

Reviewer 2:

This is a reliability study validating the consistency and degree of agreement from video analysis of soccer matches. The authors acknowledge the limitations of their study which is a plus for the review report. However, the importance and impact of their work is not efficiently communicated and it is suggested they work to improve that aspect. In that context, the Discussion could be improved by linking the impact of correctly estimating an ACL injury from a video analysis to the next steps, for instance providing feedback to both coaches and athletes about risk factors, technique correction and perhaps early detection of other risky situations.

R: We agree that the practical implications of our work were not sufficiently emphasized. We have now added a paragraph in the Discussion detailing how reliable systematic video analysis can directly impact clinical and coaching practices, specifically by providing practical feedback to coaches and athletes regarding risk factors, technique correction, and the early detection of risk movement patterns.

"Establishing this high reliability is not merely a methodological exercise, it has significant clinical and practical implications. Consistent and precise video analysis allows sports medicine professionals to accurately identify risky movement patterns and contextual triggers that precede ACL injuries. This robust data can then be translated into practical feedback for coaches and athletes, who can implement targeted technique corrections during training, adjust tactical behaviors that expose players to vulnerable positions, and ultimately improve the early detection of other risk situations, creating a more effective feedback loop for injury mitigation strategies."

Specific other comments and observations follow below:

- Abstract: no comments
- Introduction: no comments
- Methods:

L58: a symbol indicating trademark properties or copyright ones for the platform should be included here.

R: Thank you for this observation. We have added the registered trademark symbol (®) to the Noisefeed platform in the Methods section.

L76: "consistent with previous literature, was estimated at 40 ms following initial contact (IC)...": the authors should report the studies that provided evidence about the suggested time period of the 40ms to justify its use. Why wasn't this 40ms time period used in all analyses? L72-73 report a shorter by almost 50% time period for analysis, which creates a serious discrepancy in how the videos were analyzed.

R: We agree with the need to include references to support this point. Therefore, we have added the appropriate references [Della Villa, et al. (2020); Buckthorpe, et al. (2024)] to justify the 40-ms timeframe for the visible tissue failure following initial contact. Regarding the reported discrepancy, we respectfully clarify that two distinct time metrics are involved here. The "12-15 seconds before to 3-5 seconds after" refers to the total duration of the video clip assessed to evaluate the contextual, situational, biomechanical, and anatomical variables. Conversely, the "40 ms" refers to the specific micro-timeframe (milliseconds) within that clip where the actual biomechanical failure of the ligament occurs.

L77-78: This needs to be clarified more. If understood correctly, the majority of the video footage on which the reliability analysis was based on, came from 2D data, i.e. single camera angle? if so, the authors must inform readers about their methodological procedures in more detail.

R: We have added more details to the "Video Acquisition and Processing" subsection. We clarified that all video footage was based on standard 2D broadcast data, from the main camera angle, which was subsequently supplemented by slow-motion or regular-speed replays from alternative angles, only when these were provided by the television broadcaster.

"Video footage relied on standard 2D broadcast data from the main broadcast camera angle. When available, multiple 2D camera angles and slow-motion replays were added to the sequence."

L82-95: it is suggested to present the investigated variables in a more organized way; divide them in contextual and biomechanical and present them accordingly.

R: This is a very helpful suggestion for easier reading. We have reorganized the Observational Checklist and Variables subsection, clearly dividing and presenting the variables under categories: *"Contextual and situational variables"* and *"Biomechanical and anatomical variables"*.

L114-115: the authors need to define the professional background of the raters and if they received any level of training with the software or the checklist.

R: We thank the reviewer for pointing this out. We have updated the Methods section to explicitly detail the diverse multidisciplinary professional backgrounds of our raters (SO has a background in biomedical engineering; KS in sports science; and MC in epidemiology, biostatistics, and health research). The familiarization process with the software and checklist has also been detailed in the "Video Analysis Procedures" section.

"The evaluation team members have diverse professional backgrounds, including biomedical engineering (SO), sports science (KS), and epidemiology, biostatistics, and research in health (MC, currently a medical student). Before data collection, the raters underwent a familiarization process at a dedicated consensus meeting to review the checklist, align operational definitions, and establish standardized evaluation criteria."

L118: are there references to support the minimum of 30 days interval?

R: We appreciate the reviewer's comment. The 30-day interval (approximately 4 weeks) was selected as a conservative "washout" period to completely mitigate recall bias (memory effect). Sim & Wright (2005) reported that an interval of 2 to 14 days is usual, but that this depends on the variable being measured. In this case, as this is an observational reliability study using video analysis, it was considered that an interval of 2 to 4 weeks would be sufficient to ensure raters do not simply remember their previous assessment. We have updated the manuscript to clarify this rationale and added a methodological reference (Sim & Wright, 2005) that highlights the need for an adequate time interval to minimize memory effects in intra-rater reliability studies.

▪ Results:

L139: The authors do not report the outcome of the normality assessment analysis and if normality was violated, what measures did they take.

R: We have added a sentence at the beginning of both subsections of the Results to report that the preliminary assumptions of normality (Shapiro-Wilk test) were met for the quantitative variables.

"For quantitative variables, preliminary assessment confirmed that the data met the assumptions of normality (Shapiro-Wilk test, $p > 0.05$)." and "...again after confirming the assumption of normality (Shapiro-Wilk test, $p > 0.05$)."

L141-143: a similar statement is provided in the Discussion. To avoid repetitions, it is recommended that it is removed from this section.

R: We agree with the reviewer. To avoid repetition with the Discussion, we have removed the referenced sentence from the Results section.

"Overall, the systematic video analysis checklist demonstrated high reliability, though the visual estimation of player speed proved to be a significant limitation in both assessment settings."

L152-153: "body location" is an anatomical detail rather a biomechanical one; further, the word location is a general word and if the authors desire to use appropriate terminology, then perhaps they consider options like body part, affected area or affected body part, or anatomical area.

R: We would like to thank the reviewer for this terminological correction. We have replaced the general term "*body location*" with the more precise term "*anatomical area*" throughout the manuscript.

L156: In my opinion, including visual speed estimation in the video analysis checklist is erroneous. Subjective visual perception is highly unreliable and since the observer and the phase he/she observes are not moving in the same speed, this is problematic. Present results confirm the unsuitability of this parameter.

R: We agree with this perspective. Subjective visual perception of speed is highly problematic, as our findings confirmed. As also suggested by another reviewer, we have explicitly stated in the Discussion section that this parameter is unsuitable for visual estimation and recommended its removal or replacement with objective tracking data in future checklist.

"Consequently, it is strongly recommended that visual speed estimation be removed or substantially modified in future versions of systematic video analysis checklists, as it cannot be treated with the same level of confidence as contextual variables. Researchers should replace subjective visual assessment with objective data driven from synchronized optical tracking technologies, GPS wearable microtechnology, or specialized computer vision algorithms."

Nevertheless, we believe that the inclusion of horizontal and vertical speed variables in this reliability study, as well as the poor results obtained, was highly important in drawing these conclusions.

- Discussion:

L179: the study has a single purpose, so the adjective primary is unnecessary.

R: We have removed the word "*primary*", as suggested.

L238-242: This suggestion goes without saying. However, integrating qualitative video analysis with numerical data, changes the nature of this assessment method and transforms it into something else. Perhaps the authors could suggest the assessment of acceleration, which might be deduced by the change in body posture and the respective frames. So, perhaps by combining two pieces of information, two observers might agree better with each other. Of course, this scenario remains to be validated.

R: We thank the reviewer for this excellent and insightful methodological suggestion. We agree that, in the absence of objective tracking data, deducing acceleration or deceleration indirectly through qualitative changes in body posture across consecutive frames could be an alternative to the highly unreliable simple visual speed estimation. We have integrated this possibility into the Discussion section, but we made sure to mention that it requires full validation.

"Alternatively, in the absence of objective tracking data, future observational methodologies could explore deducing acceleration or deceleration indirectly by evaluating qualitative changes in body posture across consecutive frames. However, whether combining these pieces of observational information improves intra- and inter-rater agreement remains a hypothesis and requires formal validation."