# SOCCER INJURY PREVENTION

Matt D, Saroj U, Sam A, Connor H, Anjali J

Soccer is played by 4% of the world population. (2014 FIFA Census) This is over 320,000,000 people!

Soccer is regarded as a high-intensity sport.

Exposes the players to physical and physiological demands.

# Soccer Stats

Running longer (7-10 miles) per match

90+ minutes on the field

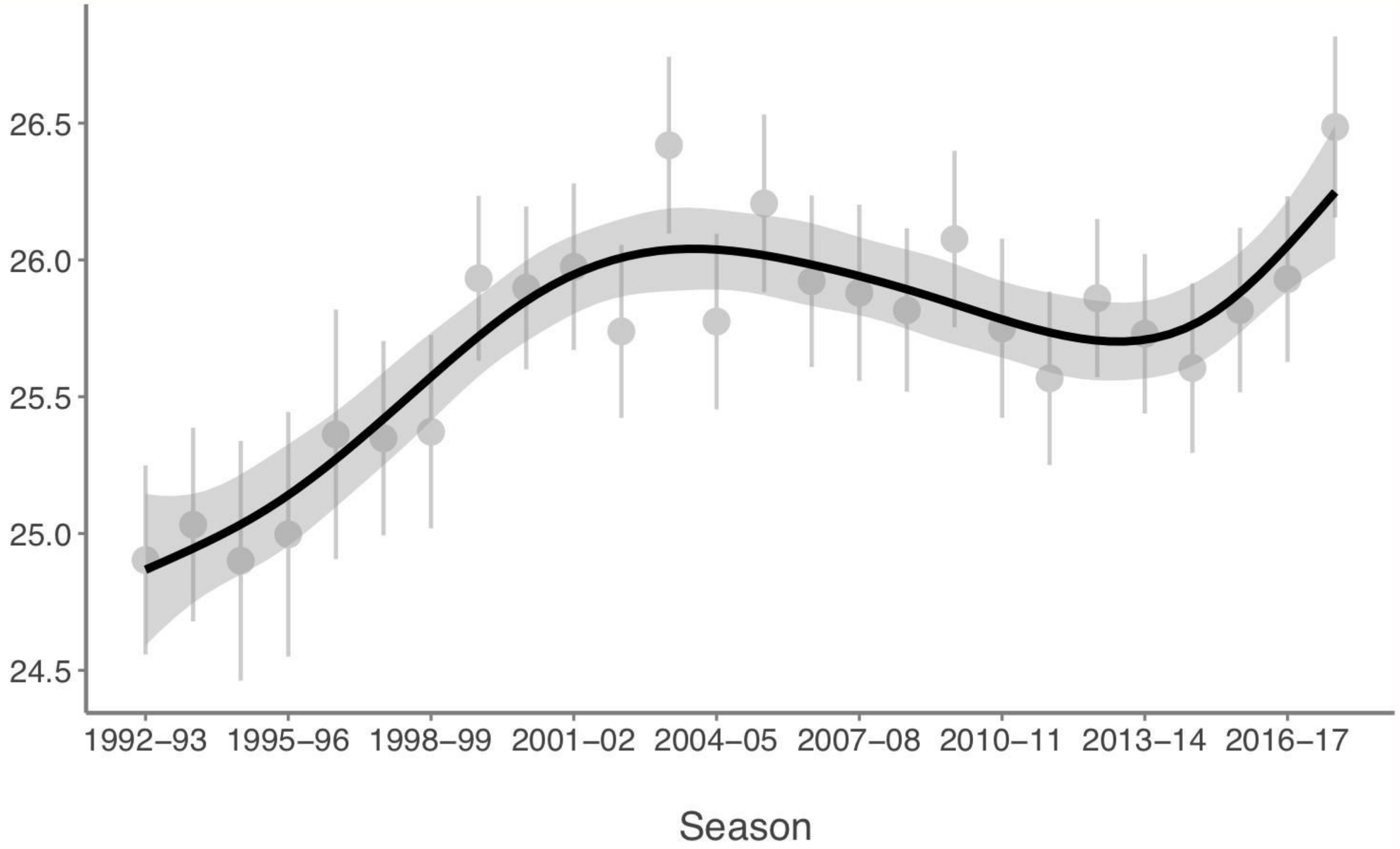
No micromanagement

Longer careers (avg age increasing)

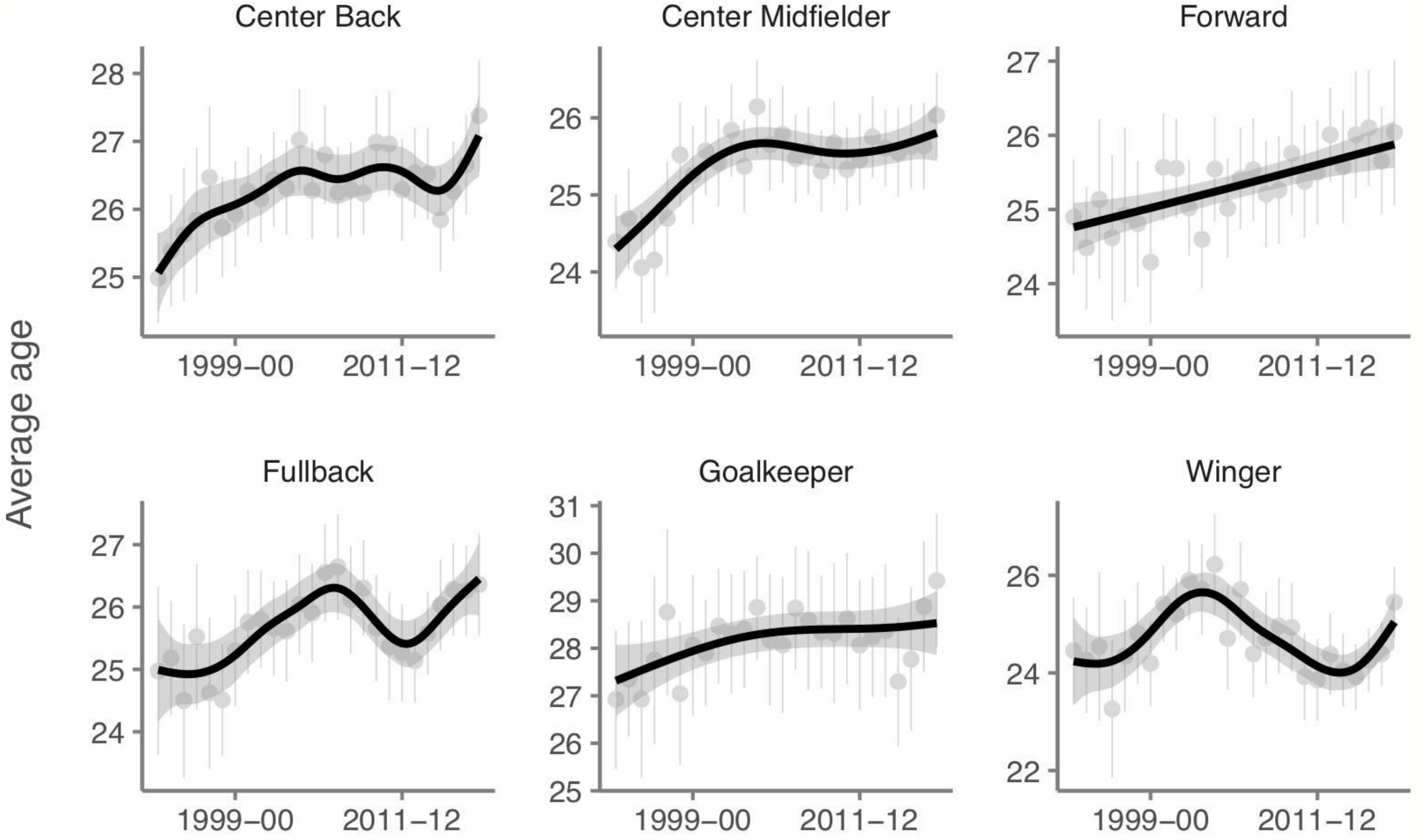
# INSPIRATION

In the last decade, the number of studies about machine learning algorithms applied to sports has rapidly increased.

# AVERAGE AGE PER SEASON



# AVERAGE AGE PER POSITION



# PROBLEM STATEMENT

**Problem:**

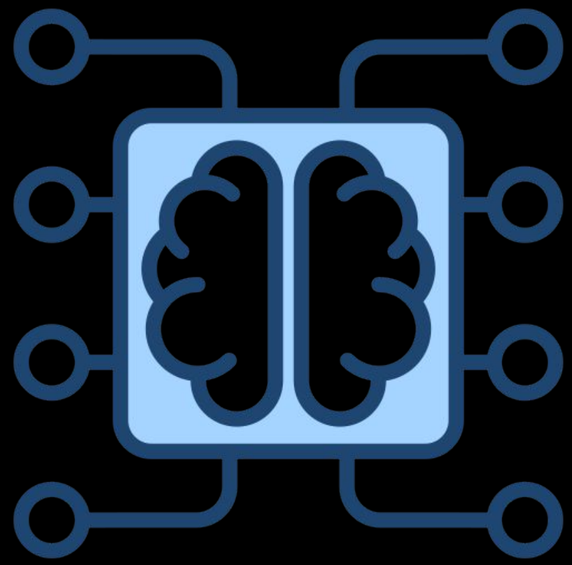
Soccer players continuously face a risk of injury which can lead to major setbacks.

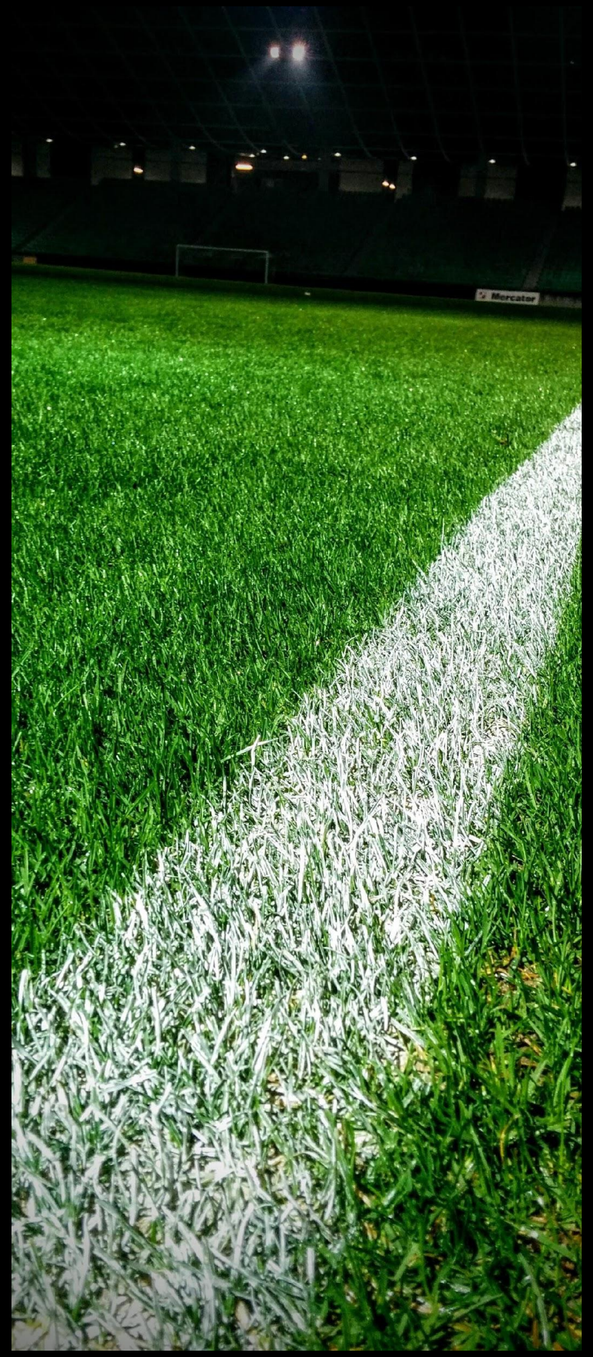
Injuries in sports significantly impact the team performance and club finances.

# GOAL/SOLUTION

**Solution:**

Implement machine learning to improve predictive accuracy in these injuries. This will provide a guideline describing a correct approach for sports science predictions driven by data.





# Research Question

“What are the primary risk factors influencing soccer injuries, and to what extent can machine learning algorithms accurately forecast and mitigate them using comprehensive player data?”

# CONNECTION TO SOCIAL GOOD

- Proactively identify & mitigate injury risks
- Enhance player well-being & overall performance
- Long-term player retention
- Cross-sport applications

# Relevant Literature

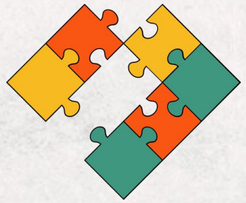
High accuracy using machine learning primarily limited by sample size and time constraints

**Assess Injury Risk in Elite Youth Football Players**

-Identified acute vs overuse injuries

# Gaps in Literature

Limited attention to injury dynamics



Insufficient multivariate patterns

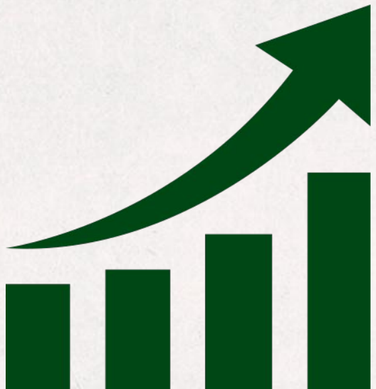
Lack of accurate/interpretable models

Comprehensive game data

- Time Played
- Home/away

# Literature Findings

Decision Trees & SVM show high promise for improving predictive accuracy.



# Goal

Aim to address these gaps & develop more robust multivariate injury prediction models.

# Theory

**Theoretical Framework:**

Player well-being & injury occurrence are influenced by a complex factors.

**Data-driven Approach:**

Unravel intricate patterns within factors to enhance injury prediction accuracy.

# Hypothesis

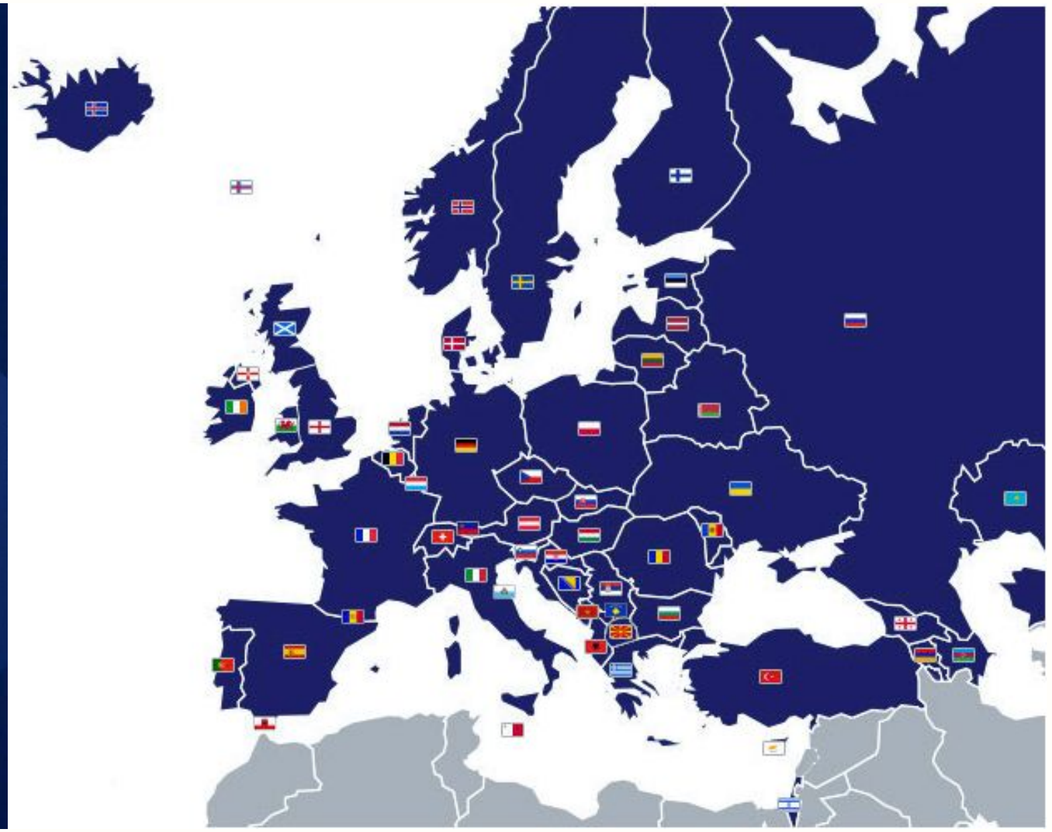
Conceptual Variables:

Individual characteristics, training intensity, workload, and possible environmental conditions.

Hypothesis:

Older players with a history of injuries and higher training intensities will demonstrate an increased likelihood of injury.

# DATA



- Transfermarkt tracks detailed player stats to measure their estimated market value
- Provides comprehensive dataset of player game and injury history

# Data Processing

**Collection Challenges:**

- Limited time, resources, and privacy constraints hinder acquiring granular data for comprehensive analysis.

**Quality Concerns:**

- Datasets are incomplete with missing values (NaN), excluded illness/surgery data, and sensitive information omitted, reducing comprehensiveness.

# DATA

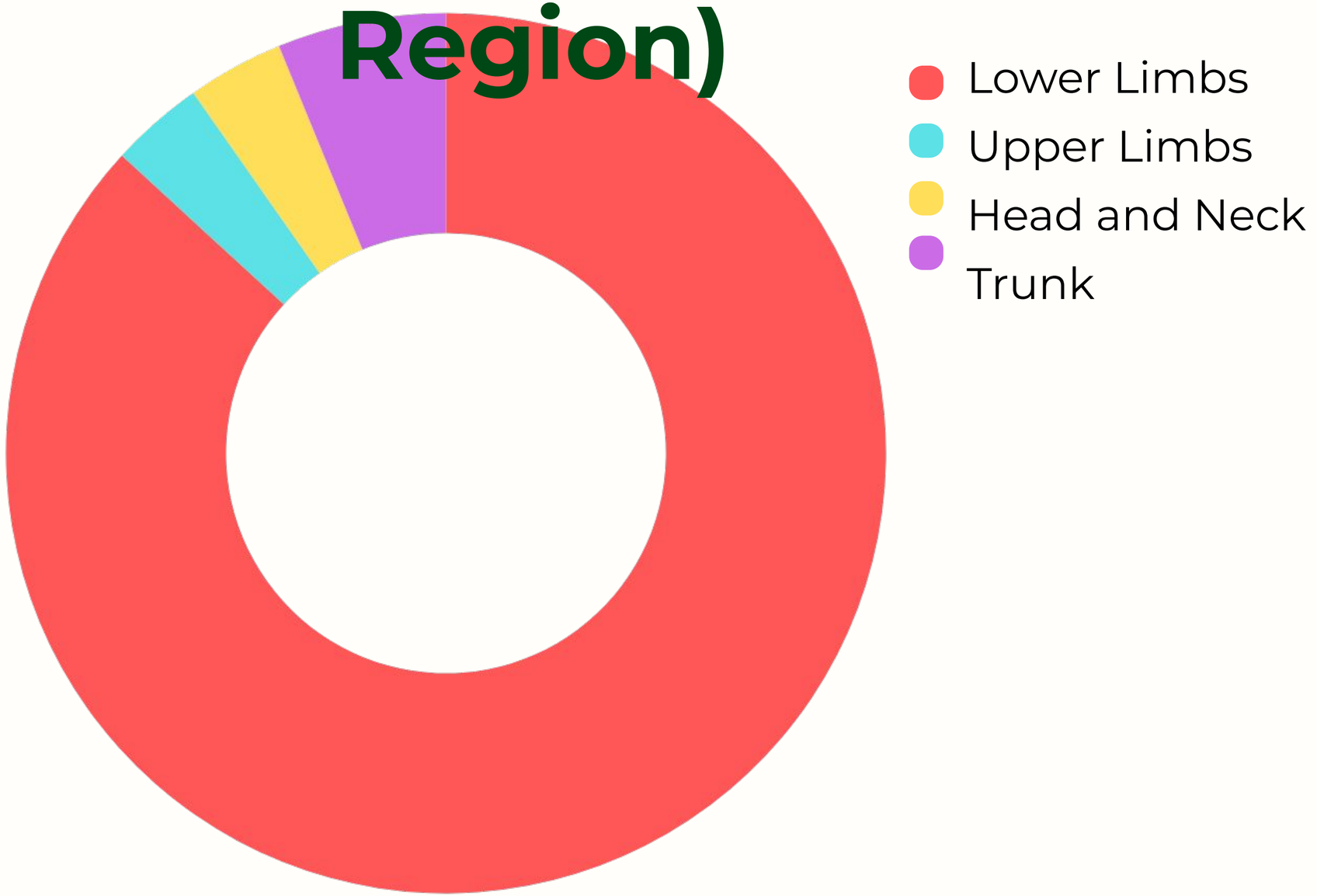

- Injuries from over 1,300 male players

- 98 Professional European teams

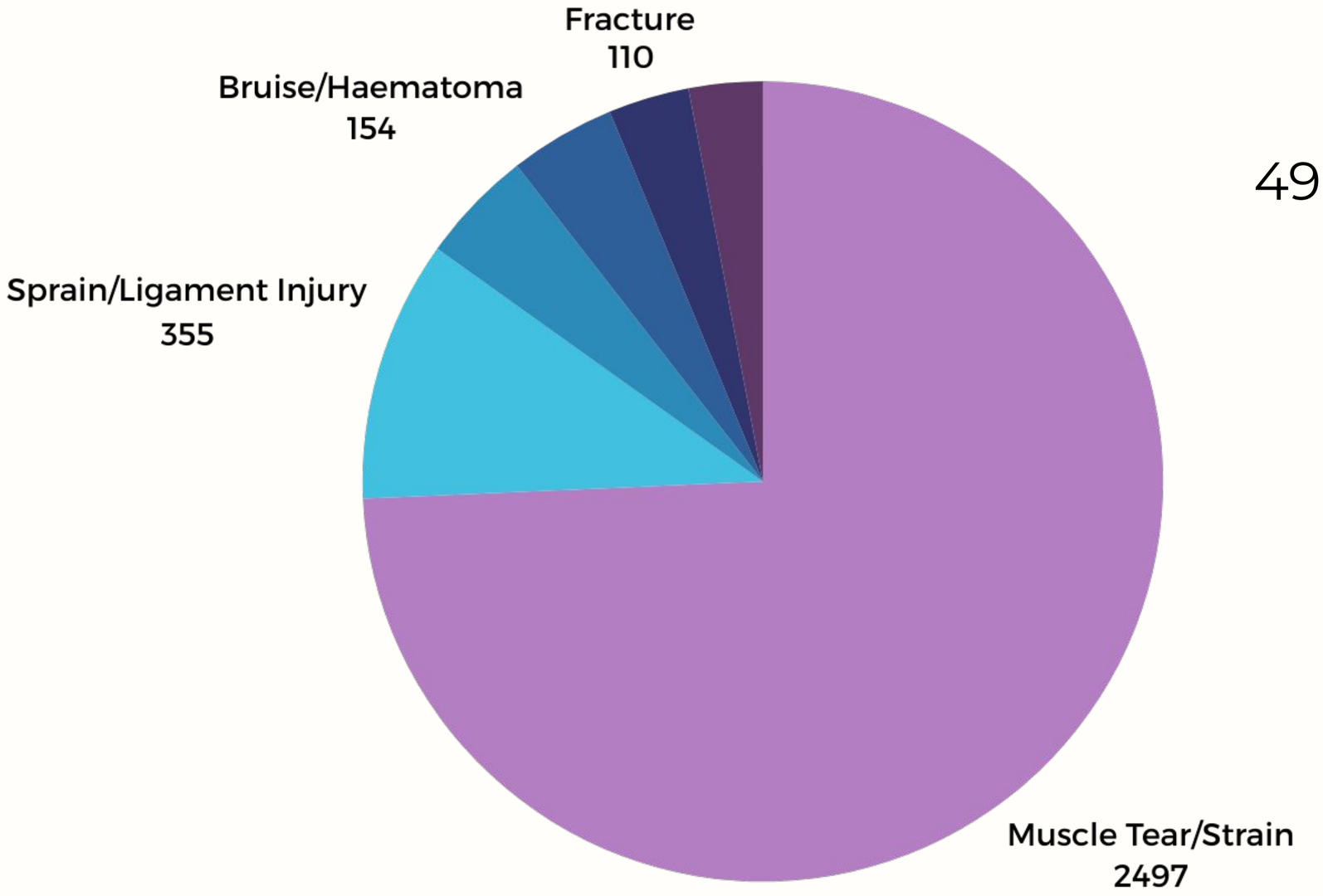
- From the years 2009-2019

# DATA FINDINGS

# Most Occurring Injuries (Body Region)



# Most Occurring Injuries (Type)




49% of Players experienced a recurring injury

# VARIABLES/FEATURES

## Player

- Age
- Position
- Height/Weight
- BMI*(Body Mass Index)

## History

- Seasons played
- Number of Injuries

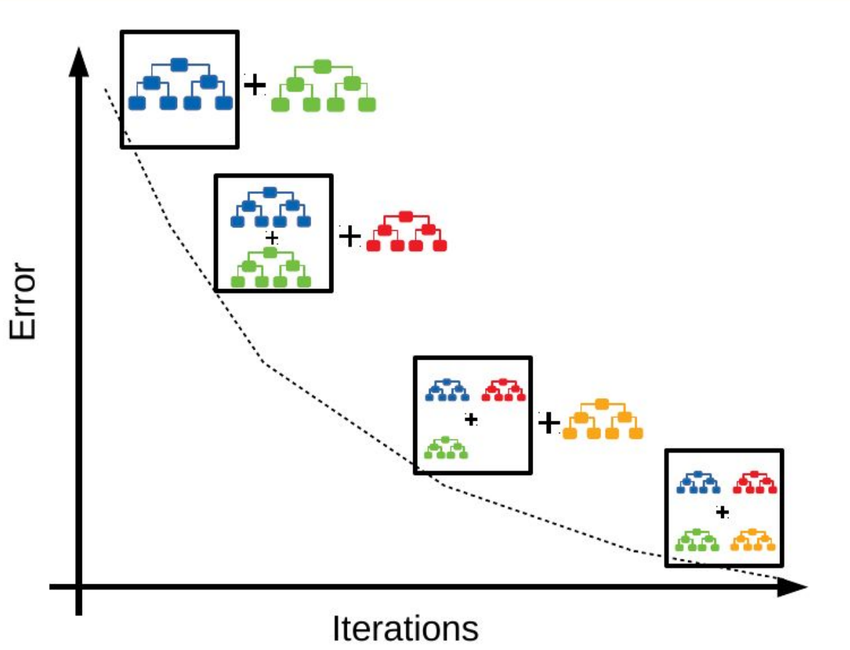
## Injuries

- Injury type
- Injury severity
- Injury Reccurence

# Model Selection

## Gradient Boosting

- Interpretability
- Variety of data types
- Parameter tuning



# ANALYSIS

Recall: .79

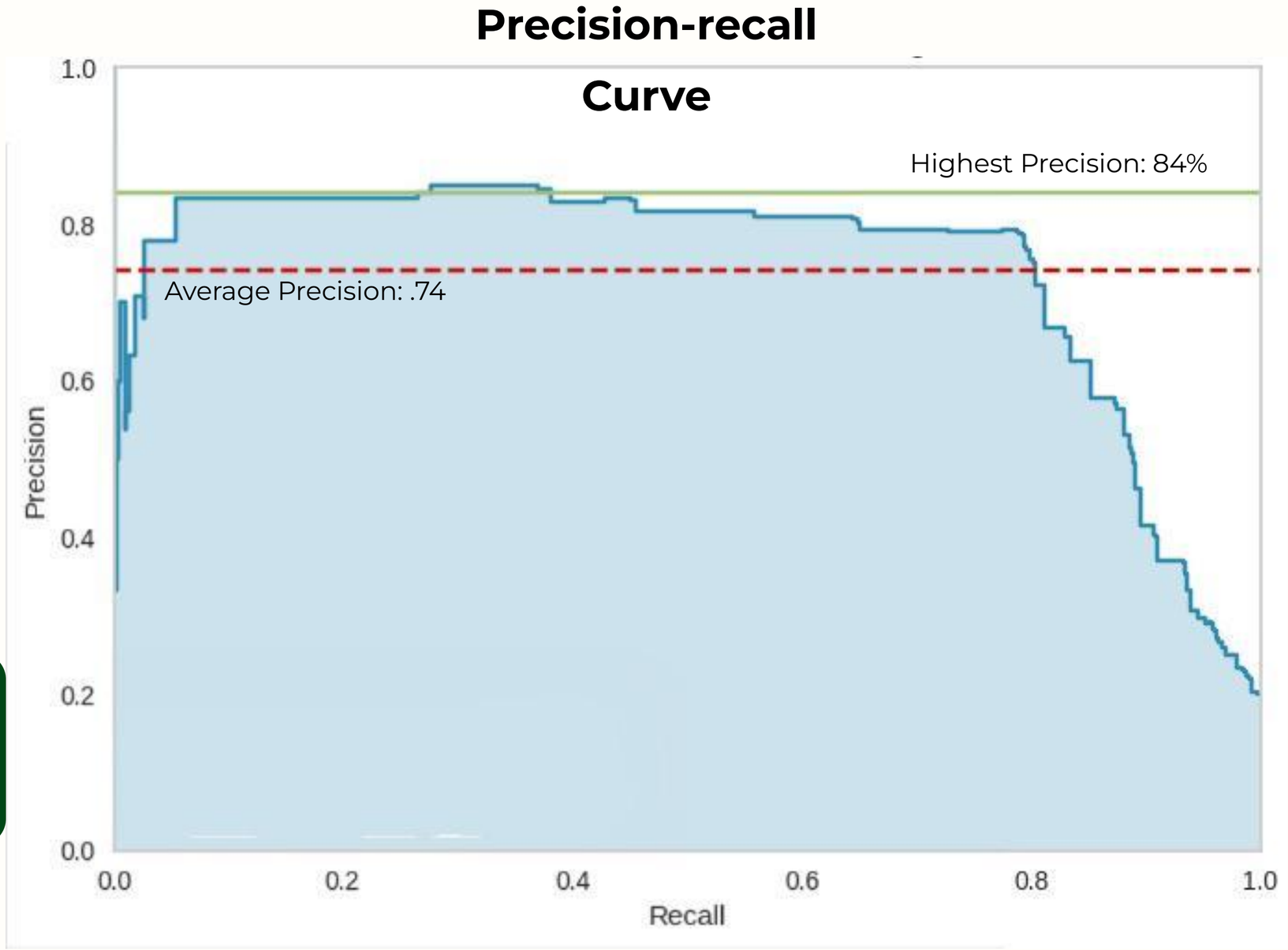
Accuracy: .80

Precision: .81

F1 Score: .69

Baseline Accuracy: .72

Age, Club Net-Worth and Reccurance were significant



# Use of our project


The current state of our project is customized for stakeholders in management and coaching roles, enabling them to make decisions related to player valuation and playing time.

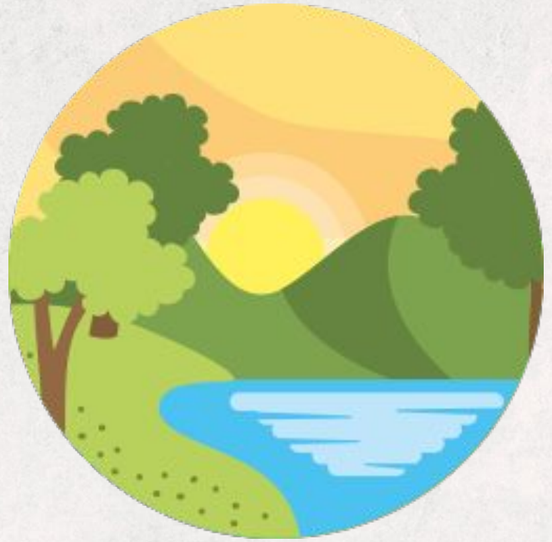
# Potential Project Impact

Similar non-contact injuries among other athletes (cyclists and runners) could also be predicted with minor adjustments.

# Limitations and Drawbacks


Lack of GPS & Training data

- Intensity
- Speed
- Strength



No environmental & weather conditions

- Rain, Snow, Hot, Cold

# Plan of Action

- Teams to provide us with data during the off-season (June)

  - Transition into a neural network model
    - Complex Relationships
    - Non-linear Relationships
    - Scalability

- Share model with sports scientists/trainers to work on injury prevention plans. (Strength, Flexibility, Rest)

# THANK YOU!

# Citations


Owen, A., Wong, D. P., Dellai, A., Paul, D.J., Orhant, E., & Collie, S. (2013). Effect of an injury prevention program on muscle injuries in elite professional soccer. The Journal of Strength and Conditioning Research, 27(12), 3275-3285. https://doi.org/10.1519/jsc.0b013e318290cb3a

Piłka, T., Grzelak, B., Sadurska, A., Górecki, T., & Dyczkowski, K. (2023). Predicting Injuries in Football Based on Data Collected from GPS-Based Wearable Sensors. Sensors, 23(3), 1227. https://doi.org/10.3390/s23031227

Ehrmann, F., Duncan, C. S., Sindhusake, D., Franzsen, W.N., & Greene, D. (2016). GPS and Injury Prevention in Professional Soccer. The Journal of Strength and Conditioning Research, 30(2), 360-367 https://doi.org/10.1519/jsc.0000000000001093

Mandorino, M., Figueiredo, A., Cima, G., & Tessitore, A. (2021, November 28). A data mining approach to predict Non-Contact Injuries in young Soccer players. Sciendo. https://sciendo.com/article/10.2478/ijcss-2021-0009

# Additional References

- MLS Injury Data
- Player Strength based on age
- Age trends of players
- FIFA Census
- Transfermarkt.us