedback opportunities*. While physical learning remains irreplaceable, integrating real-time analytics ensures that each player receives tailored performance insights, even in coach-scarce environments. Such systems democratize access to performance evaluation, promoting fairness and inclusivity in athlete development.

From an ethical standpoint, concerns surrounding privacy and data usage have been acknowledged. The framework prioritizes ethical design principles, ensuring that video data is sourced responsibly and algorithmic decisions maintain cultural sensitivity, particularly regarding attire and player behavior norms within Nepalese society. The potential risks of promoting uniform tactical approaches or reinforcing narrow performance benchmarks will be mitigated through continuous dataset diversification and localized user testing, preserving individuality and cultural authenticity.

Furthermore, psychological theories such as bounded rationality, proposed by *Herbert Simon*, affirm that amateur players possess finite cognitive capacities to process complex match events and technical nuances. Real-time analytical tools streamline this decision-making process by highlighting the most relevant performance indicators, reducing cognitive load, and supporting timely adjustments. The paradox of choice, another well-documented behavioral economics concept, suggests that athletes can become overwhelmed when presented with excessive, unfiltered performance data. By curating personalized, context-relevant metrics, the proposed system optimizes information delivery, enhancing decision-making efficiency without contributing to mental fatigue.

For these reasons, the proposed video analysis framework is justified as a technically viable, psychologically beneficial, and ethically responsible solution to bridge the analytical resource gap within Nepal's amateur tennis community.

Statement of Purpose

The absence of accessible, real-time analytical tools limits the capacity for data-driven coaching and comprehensive performance evaluation in tennis. This project democratizes advanced analytical metrics by employing *open-source deep learning components* to generate actionable insights at minimal expense.

Research Question

1. Can object detection modules, pose estimation networks, and object tracking algorithms be integrated to produce accurate, near real-time analytics for tennis matches using standard high-definition video recordings?
2. How does real-time video analysis feedback affect the motivation, confidence, and performance anxiety of amateur tennis players at the grassroots level?
3. What are the usability and accessibility challenges associated with implementing AI-based sports analytics systems for non-professional players and local coaching setups?

Literature Review

Desk-Based Research Methodology

Recent developments in artificial intelligence (AI) and deep learning have transformed the landscape of sports analytics, providing novel opportunities for automating performance assessments in dynamic environments such as tennis. Existing literature offers extensive insights into the capabilities of AI-based video analysis systems, particularly regarding object detection, tracking, and sequential pattern analysis.

Desk-based research has emphasized the direct application of AI models to sports video analytics. Early systems employed conventional statistical methods for performance measurement, which were later enhanced by the introduction of Convolutional Neural Networks (CNNs). Landmark innovations such as the R-CNN family and the You Only Look Once (YOLO) series significantly improved object detection speed and accuracy in complex scenes. Among these, *YOLO has demonstrated superior efficacy for sports video analysis,* providing real-time detection of rapidly moving objects like tennis balls and athletes.

Advancements in pose estimation algorithms, particularly the High-Resolution Network (HRNet), have further contributed by enabling precise biomechanical analysis through joint coordinate extraction. Multi-object tracking (MOT) frameworks, including Deep SORT and ByteTrack, have maintained consistent object identities across frames, managing occlusions and re-identification challenges frequently encountered in tennis match footage.

Secondary desk-based studies have explored the broader implications of AI-driven systems within sports and adjacent industries. *Reviews of AI applications in athletic retail, transportation, and automated retail environments have contextualized the scalability and technical feasibility of vision-based analytics frameworks.* These studies have confirmed the potential for AI systems to

deliver high processing performance on consumer-grade hardware, reinforcing the viability of accessible performance evaluation tools for grassroots sports.

Ethical and governance considerations have also emerged, with scholars discussing the risks of data privacy violations, bias in AI-generated insights, and the absence of standardized regulatory guidelines in competitive sports contexts. Organizations such as the International Tennis Federation (ITF) and the International Olympic Committee (IOC) have yet to establish formal protocols governing AI integration, leaving an unregulated space in which developers must self-regulate ethical standards.

A noticeable research gap persists in the availability of open-source, real-time AI systems specifically tailored to tennis and other racket sports. Additionally, minimal attention has been given to integrating data literacy training into youth development programs, despite evidence suggesting significant long-term performance benefits. Addressing this gap, the proposed study contributes to an underexplored domain by developing a real-time, AI-based video analysis framework for tennis, prioritizing affordability and accessibility.

Case Study

Several empirical case studies have highlighted the tangible benefits of AI in professional tennis. Notably, data-driven insights have been employed for tactical planning, injury management, and training optimization.

For example:

- *Naomi Osaka's coaching team* used predictive models for opponent strategy analysis, enabling dynamic match preparations based on historical and real-time opponent behavior data.
- *Andy Murray's post-injury rehabilitation program* was guided by biomechanics systems that provided precise feedback on his joint movements and stress points, accelerating recovery while minimizing reinjury risk.

These applied case studies have demonstrated the practical implications of AI tools in elite tennis, offering evidence of AI's capability to influence player outcomes, reduce injury rates, and optimize tactical strategies at the highest competitive levels.

Parallel studies in sports psychology have underscored the importance of feedback systems in reducing performance anxiety and enhancing motivation among amateur athletes. According to recent reviews, access to objective performance metrics increases player confidence and reduces anxiety associated with subjective evaluations by coaches, particularly in high-pressure competitive or open sessions. The integration of AI feedback systems in other sports has demonstrated marked improvements in athlete confidence and psychological readiness. This

reinforces the psychological value proposition of integrating data-driven video analysis tools into amateur and grassroots tennis.

Integration

The proposed prototype integrates multiple deep learning-based analytical modules to construct a unified video analysis framework tailored for tennis performance evaluation. A sequential data processing pipeline has been designed, starting with object detection, followed by pose estimation, multi-object tracking, and stroke classification. YOLOv8, a state-of-the-art object detection model with a Cross Stage Partial Darknet-53 backbone, will be fine-tuned to identify player and ball regions in real-time from high-definition video footage.

Following detection, spatial court keypoints will be identified through a CNN model implemented using the PyTorch framework. These spatial references will aid in contextualizing player positioning and shot locations. Subsequently, a High-Resolution Network (HRNet) will be deployed for pose estimation, extracting precise skeletal keypoint coordinates of the athletes, enabling biomechanical performance assessments and stroke categorization.

To maintain consistent identity management of multiple tracked objects — including two players and the ball — multi-object tracking algorithms such as Deep SORT or ByteTrack will be integrated. These algorithms will address challenges related to object occlusion, re-identification, and rapid movement by assigning stable tracking IDs across consecutive video frames.

Video frame manipulation, including resizing, extraction, and normalization, will be executed through the Open-Source Computer Vision Library (OpenCV). For shot classification, a hybrid temporal classifier combining a ResNet-50 CNN and a two-layer LSTM network will be implemented. This module will leverage both spatial and sequential data to categorize strokes into predefined classes such as forehand, backhand, and serve.

All modules will be containerized *using Docker, enhancing modularity, ease of integration, and scalability.* Inter-module communication will be facilitated through RESTful API endpoints, promoting system flexibility and allowing asynchronous data processing. Analytical outputs, including trajectory plots, positional heatmaps, rally summaries, and player movement analytics, will be consolidated into an interactive dashboard, providing real-time visual feedback to users. This visualization platform will deliver performance indicators in a format accessible to coaches, athletes, and analysts.

Project Plan

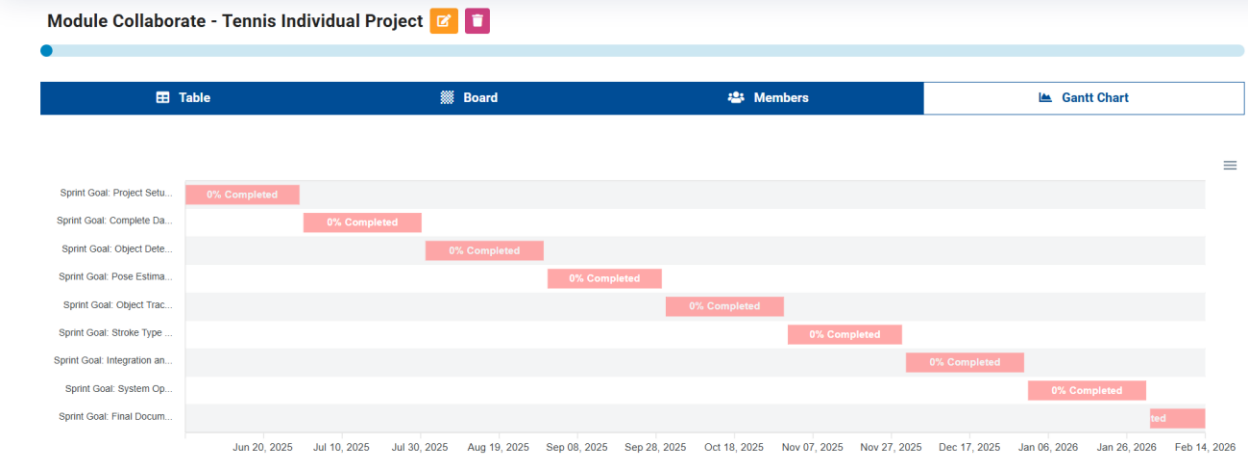


Figure 1 Project Plan

Figure 2 Project Plan Table

Methodology

Data Collection and Annotation

High-definition singles match videos recorded at 15 to 30 frames per second will be sourced from publicly available video repositories such as YouTube or tennis archives. Additionally, video recordings may be produced independently using standard DSLR or mirrorless cameras to supplement the dataset. *To ensure annotation accuracy, trained personnel will manually generate labels for each video frame, including bounding boxes for players and ball instances, skeletal key points for major joints, and stroke labels indicating forehand, backhand, or serve actions.*

Annotations will adhere to consistent labeling guidelines, with inter-rater reliability periodically evaluated to maintain dataset integrity. These annotated datasets will be curated into structured JSON or XML formats suitable for supervised learning processes.

Data Preprocessing

Video data will undergo preprocessing through the OpenCV library. Each video will be divided into individual frames and subsequently resized to a uniform resolution to ensure consistent input dimensions across model pipelines. Pixel normalization will be applied to standardize the input range and improve model convergence during training.

To support robust supervised learning, datasets will be partitioned into training (70%), validation (15%), and testing (15%) subsets. Data augmentation techniques such as random horizontal flipping, rotation, cropping, and brightness adjustments will be applied to artificially increase dataset diversity and reduce overfitting risks.

Model Selection and Training

A YOLOv8 object detection model will be fine-tuned through transfer learning on the annotated tennis dataset. The model's pre-trained weights on COCO or Open Images datasets will serve as the initialization point, with additional training conducted on the domain-specific tennis dataset to optimize detection performance for player and ball instances.

For pose estimation, the High-Resolution Network (HRNet) will be adapted by retraining its final regression layers on annotated joint coordinates. HRNet's ability to maintain high spatial resolution throughout the feature extraction process makes it suitable for precise skeletal keypoint detection, especially under partial occlusions and rapid player movement.

Stroke classification will employ a hybrid architecture combining a ResNet-50 CNN for extracting spatial features from player poses and a two-layer Long Short-Term Memory (LSTM) network for modeling the temporal sequences of joint movements. Hyperparameter optimization will be conducted using a grid search approach, varying batch sizes, learning rates, and optimizer

configurations (Adam, SGD, or RMSprop) to determine the most efficient model settings based on validation performance.

Shot Type Classification

Temporal features derived from sequential player joint coordinates and ball trajectory vectors will be extracted and fed into supervised classification algorithms. The LSTM module will capture the sequential dependencies between body postures and stroke types over consecutive frames, enabling reliable categorization of shots into forehand, backhand, and serve categories.

Evaluation Metrics

Framework performance will be quantitatively assessed through several established evaluation metrics.

- Mean Average Precision (mAP) will measure the precision and recall of object detection outputs (players and ball).
- Percentage of Correct Keypoints (PCK) will evaluate the accuracy of skeletal keypoint predictions relative to ground truth positions.
- Multiple Object Tracking Accuracy (MOTA) will assess the consistency and correctness of player and ball identity tracking across frames.
- F1 Score will be computed for stroke classification accuracy, balancing precision and recall.
- Frames Per Second (FPS) processing speed will be monitored to confirm real-time system capability, targeting a minimum of 30–40 FPS on high-definition input streams.

Conclusion

The proposed framework combines open-source deep learning tools to deliver affordable, near real-time analytics for tennis performance evaluation. By automating object detection, pose estimation, and multi-object tracking from standard video footage, the system aims to enhance accessibility to advanced analytical insights for amateur players, community programs, and training institutions.

References

Papers with Code - CSPDarknet53 Explained. (n.d.). https://paperswithcode.com/method/cspdarknet53

Why tennis remains the most popular racket sport | 40LOVE. (n.d.). *The Times Nepal*. https://www.40love.co/blog/tennis-popularity

Takahashi, H., Okamura, S., & Murakami, S. (2022). *Performance analysis in tennis since 2000: A systematic review focused on the methods of data collection*. https://revistaseug.ugr.es/index.php/IJRSS/article/view/33263

Saha, G. (2022, March 25). Top 10 Open source Deep learning tools. *Open Source For You*. https://www.opensourceforu.com/2022/03/top-10-open-source-deep-learning-tools/

Gino, S. (2023, March 10). Sport-Analytics with YOLO et al. - Shahar Gino - Medium. *Medium*. https://sgino209.medium.com/sport-analytics-with-yolo-et-al-951b3f26221b

Rukundo, S., Wang, D., Wongnonthawitthaya, F., Sidibé, Y., Kim, M., Su, E., & Zhang, J. (2025, March 11). *A survey of challenges and sensing technologies in autonomous retail systems*. arXiv.org. https://arxiv.org/abs/2503.07997

Andy Murray and hip resurfacing | Edwin Su MD | Orthopaedic Surgeon New York NY. (n.d.). Andy Murray and Hip Resurfacing. https://www.edwinsu.com/andy-murray-and-hip-resurfacing.html

Li, X., & Xia, M. (2024). Dynamic calibration of self-efficacy to cognitive load: the longitudinal mediation effect of state anxiety. *BMC Psychology*, *12*(1). https://doi.org/10.1186/s40359-024-02254-y

Arastey, G. M. (2019, November 22). *The Brentford FC story: running a football club through data | Sport Performance Analysis*. Sport Performance Analysis. https://www.sportperformanceanalysis.com/article/2018/6/8/the-history-of-brentford-football-analytics

A Kathmandu tennis school aims to produce ace players. (2021, November 11). *The Kathmandu Post*. https://kathmandupost.com/art-culture/2021/11/11/a-kathmandu-tennis-school-aims-to-produce-ace-players#:~:text=In%20a%20bid%20to%20produce,popularise%20the%20sport%20in%20Nepal.

Wikipedia contributors. (2025, June 10). *Herbert A. Simon*. Wikipedia. https://en.wikipedia.org/wiki/Herbert_A._Simon

Appendix

Glossary of Terms

- Convolutional Neural Network: A deep learning architecture for image processing.
- You Only Look Once Version Eight: A real-time object detection model.
- High-Resolution Network: A network architecture preserving spatial resolution for precise feature localization.
- Long Short-Term Memory network: A recurrent network capable of capturing long-term dependencies in sequence data.
- Representational State Transfer Application Programming Interface: A protocol for communication between software modules.
- Frames Per Second: The number of video frames processed per second by the system.