

Peer-Review

Panchal, Dhruv. 2026. "Automated Fault Attribution for Formula 1 Collisions: Evidence for the Limits of Telemetry-Based Classification." *Journal of High School Science* 10 (3): 105–29. <https://doi.org/10.64336/001c.164963>.

you are asking the wrong questions (using the wrong features) of/in the model. See chatgpt file attached.

Run your model again - this time with relevant features. I am not saying you will extract major improvement; but your model will now be relevant to your title/objective and question.

Response to Reviewers

Reviewer 1

Reviewer comment: *"you are asking the wrong questions (using the wrong features) of/in the model. See chatgpt file attached. Run your model again - this time with relevant features. I am not saying you will extract major improvement; but your model will now be relevant to your title/objective and question."*

Author response: The manuscript now addresses this through a new set of FIA-aligned interaction features, introduced in Section 3.7, that capture things like closing speed between the cars, how much their paths overlapped, who had the inside line, and the gap between their braking points amongst others.

Reviewer 2

Reviewer comment: *"Please verify that all links to the references point to the correct source."*

Author response: All 17 references point to the correct source.

Open comment 2

Reviewer comment: *"The authors must disclose and acknowledge any assistance received in the preparation of this manuscript, including but not limited to editorial, technical, analytical, or writing support. All such contributions must be clearly stated in the Acknowledgments section."*

Author response: The Acknowledgments section now contains the specific contributions made by Claude Code.

Open comment 3

Reviewer comment: *"Include enough recent references along with foundational ones. Ensure references directly support your claims. Avoid 'padding.' Use credible sources (peer-reviewed journals, reputable books, official reports)."*

Author response: We added more references to code and models that were used. We avoided 'padding' references by only adding those that were used and not just related.

Thank you for addressing my comments. Several deficiencies remain. Please address my ongoing concerns below

1. Validation of FIA-Aligned Features Against Actual Steward Reasoning

The introduction of the FIA-26 interaction feature set is a substantial improvement over the original manuscript and directly addresses concerns regarding the use of features that were insufficiently connected to the fault-attribution process. However, the manuscript currently assumes that the engineered FIA-26 variables represent the criteria actually used by FIA stewards without providing evidence for this assumption. Features such as corner ownership, right-to-contest-corner metrics, leave-space obligations, aggression measures, and avoidance indicators are plausible proxies for steward reasoning, but their correspondence to real steward decision-making remains unvalidated. To strengthen the central argument, the authors should consider comparing the FIA-26 features against the content of actual steward reports. Written steward decisions frequently reference concepts such as significant overlap, entitlement to racing room, avoidable contact, late braking, failure to leave space, and loss of control. Demonstrating that the proposed features correlate with the occurrence or emphasis of such concepts in steward reports would provide important evidence

that the FIA-26 feature set captures the intended decision process rather than merely representing an alternative set of telemetry-derived variables.

If such validation is not feasible because steward reports do not contain sufficient detail, are not consistently available, or cannot be systematically analyzed, then this should be explicitly acknowledged as a limitation. In that case, the FIA-26 framework should be presented as a conceptual and theoretically motivated approximation of steward reasoning rather than a validated representation of the steward decision process itself. Clarifying this distinction would improve the scientific rigor of the manuscript and appropriately calibrate the strength of the conclusions that can be drawn from the feature-engineering framework.

2.Feature Importance, Interpretability, and Scientific Insight

The revised manuscript introduces 26 FIA-aligned interaction features and demonstrates that these features generalize somewhat better than the original 225 statistical telemetry features. However, the manuscript does not provide sufficient insight into which of the FIA-26 features actually contribute to model behavior. As currently written, readers are unable to determine whether the modest predictive signal arises from a small subset of meaningful interaction variables or from diffuse contributions across many engineered features.

The authors should provide feature-importance analyses and model interpretability results. For tree-based models, permutation importance or impurity-based importance measures would be informative, while SHAP analysis would be particularly valuable because it would reveal how specific FIA-aligned concepts influence model predictions. Such analyses would substantially increase the scientific value of the study, regardless of predictive performance, by identifying which aspects of driver interaction contain the strongest fault-attribution signal. Given that the manuscript is increasingly framed as an investigation into the limits of telemetry-based fault attribution, understanding which features matter is arguably more important than marginal changes in classification accuracy.

3.Reframe the Manuscript as a Feasibility Study and Emphasize the Negative Result

In my view, the revised manuscript is scientifically stronger than the original because it no longer presents a machine-learning success story but instead exposes an important limitation of the problem itself. The results support several conclusions that deserve greater emphasis throughout the Abstract, Discussion, and Conclusion.

First, the manuscript demonstrates that naïve telemetry features do not generalize across seasons. Second, FIA-aligned interaction features are more appropriate than generic statistical summaries because they better reflect the relational quantities considered during steward adjudication. Third, even after introducing domain-relevant features, predictive performance remains near chance under deployment-like evaluations such as leave-one-season-out validation. Taken together, these findings suggest that telemetry alone is insufficient for robust fault attribution.

I encourage the authors to reorganize the framing of the paper around this central result. The most important contribution may not be the development of a classifier but rather the demonstration that fault attribution appears difficult to predict from telemetry alone, even when rule-aligned features are used. This is a meaningful scientific finding because it provides guidance to future researchers regarding the limitations of telemetry-only approaches.

The Discussion should further explore the possibility that fault attribution depends on information fundamentally absent from telemetry data. FIA stewards do not rely solely on speed, braking, and positioning. They also consider video evidence, broader spatial context, driver intent, racing-line expectations, situational awareness, precedent from prior rulings, and subjective interpretation of regulations. Many of these factors are not available within the FastF1 telemetry channels used in this study. Consequently, the observed performance ceiling may reflect an information limitation rather than a machine-learning limitation.

I therefore recommend that the authors explicitly state that telemetry-based fault attribution appears fundamentally limited and that future systems will likely require richer multimodal inputs,

including video, track geometry, steward reports, contextual race information, and other sources beyond telemetry alone. Such a reframing would align the manuscript more closely with its actual findings and would significantly strengthen its scientific contribution.

Comment 1: Validation of FIA-Aligned Features Against Actual Steward Reasoning

“The introduction of the FIA-26 interaction feature set is a substantial improvement over the original manuscript and directly addresses concerns regarding the use of features that were insufficiently connected to the fault-attribution process. However, the manuscript currently assumes that the engineered FIA-26 variables represent the criteria actually used by FIA stewards without providing evidence for this assumption. Features such as corner ownership, right-to-contest-corner metrics, leave-space obligations, aggression measures, and avoidance indicators are plausible proxies for steward reasoning, but their correspondence to real steward decision-making remains unvalidated. To strengthen the central argument, the authors should consider comparing the FIA-26 features against the content of actual steward reports. Written steward decisions frequently reference concepts such as significant overlap, entitlement to racing room, avoidable contact, late braking, failure to leave space, and loss of control. Demonstrating that the proposed features correlate with the occurrence or emphasis of such concepts in steward reports would provide important evidence that the FIA-26 feature set captures the intended decision process rather than merely representing an alternative set of telemetry-derived variables. If such validation is not feasible because steward reports do not contain sufficient detail, are not consistently available, or cannot be systematically analyzed, then this should be explicitly acknowledged as a limitation. In that case, the FIA-26 framework should be presented as a conceptual and theoretically motivated approximation of steward reasoning rather than a validated representation of the steward decision process itself. Clarifying this distinction would improve the scientific rigor of the manuscript and appropriately calibrate the strength of the conclusions that can be drawn from the feature-engineering framework.”

Response:

We did not attempt a systematic content analysis against the steward decision PDFs because those reports vary widely in length and specificity and rarely quantify the concepts they invoke (e.g., “significant overlap,” “avoidable contact”), which would make a content analysis at our sample size (164 incidents) unreliable rather than informative. This is now stated explicitly.

Comment 2: Feature Importance, Interpretability, and Scientific Insight

“The revised manuscript introduces 26 FIA-aligned interaction features and demonstrates that these features generalize somewhat better than the original 225 statistical telemetry features. However, the manuscript does not provide sufficient insight into which of the FIA-26 features actually contribute to model behavior. As currently written, readers are unable to determine whether the modest predictive signal arises from a small subset of meaningful interaction variables or from diffuse contributions across many engineered features. The authors should provide feature-importance analyses and model interpretability results. For tree-based models, permutation importance or impurity-based importance measures would be informative, while SHAP analysis would be particularly valuable because it would reveal how specific FIA-aligned concepts influence model predictions. Such analyses would substantially increase the scientific value of the study, regardless of predictive performance, by identifying which aspects of driver interaction contain the strongest fault-attribution signal. Given that the manuscript is increasingly framed as an investigation into the limits of telemetry-based fault attribution, understanding which features matter is arguably more important than marginal changes in classification accuracy.”

Response:

We have added a dedicated feature-importance and interpretability analysis covering all 26 FIA-aligned features (our existing permutation-importance results covered only an earlier 8-feature subset suggested in the first review round, not the full FIA-26 set used in the current results). We computed 1 permutation importance, defined as the mean drop in held-out balanced accuracy across 30 repeats per fold, averaged over the same stratified 10-fold cross-validation used for Table 9, and

2 SHAP (TreeExplainer) values on the same Random Forest configuration (150 trees, max depth = 5).

Comment 3: Reframe the Manuscript as a Feasibility Study and Emphasize the Negative Result

“In my view, the revised manuscript is scientifically stronger than the original because it no longer presents a machine-learning success story but instead exposes an important limitation of the problem itself. The results support several conclusions that deserve greater emphasis throughout the Abstract, Discussion, and Conclusion. First, the manuscript demonstrates that naïve telemetry features do not generalize across seasons. Second, FIA-aligned interaction features are more appropriate than generic statistical summaries because they better reflect the relational quantities considered during steward adjudication. Third, even after introducing domain-relevant features, predictive performance remains near chance under deployment-like evaluations such as leave-one-season-out validation. Taken together, these findings suggest that telemetry alone is insufficient for robust fault attribution. I encourage the authors to reorganize the framing of the paper around this central result. The most important contribution may not be the development of a classifier but rather the demonstration that fault attribution appears difficult to predict from telemetry alone, even when rule-aligned features are used. This is a meaningful scientific finding because it provides guidance to future researchers regarding the limitations of telemetry-only approaches. The Discussion should further explore the possibility that fault attribution depends on information fundamentally absent from telemetry data. FIA stewards do not rely solely on speed, braking, and positioning. They also consider video evidence, broader spatial context, driver intent, racing-line expectations, situational awareness, precedent from prior rulings, and subjective interpretation of regulations. Many of these factors are not available within the FastF1 telemetry channels used in this study. Consequently, the observed performance ceiling may reflect an information limitation rather than a machine-learning limitation. I therefore recommend that the authors explicitly state that telemetry-based fault attribution appears fundamentally limited and that future systems will likely require richer multimodal inputs, including video, track geometry, steward reports, contextual race information, and other sources beyond telemetry alone. Such a reframing would align the manuscript more closely with its actual findings and would significantly strengthen its scientific contribution.”

Response:

We have revised the Abstract, Discussion, and Conclusion to foreground the negative result as the paper's central contribution, rather than the classifier itself.

Thank you for addressing my comments. Accepted.