

Final Project Report

Football Talent Identification and Match Outcome Prediction System

Course:

Instructor:

Student: Mobin Golkarian

Student ID:

University / Faculty:

This report presents the design and implementation of a data-driven application for football talent identification. The proposed system consists of three main stages: (1) talent identification based on textual and physical data, (2) movement quality analysis using video-based pose estimation, and (3) match outcome prediction using team-level statistics.

1. Problem Statement

Football talent identification is traditionally based on subjective evaluation by coaches and scouts. Such assessments can be time-consuming and inconsistent. This project aims to provide an objective and repeatable framework that leverages data analysis and movement assessment to support decision-making in talent identification.

2. Project Objectives

- Identify football talent level (Beginner, Semi-Professional, Professional)
- Analyze player movement quality from video using pose estimation
- Predict match outcomes (Home Win, Draw, Away Win)
- Develop a lightweight and fast application with an intuitive user interface

3. Datasets

Two datasets with an Iranian football context are used in this project. The datasets are synthetically generated based on realistic ranges and can be replaced with real-world data.

Dataset	File Name	Purpose	Records
Athletes	athletes_iran_football.csv	Text-based talent identification	500
Matches	matches_iran_football.csv	Match outcome prediction	600

4. Methodology

4.1 Text-Based Talent Identification

Player attributes such as age, height, weight, training experience, sprint time, VO2max, agility score, shot power, passing ability, and stamina are used as input features. A machine learning classifier predicts the talent level along with class probabilities.

4.2 Video-Based Pose Analysis

Short video clips of football actions (running, shooting, change of direction) are analyzed. Pose landmarks are extracted and used to compute simple technique indicators such as knee angle, hip angle, and movement stability. Frame sampling is applied to ensure fast processing.

4.3 Match Outcome Prediction

Match prediction is performed using team-level features including home and away teams, team ratings, recent form, and expected goals (xG). The output includes the predicted result and the probability of each possible outcome.

5. Implementation and User Interface

The application is implemented using Streamlit and provides separate tabs for each stage of the system. Pre-trained models are loaded at runtime to ensure fast response and low computational cost.

6. Results and Evaluation

Experimental results on the provided datasets demonstrate that the system can achieve reasonable performance for both talent identification and match outcome prediction. These results are illustrative and can be improved using real-world data.

7. Conclusion and Future Work

- Replace synthetic datasets with real football performance data
- Extend pose analysis to football-specific actions such as kicking and jumping
- Improve model accuracy using larger datasets and advanced features