

Ignacio Rodriguez¹, Ria Verma¹, Romina Cano¹, Jumpei Sakamoto¹, Tomas Guiloff¹, Henry Wang^{1*}, Christina Chase¹, A.E. Hosoi¹, Johsan Billingham²

¹ MIT Sports Lab, Cambridge, MA, USA; ² Fédération Internationale de Football Association, Zürich, Switzerland

*Corresponding Author: hwang21@mit.edu

Introduction

- Automatic detection of touches on the ball during football matches from tracking data or sensor measurement is an increasingly common data stream.
- Applications for referee assistance (offside kickpoint detection) motivate ensuring touch data are accurate.
- This work proposes a reproducible way for governing bodies and providers to measure ball-touch timestamp accuracy and tune tolerance to match operational risk of errors, complete with templates for data quality reporting.**

Methodology: Criterion Touch Dataset Creation

- Using game footage synced with UTC time, we tagged ball touches from randomly selected two-minute chunks of three FIFA Intercontinental Cup 2024 (FIC24) knockout stage matches.
- For each chunk, **two annotators recorded all touches**, with a **third corrector or validator** resolving complete disagreements. Minor disagreements were resolved using the mean timestamp between annotators or the later timestamp, depending on the scenario. (Fig.1)

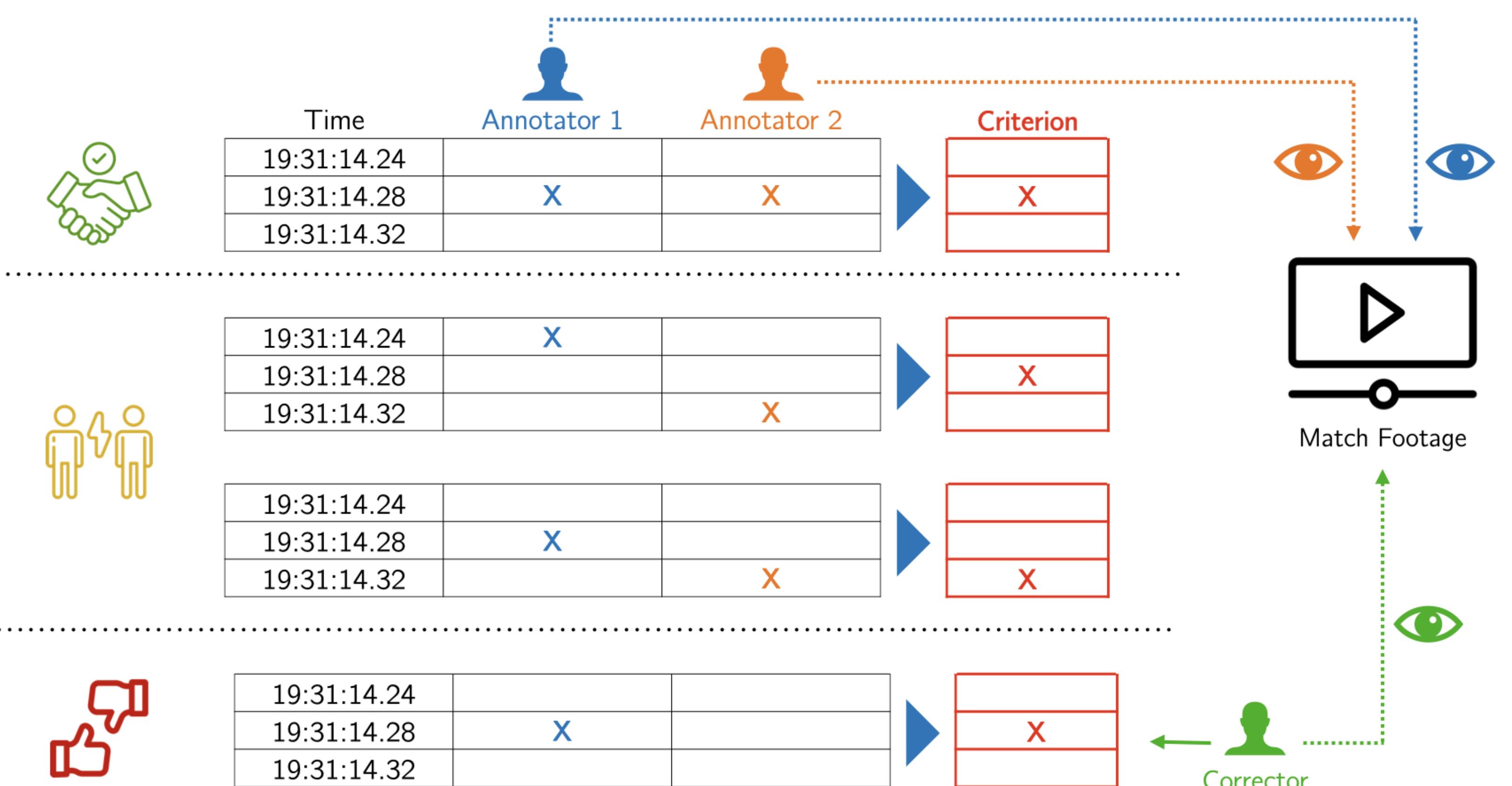


Fig 1. Annotation process for the criterion dataset, split by complete agreement, minor disagreement, and complete disagreement cases (top to bottom).

Methodology: Comparing Touch Timestamps

- The criterion dataset was used as the standard for comparing touch datasets from two external providers.
- Each touch event** is represented as a **symmetric bump** (a Gaussian kernel) **centered on its timestamp**. The bump's temporal width (σ) controls matching strictness, facilitating a continuous comparison.
- The overlap between datasets is maximized and computed using an assignment algorithm (Hungarian algorithm) that matches criterion touches to provider touches [1].
- This methodology **rewards events that are close in time**, making it less sensitive to small differences and providing a more nuanced comparison with touch data quality reported as:
 - Accuracy:** How well do the provider's touches match the criterion overall?
 - Precision:** How many provider touches are actual touches?
 - Recall:** How many actual touches are detected by the provider?

a) Illustration of three touch timestamps modeled as bumps

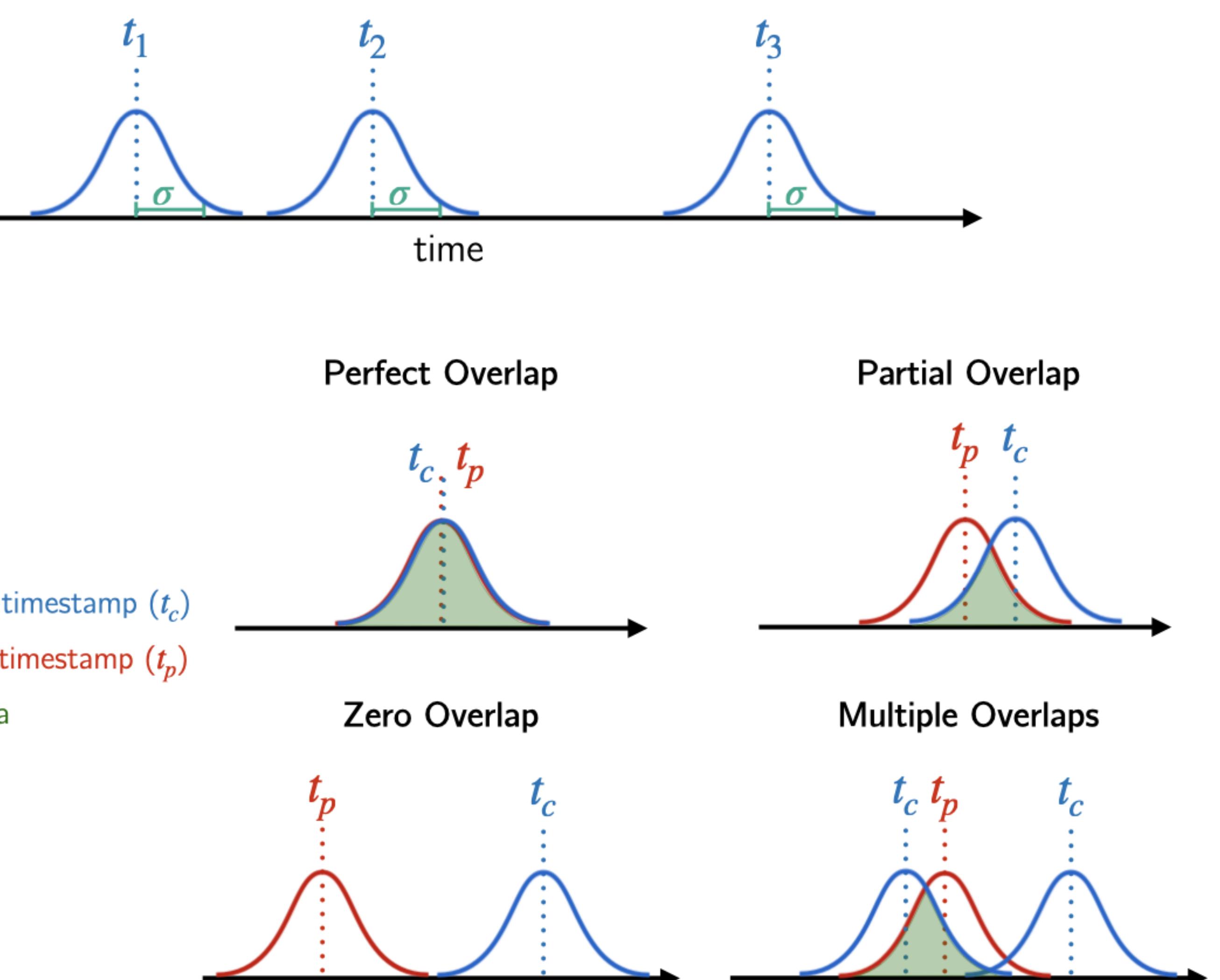


Fig 2. (a) Illustration of touch timestamps modeled as Gaussian bumps. (b) Possible modes of overlap between criterion and provider touches.

Results

- The criterion dataset creation procedure requires about **one person-hour** to manually tag five minutes of match footage.
- A **standardized test manual and data quality report template** provide a procedure to collect, assess, and report touch data performance from real matches.

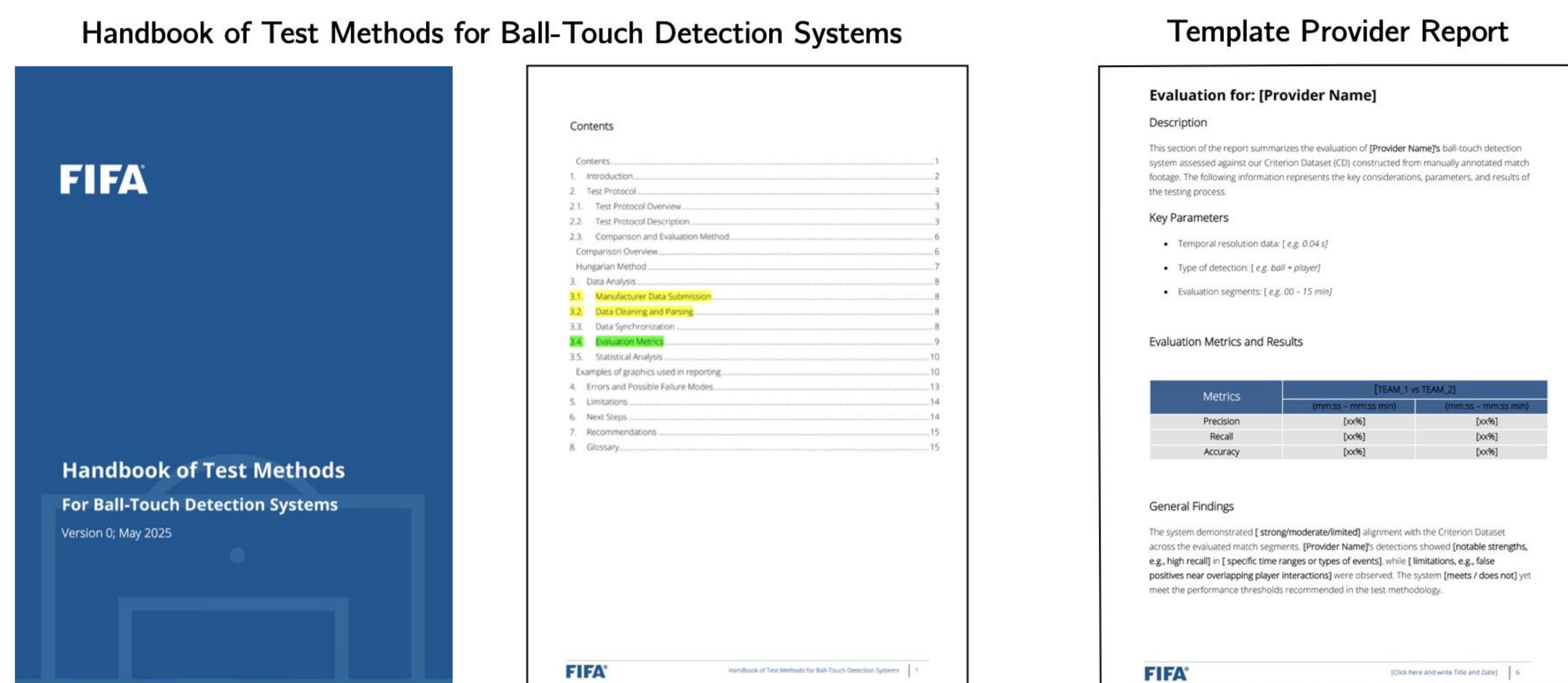


Fig 3. A handbook of test methods (left) and provider-specific report templates (right) were created to reproduce the methodology.

Limitations & Next Steps

- The manual tagging process remains resource intensive and spans a small fraction of total match time.
- The precision of the criterion dataset is **limited by the video frame rate** and the **accuracy of time alignment** between the video and UTC time used for syncing with provider touch data.
- The “correct” choice for bump width (σ) is not obvious and needs to be selected carefully by the practitioner.
- We suggest primary next steps are the release of an open-source framework to **facilitate the application of this methodology across more providers and match contexts across the industry**.

Acknowledgements

This project was done as part of the MIT course 2.98, Sports Technology: Engineering and Innovation. We thank Peko Hosoi, Christina Chase, and the entire teaching team as well as Katie Mills for their mentorship and expertise.

References

- [1] Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2), 83-97.