7. Supplementary

7.1. Video assistant referee system software

The design of both the VARS annotator and the VARS interface draws inspiration from the VAR room. To enhance the user experience, a grid layout is used to display all available perspectives synchronously. The objective is to have an easy to use interface to annotate and to predict different properties of an action.

VARS annotator. To increase the speed of the annotation task, we build a VARS annotator (Figure 6a), which shows all the available clips of an action simultaneously. The VARS annotator allows for individual adjustment of the annotated moment for each clip to achieve temporal alignment, speed adjustment for replays, and annotation of all the properties.

VARS interface. The VARS interface has the same interface as the annotator. It enables easy access to all available perspectives for a particular action. The multi-task VARS, which achieved the best results on the test set, is built directly into the interface, allowing for immediate analysis of selected videos. The VARS interface offers top two predictions for the type of foul classification, as well as the offense and severity classification for the selected videos. Furthermore, for each prediction, the VARS interface shows the confidence score of his prediction.

In Figure 6b, we can see an example of how the VARS interface looks. In this example, the VARS correctly predicts the type of foul, and tells that the action was indeed a foul, which leads in a penalty for the attacking team in this example. For the severity classification, the VARS was uncertain whether to assign no card or a yellow card.

7.2. Dataset

7.2.1 Property explanations

This section explains in detail each property. Furthermore, we illustrate multiple examples of our dataset in Figure 7.

Was it a foul? The task of identifying fouls is a critical and challenging aspect of the role of a referee and VAR. They must determine whether an action is a foul or not, in accordance to the laws of the game [27]. For each action, we determined whether an action is an (i) offence (an action which breaks/violates the Laws of the Game [27]), (ii) no offence (did not break/violate the Laws of the Game), or (iii) between (if the action lays inside a grey area). In some cases, both decisions may be correct and the final decision depends on the interpretation of the rules of an individual.

It is worth mentioning that, for each clip of the same foul, we make the same annotation. During the annotation process, we looked at all the clips and took a global and final decision which is the same for each clip of the same foul.

Was there any contact? Another important property which we annotated was if there was any contact between two players during an action. We annotated for each foul (i) with contact (if there was contact between players), or (ii) without contact (if there was no contact). This property is important because a foul with contact such as a tackling, holding, elbowing result in a direct free kick, while a foul without contact such as a simulation or dangerous play will result in an indirect free kick.

Did the player touch the ball with his hand/arm? This property annotates if the player touches the ball, deliberately or not, with his hand or arm. We annotate (i) handball (if a players touches the ball with his hand/arm), or (ii) no handball (if the ball did not touch the hand/arm). This property only states whether the ball touched the hand or arm and does not indicate whether the handball is punishable or not. An important note to make is that the upper boundary of the arm is in line with the bottom of the armpit [27].

Was the upper or under body used in the action? This property annotates which part of the body was used during an action. We differentiated between (i) under body (which corresponds to the use of the food or the leg), or (ii) upper body (which corresponds to the use of shoulder or the use of arms).

With which part of the upper body was the action made? In the case where we annotated the previous property with ”upper body”, we further split between (i) use of shoulders, or (ii) use of arms.

Class of the action. This property annotates the type of action. In total, we have 9 different classes:

1. **Tackling.** The sliding movement of a player towards an opponent who is in possession of the ball and legally or illegally using his foot or leg to try to take the ball away.

2. **Standing tackling.** The movement (not sliding) of a player towards an opponent who is in possession of the ball and legally or illegally using his foot or leg to try to take the ball away.

3. **Holding.** Occurs when a player’s contact with an opponent’s body or equipment impedes the opponent’s movement [27].

4. **Pushing.** The action of using the upper body to push an opponent away.

5. **Challenge.** Physical challenge against an opponent, using the shoulder and/or the upper arm [27].

6. **Elbowing.** The use of arms (and frequently the elbows) as a tool or a weapon to gain an unfair advantage in aerial challenges, physical battles, to create space or to intimidate other players.

(a) **VARS annotator.** The annotator may browse simultaneously the synchronised videos either at regular speed or frame by frame. He can annotate all 10 properties and adjust the annotated point of contact for each clip separately and temporal align the different clips by modifying the speed and offset of the clips.

(b) **Video Assistant Referee System interface.** The interface of the VARS shows the ground truth of the action and his top 2 predictions for the foul classification task, and the offence and severity classification task. In this example, the VARS correctly predicts the foul resulting in a penalty.

Figure 6. **Views of the VARS interfaces.** (a) shows the interface for the annotation process. (b) shows the interface of the VARS for displaying its results.

7. **High leg.** A movement where a player swings his foot close to and above the waist of an opponent.

8. **Dive.** An action which creates a wrong/false impression that something has occurred when it has not, com-

mitted by a player to gain an unfair advantage. [27]

9. **Don't know.** Corresponds to anything which can not be classified in one of the classes above.

How severe was the foul? For each foul, we annotated the severity of the foul by a scale from 1 to 5:

- **1:** a careless foul which is when a player shows a lack of attention or consideration when making a challenge or acts without precaution. No disciplinary sanction is needed. (No card) [27]
- **2:** a borderline foul between careless and reckless. We are in a grey area where both "no card" or "yellow card" would be correct.
- **3:** a reckless foul which is when a player acts with disregard to the danger to, or consequences for, an opponent and must be cautioned. (Yellow card) [27]
- **4:** a borderline foul between reckless and violent. We are in a grey area where both "yellow card" or "red card" would be correct.
- **5:** a violent foul where a player exceeds the necessary use of force and/or endangers the safety of an opponent and must be sent off. (Red card) [27].

Did the player try to play the ball? When a player commits an offence against an opponent within their own penalty area which denies an opponent an obvious goal-scoring opportunity and the referee awards a penalty kick, the offender is cautioned if the offence was an attempt to play the ball; the offender is sent off if there was no possibility to play the ball [27]. We annotated, (i) "Yes", if the player tried to play the ball, or "No", if there was no possibility to play the ball.

Did the player play the ball? The final property annotates (i) "Yes'' if the defender touches the ball, (ii) "No" when the defender did not play the ball, or (iii) "Maybe" in the case where the quality of the video or the viewpoint on the foul is not sufficient to determine if the player touched the ball or not.

7.2.2 Dataset distribution

The distribution of all classes is provided in Tables 8 and 9. Most of the properties in the dataset have a high degree of imbalance, particularly the "Handball" property that determines if a player has made contact with the ball using their arm or hand. Nearly 99% of the actions recorded in the dataset do not involve any handball. Similar imbalances are observed in the "Contact", "Try to play the ball", "Played the ball", and "Offence" properties.

The "Bodypart" and "Upperbody part" properties are relatively less unbalanced, with a distribution of approximately 66% for the superior class and 34% for the inferior class.

8. Experiments

8.1. Per class analysis

Foul classification task

The performances for each class are summarized in the confusion matrix shown in Figure 8. Our analysis shows that the performance varies considerably across classes. The VARS often confuses all the illegal use of arm classes, like "Holding", "Pushing", and "Elbowing" as these fouls share some common characteristics and can involve similar physical movements. The model performs well in detecting "Tackling", but confuses it often with "Dive" as it struggles to distinguish between genuine and deceptive actions, which can be challenging due to the complex and dynamic nature of soccer games. However, the most challenging class for the VARS is "Challenge", which shares visual similarities with many other classes, making it difficult for the system to generalize properly during training.

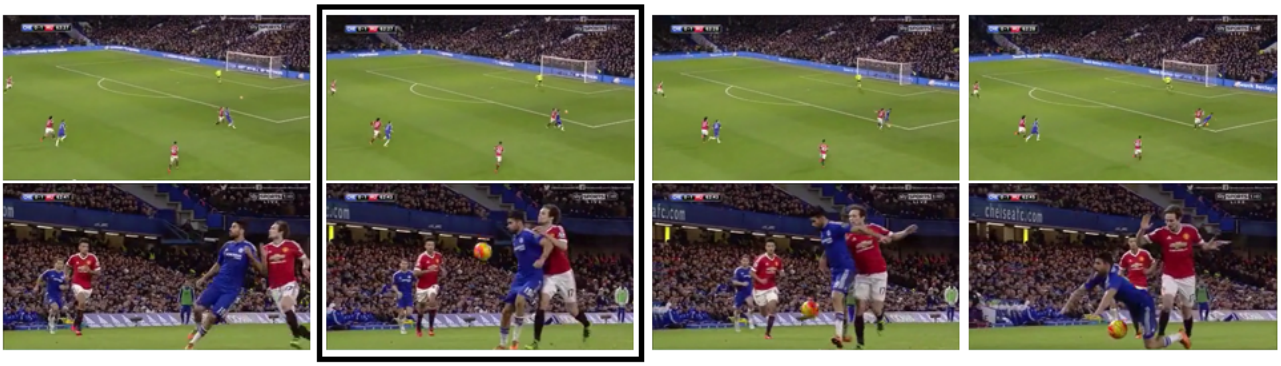
Offence and severity classification task

Figure 9 displays the confusion matrix for the offence and severity classification, revealing that the model frequently confuses classes with their neighboring classes. For instance, when the ground truth is "Offence + No card", the VARS often mistakes it for "No offence" or "Offence + Yellow card".

Indeed, the visual similarity between all the classes, especially with the neighbor classes, is very high. Small details, which very often can only be seen in a couple of frames, can differ between the actual class. Factors such as the speed of the foul, the point of contact, or the intention of playing the ball are critical criteria for deciding which class an action corresponds to. However, these criteria can be challenging to spot for a model and to differentiate between the different classes. Furthermore, there is only a small number of instances of "Offence + Red card" in the dataset, making it more challenging for the model to generalize. Despite all these difficulties, the VARS is still able to achieve an accuracy of 0.43.

(a) “Offence”, “Tackling”, “Yellow card”, “With contact”, “Under body”, “/”, “Played the ball”, “Tried to play the ball”, “No handball”

(b) “Offence”, ”High leg”, ”Red card”, “With contact”, “Under body”, ”/”, “Ball is not played”, “Tried to play the ball”, “No handball” and “No handball offence”



(c) “Offence”, “Challenge”, “No card”, “With contact”, “Upper body”, “Use of shoulder”, “Ball is not played”, “Tried to play the ball”, “No handball” and “No handball offence”

(d) “Offence”, “standing tackling”, “Yellow card”, “With contact”, “Under body”, ”/”, “Did not play the ball”, “Tried to play the ball”, “No handball”

Figure 7. **Dataset overview and ground truth.** We annotated the exact frame where the point of contact happens (depicted by the back box).

Fouls		Severity		Offence		Handball		Handball offence	
Class	**Prob.**	**Class**	**Prob.**	**Class**	**Prob.**	**Class**	**Prob.**	**Class**	**Prob.**
St. tackling	0.43	No card	0.55	Offence	0.85	Yes	0.99	Yes	0.82
Tackling	0.15	Yellow card	0.26	No offence	0.10	No	0.01	No	0.18
Challenge	0.13	NC/YC	0.15	Between	0.03				
Holding	0.12	YC/RD	0.02						
Elbowing	0.05	Red card	0.01						
High leg	0.03								
Pushing	0.02								
Dive	0.01								

Table 8. Distribution of classes in our SoccerNet-MVFouls dataset.

Bodypart		Upperbody part		Try to play the ball		Played the ball		Contact	
Class	**Prob.**	**Class**	**Prob.**	**Class**	**Prob.**	**Class**	**Prob.**	**Class**	**Prob.**
Upperbody	0.36	Arms	0.66	Yes	0.92	Yes	0.10	With	0.97
Underbody	0.64	Shoulder	0.33	No	0.08	No	0.87	Withou	0.03
						Maybe	0.02		

Table 9. Distribution of classes in our SoccerNet-MVFouls dataset.

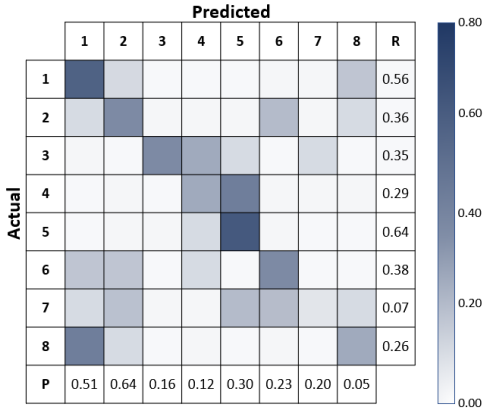


Figure 8. **Confusion matrix for the type of foul classification.** The VARS demonstrates good performance in classifying "Standing Tackling", "Tackling", and "Elbowing". However, the model struggles with "Challenge" and frequently confuses it with other classes. 1: Standing Tackling, 2: Tackling, 3: High Leg, 4: Pushing, 5: Elbowing, 6: Holding, 7: Challenge, 8: Dive, R: Recall and P: Precision.

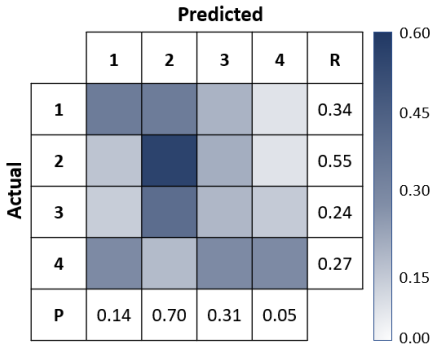


Figure 9. **Confusion matrix for the offence and severity classification.** The VARS shows good performance for the "Offence + No Card" class. The model confuses the classes "No Offence" and "Offence + Yellow Card" with "Offence + No Card". For the class "Offence + Red Card" the model is not able to provide good results due to the low amount of samples in the dataset. 1: No offence, 2: Offence + No card, 3: Offence + Yellow card, 4: Offence + Red card, R: Recall and P: Precision.