DS5007 Project:
Predicting Recovery Times for Football Injuries using LSTM

Shubham Choudhary
IIT Palakkad
212405015@smail.iitpkd.ac.in

Achyuthkrishnan Balagopal V
IIT Palakkad
112502001@smail.iitpkd.ac.in

Introduction

Injuries are an inevitable aspect of professional football, significantly impacting team performance, player careers, and club finances. Accurately predicting recovery time following an injury is crucial for effective squad management, rehabilitation planning, and strategic decision-making. Traditional approaches rely on clinical expertise and historical averages based on injury type. This project proposes a Long Short-Term Memory (LSTM) neural network to predict recovery time for football injuries using data taken from the website Transfermarkt and published on Kaggle

Problem Statement

This project focuses on trying to predict recovery times for various injuries suffered by football players by building a Long Short-Term Memory (LSTM) neural network on data gathered from the website Transfermarkt.

Motivation

Recent advances in deep learning have shown promise in sports injury analytics. [1] demonstrated the potential of RNN-LSTM architectures for sports rehabilitation outcome prediction, while [2] achieved 91.5% accuracy using LSTMs for injury prediction, outperforming CNNs. However, these studies focus primarily on injury occurrence prediction rather than recovery time estimation—a more actionable metric for medical staff and team management. This project addresses this gap by developing an LSTM-based model to predict recovery duration for football injuries. By comparing LSTMs against CNNs and traditional machine learning baselines, we demonstrate the value of temporal modeling for accurate recovery time prediction in professional football.

Dataset

The dataset used for this project is 5.7M+ Records -Most Comprehensive Football Dataset. It was obtained by scraping the website Transfermarkt which contains comprehensive information about various football players and clubs across the world. The dataset contains more than 144,000 injury records. But a few entries are not being considered for this study. This study focuses solely on injuries sustained on the pitch. Hence player absences due to illnesses or classed as unknown are not considered for this study.

Methods

The dataset 5.7M+ Records -Most Comprehensive Football Dataset has many information which is irrelevant to this study, such as player ratings for a match, transfer value etc. These are not being considered. The dataset has to be cleaned to obtain only the relevant information. We propose following on from the work

of [1] and [2] and applying LSTM to build a predictive model for recovery time. If time permits, we would also do a comparative study on LSTM vs CNN as in [2].

Novelty

Previous works have focused mostly on the chances of occurence of injury [2] or the possibility of getting a positive outcome from certain rehabilitation techniques [1]. Our work tries to predict the recovery time for a particular injury based on factors such as injury type, player's age, height and previous injury record focusing solely on professional football players.

Expected Outcomes

- Predict recovery time for various injuries depending on factors such as age, height, previous injury record, etc.
- Compare CNN and LSTM accuracy for this problem

References

[1] Yanli Cui. An efficient approach to sports rehabilitation and outcome prediction using rnn-lstm. *Mobile Networks and Applications*, pages 1–16, 2024.

[2] Mohammad Mohsen Sadr, Mohsen Khani, and Saeb Morady Tootkaleh. Predicting athletic injuries with deep learning: Evaluating cnns and rnns for enhanced performance and safety. *Biomedical Signal Processing and Control*, 105:107692, 2025.