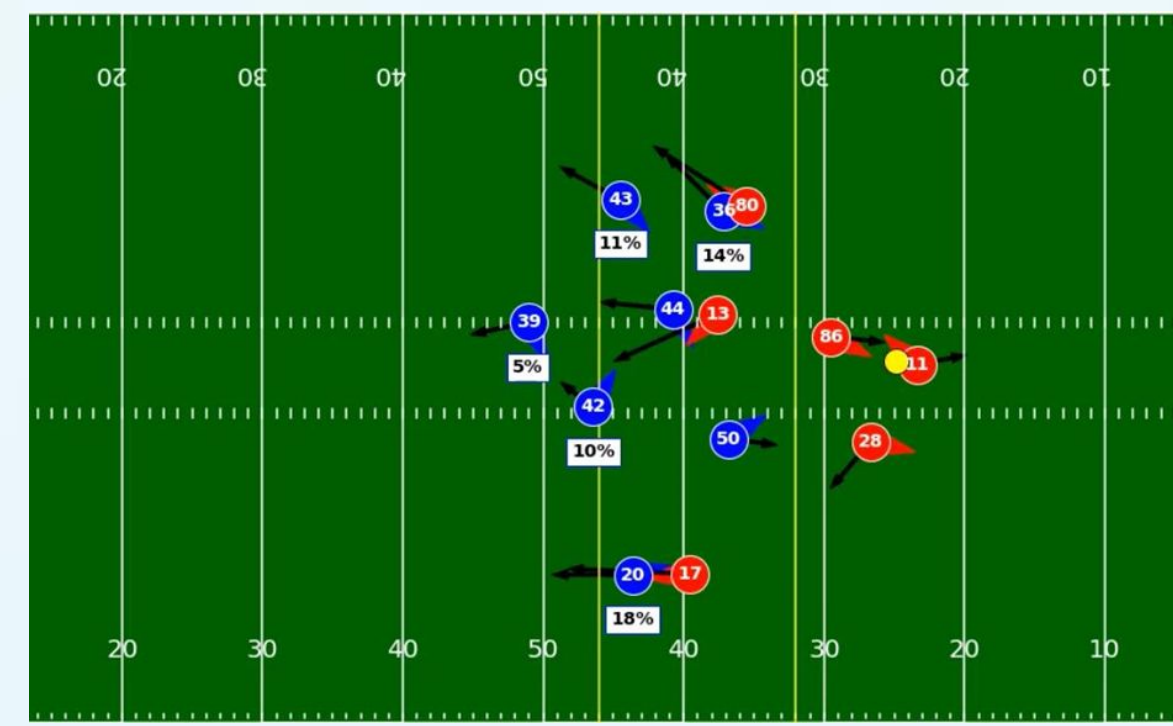


Introduction

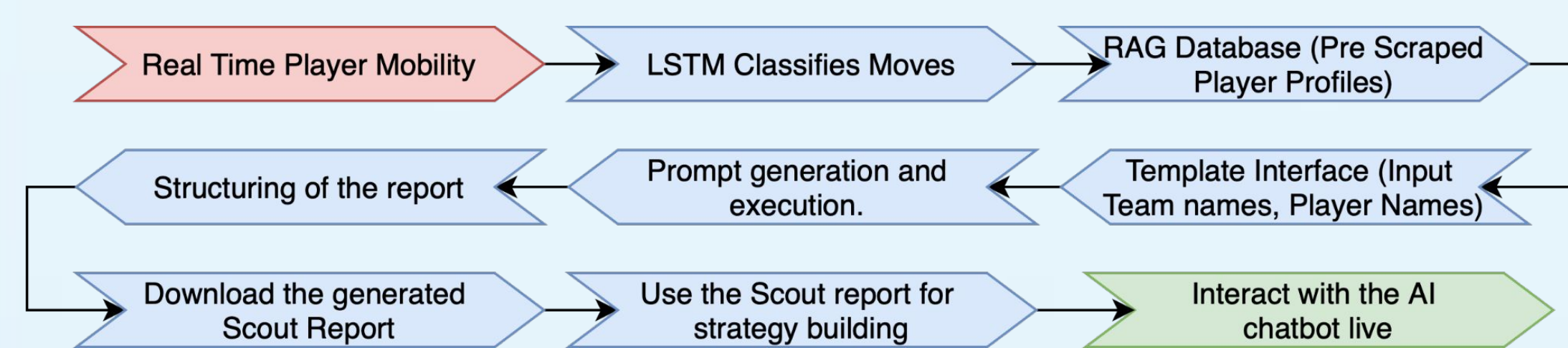
Professional American football teams are increasingly utilizing data-driven insights to enhance performance and gain a competitive edge. Advances in technology have expanded data collection opportunities, allowing coaches to access real-time information on player mobility, health, and performance. This data is crucial for making informed decisions before and during games.

To succeed, teams must consolidate vast amounts of fragmented information into actionable strategies, which requires rapid gathering and analysis due to the sport's high-pressure nature. Football's spatio-temporal dynamics necessitate a unique approach to data analysis. While other sports have utilized models like GNNs, CNNs, and reinforcement learning, American football has historically kept much of its data private. Recently, however, the NFL has begun sharing some data for public competitions, paving the way for innovative game-time analyses.

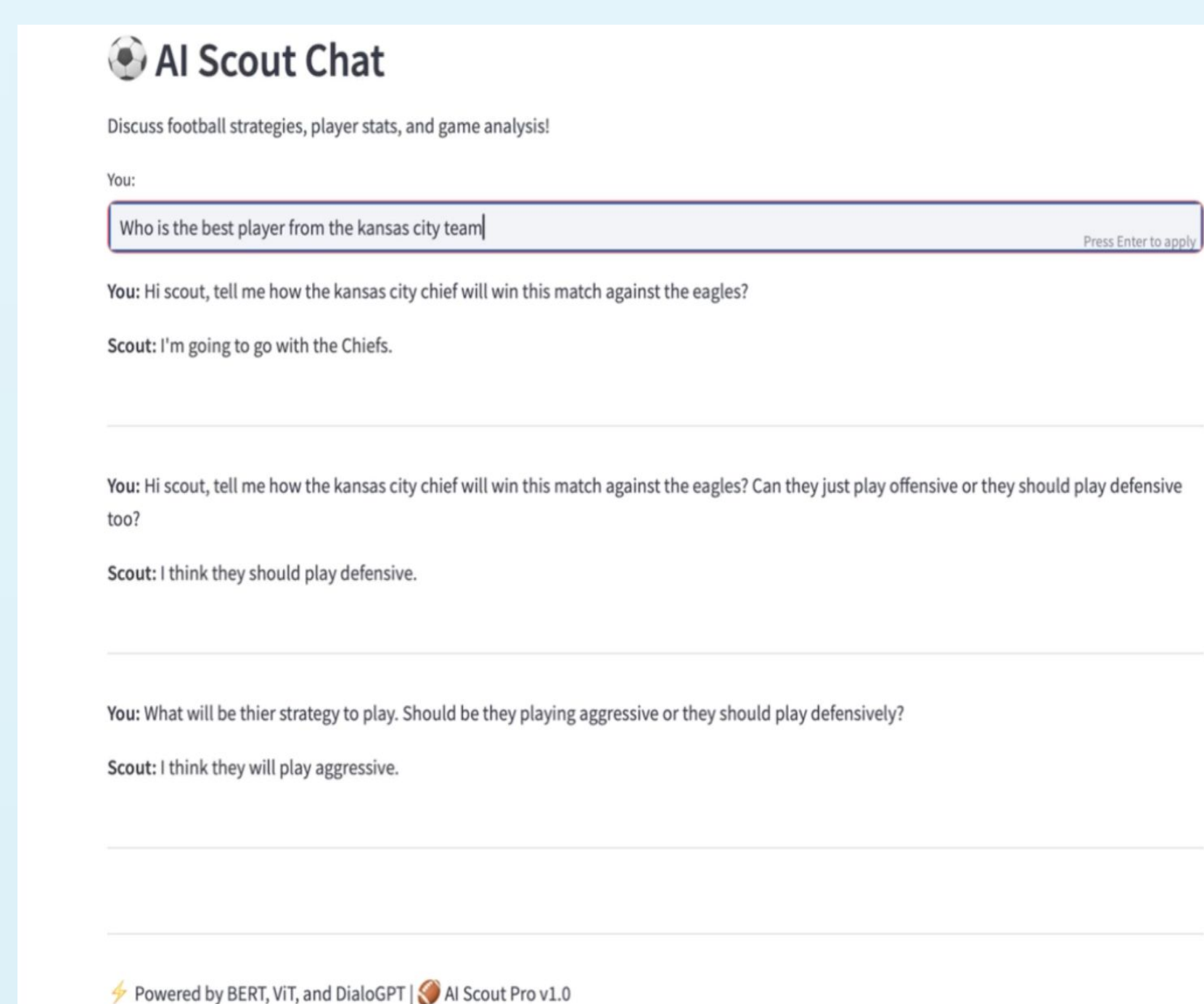
Previous research on American football has focused on specific aspects such as coverage classification, player identification, and movement trajectories. This project aims to adopt a holistic approach by developing a coaching assistant model that integrates multi-modal data—including player positional tracking, news feeds, and performance statistics—using advanced AI techniques. Our system employs deep learning for scheme analysis and a large language model (LLM) to provide domain-specific information. This enables coaches and analysts to adapt strategies effectively, optimizing play formations, assessing defensive schemes, and maximizing player utilization to improve performance and minimize risks.



Application Overview



Application Interface



Methodology

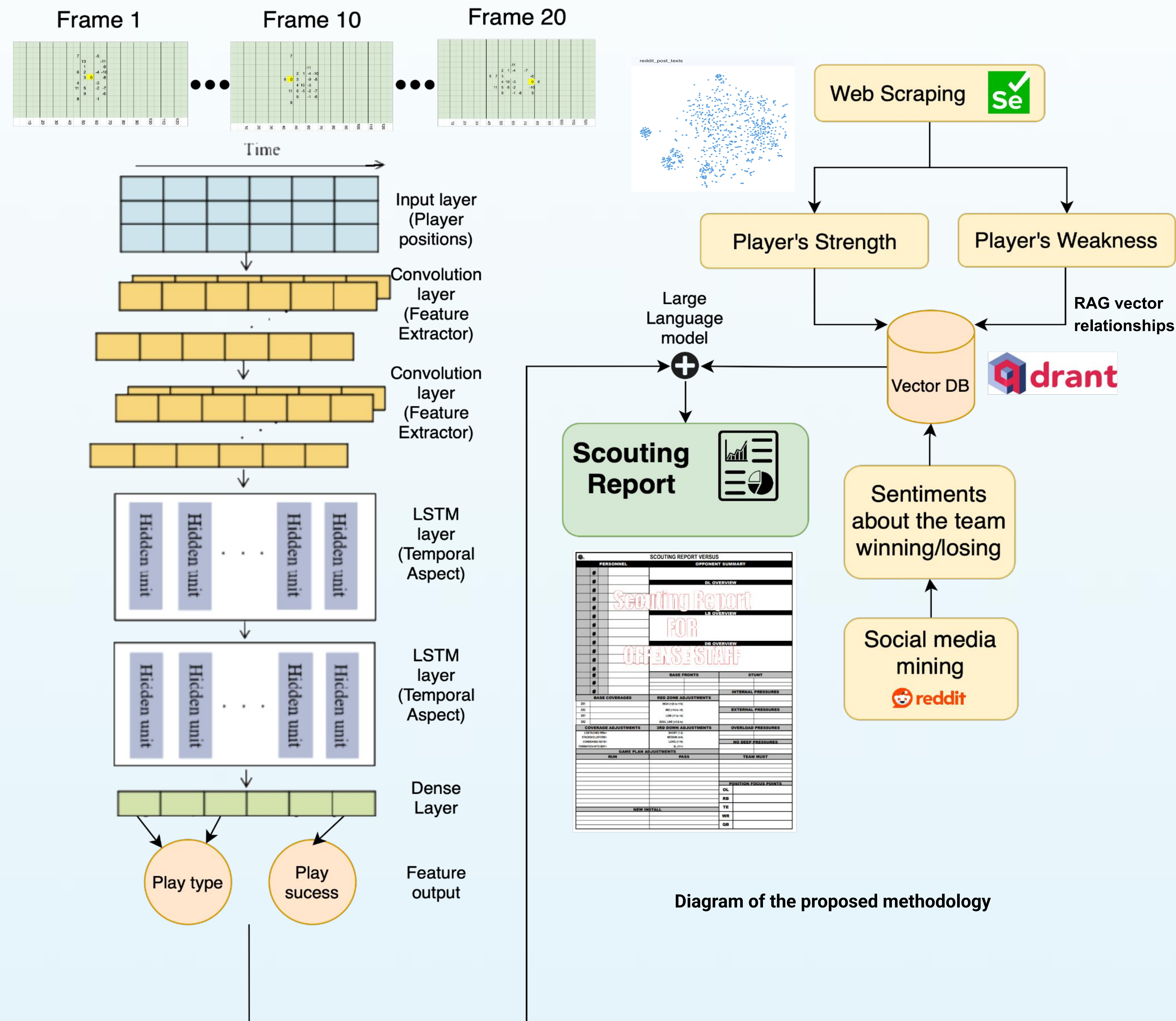


Diagram of the proposed methodology

Data Processing

- Use NFL player tracking data (Next Gen Stats).
- Focus on **offensive players** for scouting reports.
- Clean and standardize play frames.

Feature Engineering

- Extract **spatial** (player positions, formations).
- Capture **temporal** (movement over time).
- Create **interaction features** (relative positioning).

Model Architecture

- CNN** → Understands spatial patterns.
- LSTM** → Tracks movement over time.
- Combined model** → Classifies play types & predicts success.

Play Classification & Success Prediction

- Identify offensive formations & play types.
- Predict play success based on yardage & game situation.

AI-Augmented Insights (RAG Integration)

- Retrieve historical team tactics & scouting reports.
- Use **play embeddings** to enhance predictions.

Evaluation & Real-Time Application

- Test accuracy on **formation & play type detection**.
- Deploy an **interactive dashboard** for coaches & analysts.

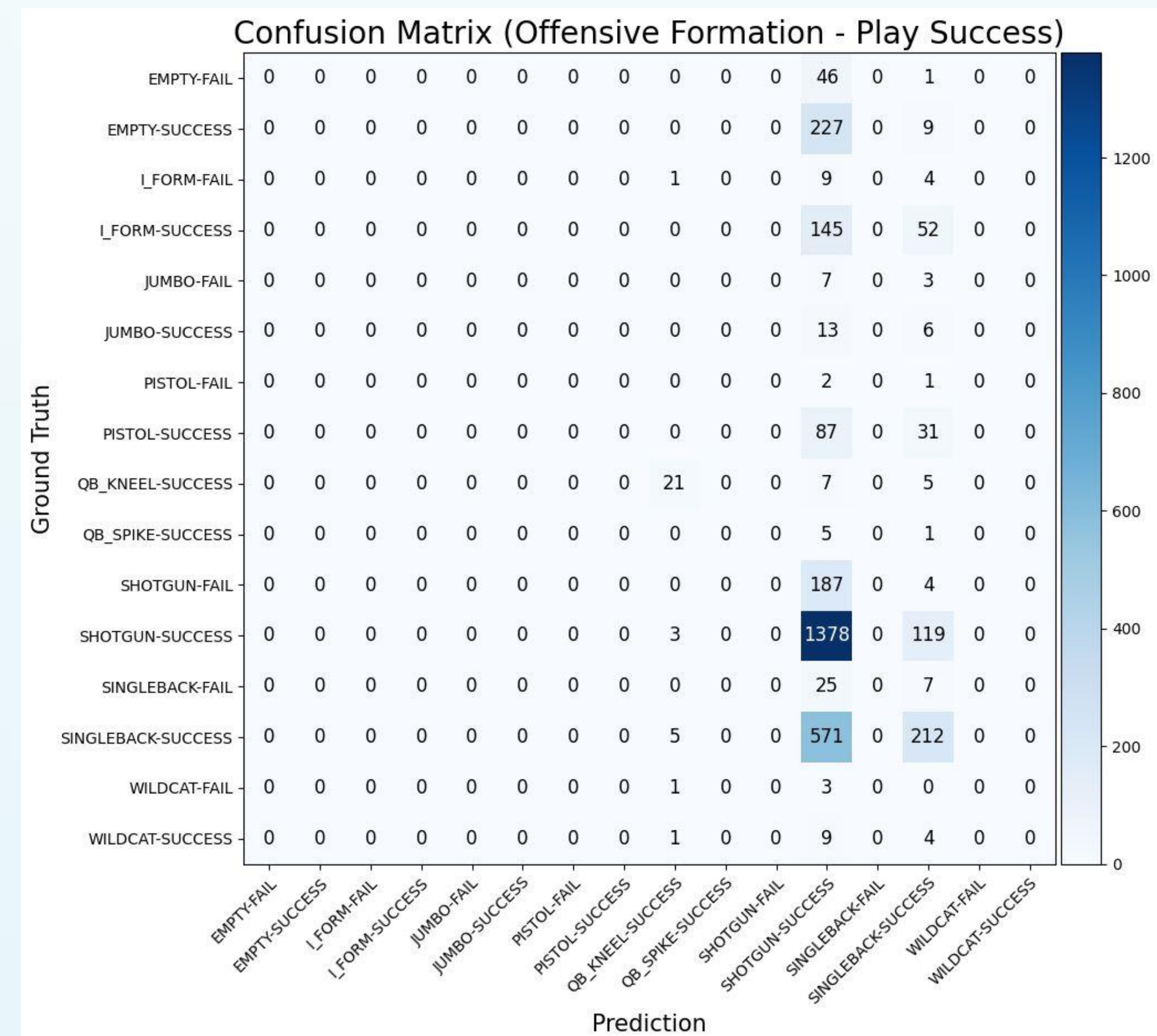
Results

CNN-LSTM Prediction Accuracies:

- Play-action type (T/F) - 0.80
- Run-pass option type (T/F) - 0.89
- Offensive formation (multi-class) - 0.56
- Receiver alignment (multi-class) - 0.41
- Defensive pass coverage (multi-class) - 0.31
- General offensive play type (multi-class) - 0.61
- Run concept (multi-class) - 0.53
- Offensive play success (T/F) - 0.91

RAG precision@10* - 98.3%

*fraction of relevant items within top ten search results



References

- Cheong, L. L., X. Zeng, & A. Tyagi (2021). Prediction of defensive player trajectories in NFL games with defender CNN-LSTM model. *MIT Sloan Sports Analytics*.
- Ding, N., K. Takeda and K. Fujii (2022). Deep Reinforcement Learning in a Racket Sport for Player Evaluation With Technical and Tactical Contexts. *IEEE Access*, 10, 54764-54772.
- Liu, H., C. Adreon, N. Wagnon, et. al. (2023). Automated player identification and indexing using two-stage deep learning network. *Scientific Reports*, 13(10036).
- Lopez, M., T. Bliss, A. Blake, P. Mooney, & A. Howard (2024) NFL Big Data Bowl 2025. <https://kaggle.com/competitions/nfl-big-data-bowl-2025>. Kaggle.
- Raabe, D., N. Reinhard, & D. Memmert (2023). Graph representation for the analysis of multi-agent spatiotemporal sports data. *Applied Intelligence*, 53, 3783-3803.
- Song, H., M. A. Jazaery, H. Ding, et. al. (2023). Explainable defense coverage classification in NFL games using deep neural networks. *MIT Sloan Sports Analytics*.
- Wang, Z., P. Velickovic, D. Hennes, et. al. (2024). TacticAI: an AI assistant for football tactics. *Nature Communications*, 15(1906).
- Westerberg, J. (2017). *A deep learning approach for action classification in American football video sequences* [Thesis, Uppsala University]. DiVA – Digitala Vetenskapliga Arkivet.

Acknowledgments

University of Missouri - Kansas City
CS5542: Big Data Analytics and Applications
Professor: Dr. Yugyung Lee