# **Group 10 Project Proposal: Tennis Ball Landing Predictions**

## Abstract

This project addresses the problem of predicting tennis shot landing positions from video. We propose a two-stage Convolutional Neural Network (CNN) approach. First, a CNN is trained to identify the frame in which the ball is hit, using manually tagged data. Second, the identified hit frame is passed into a second CNN to predict the ball's landing coordinates.

This project is motivated by a potential for genuine commercial value - this method may be applicable in many ways, for example: for automated line calling in competitive tennis, for dynamic ball machines to predict where to move, or on low-latency architecture to predict tennis outcomes before odds have adjusted on betting markets.

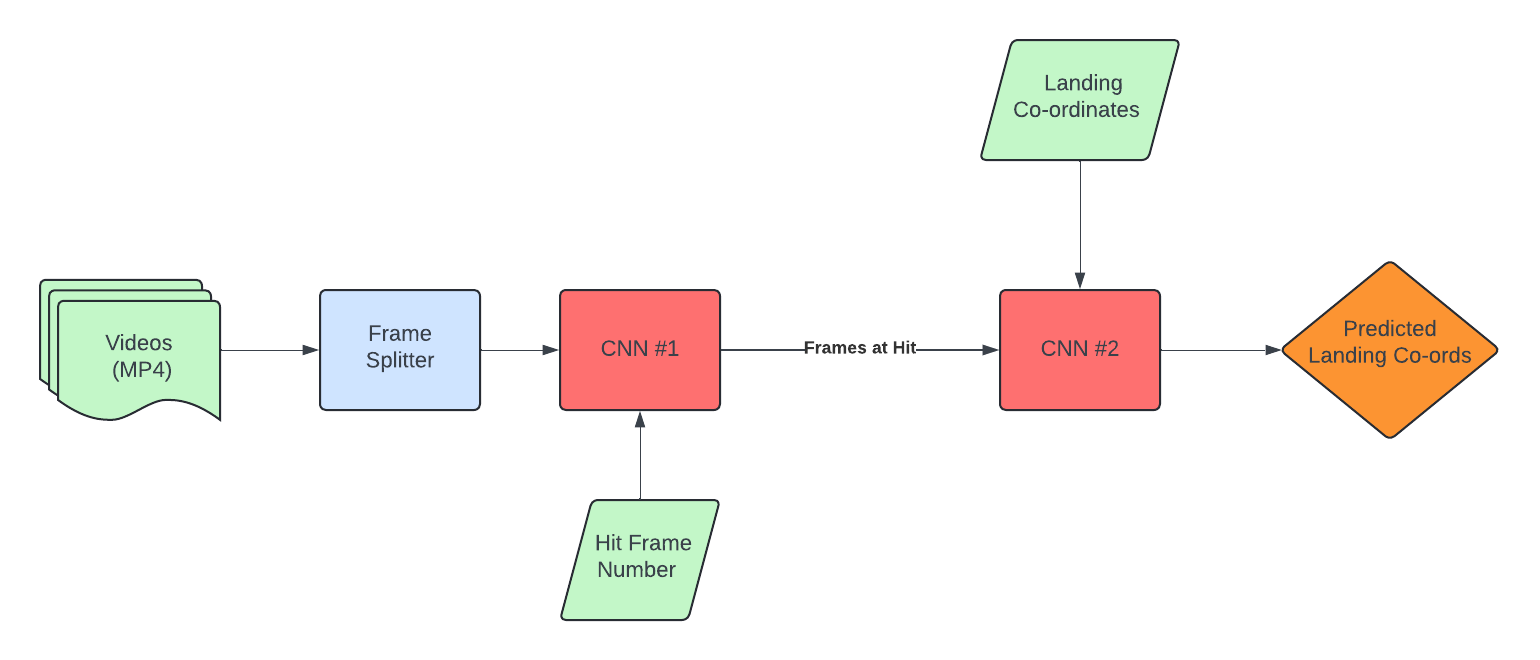
## Introduction

In recent years, the use of AI in the sports and sports betting industries has grown rapidly, aiding people in obtaining the best and most complete information possible about different sports and athletes. AI is also able to provide in-depth analysis and interpretations of the increasing mass of data that is collected by both sporting and betting companies. An example of this in tennis is TennisViz, which tracks the swings and ball speed to deliver insights into different aspects of a tennis players’ performance, like shot quality and serve effectiveness. This, however, is often done by processing the data after the shot has occurred, rather than predicting the outcome of a shot live. After looking into the market there are examples of AI models being used on data from video footage, in-racket technologies, wearable trackers, or a combination of them, but little to no usage of CNNs in shot prediction. The advantage of predicting shot outcome from the shot position is that you can take a view on the future, rather than explaining the past, as many of the existing models do. The architecture of a CNN allows the model to identify many key aspects automatically including the speed of the ball, the speed of the racket movement, distance from baseline, distance from net - even court surface. With enough diverse data, this model should be able to perform well on multiple camera angles and in any court location - just requiring a visual stream to operate on.

## Methods and Algorithms

Currently our intended architecture consists of a 2 stage CNN pipeline with the following process:

1. Video Preprocessing: As the videos capturing each full swing and ball landing from the player are in MP4 format we will need to split these into the individual image frames.
2. CNN #1 Hit Detection: The sequence of individual image frames is passed into the first CNN to identify the region of frames of which the ball is hit, making this a regression and not a classification task.
3. CNN #2 Landing Prediction: After receiving the predicted hit frame(s) from CNN #1, the second CNN takes the selected frame(s) as input and predicts where in the opponent’s section of the court the ball will land. This CNN is trained using labeled data that includes the ball’s To-Closest-Doubles-Sideline-Distance (meters) and To-Baseline-Distance (meters). These two values form a 2D coordinate that represents the landing location of the ball on the opponent’s side of the court.
4. Output: The final output of this workflow is the set of predicted coordinates.

Figure 1: Workflow of the prediction pipeline

## References

Lai, Kalin Guanlun; Huang, Hsu-Chun; Lin, Shang-Yi; Lin, Wei-Ting; Lin, Kawuu Weicheng (2024), “Tennis Shot Side-View and Top-View Data Set for Player Analysis in Tennis”, Mendeley Data, V4, doi: 10.17632/75m8vz7jr2.4

Wang, C.-Y., Lai, K. G., Huang, H.-C., & Lin, W.-T. (2024). Tennis player actions dataset for human pose estimation. *Data in Brief*, *51*, 110665. <https://doi.org/10.1016/j.dib.2024.110665>