Smart Squash Talent Identification & Match Prediction System

Final Project Report (3–4 pages)

Sport: Squash — talent identification using text data, pose detection on videos, and match prediction.

1. Problem Statement & Objectives

We build a lightweight application that supports sports talent identification for squash. The system works in three stages: (1) text-based talent recommendation using fitness and background data, (2) video-based movement assessment using pose estimation, and (3) match outcome prediction using historical-like match statistics.

Outputs: recommended squash level (Beginner/Intermediate/Advanced), technique quality score from video, and Win/Lose prediction for a match.

2. System Overview

The delivered artifact is a Streamlit application (fast UI, minimal dependencies) plus two trained ML models. Pose estimation uses MediaPipe (efficient on CPU) and extracts simple biomechanical features (knee angle, stability proxies, motion proxy).

3. Datasets (Iran-oriented)

To satisfy the requirement of an Iranian dataset, we provide two structured datasets with Iranian provinces and Persian categorical values. The datasets are synthetic but realistic: ranges and distributions follow common field tests used in Iranian sport education/training forms (e.g., agility tests, 20m sprint, beep test, reaction time) and are suitable for demonstration and evaluation.

Dataset	Rows	Columns	Notes
athletes_iran_squash.csv	800	15	Class counts: {'Beginner': 402, 'Intermediate': 386, 'A
matches_iran_squash.csv	1800	13	Includes ratings, ages, experience, winner label

4. Methods

Stage 1 — Talent Identification (Text): We train a multi-class classifier (RandomForest) over numeric fitness metrics and categorical variables (province, dominant hand, previous sport). Categorical features are one-hot encoded.

Stage 2 — Pose Detection (Video): MediaPipe Pose estimates body landmarks per frame. We compute: (i) mean and std of knee angle (hip-knee-ankle), (ii) stability proxy via std of hip/shoulder vertical movement, and (iii) motion proxy via frame-to-frame hip displacement. A heuristic score in [0,1] summarizes technique quality.

Stage 3 — Match Prediction: Using match records, we create difference features: rating_diff, age_diff, exp_diff, plus home indicator, and train a binary RandomForest classifier to predict winner (A wins vs B wins).

5. Application UI & Performance Considerations

The UI is organized into four tabs: (1) Text-based talent form, (2) Video upload and pose overlay, (3) Match prediction panel, (4) Dataset/model info. To keep the application fast and small, video processing is capped by max_frames and sample_every.

Speed: frame sampling reduces compute. Size: models are small (tree-based, depth-limited) and stored with joblib. The app runs fully offline after installing dependencies.

6. Results (Demo Metrics)

Component	Metric (demo split)	Value
Talent model	Accuracy	0.825
Match model	Accuracy	0.689

7. Conclusion & Future Work

This project demonstrates an end-to-end pipeline for squash talent identification and match prediction. Future improvements: collect real Iranian squash datasets (federation/club-level), use temporal models for rallies, integrate shot classification, and calibrate pose metrics with expert-labeled technique scores.

Deliverables: Streamlit app, datasets, trained models, and this report.