

give me 25 examples of numerical features they should have used but did not - and were available in telemetry

You're right to focus on this—the **weakness isn't just "not enough data," it's that the features don't encode the actual *interaction* between the cars**. Most of their 225 features are generic summaries (means, std, etc.), not *racing-relevant physics or geometry*.

Below are **25 concrete, telemetry-derived features** they could have constructed (using the same underlying data they already had), but didn't.

Relative position & geometry (these are critical—and missing)

These capture *who was where*, which is central to fault.

1. **Lateral offset at collision**
→ how far apart cars are across the track (inside vs outside)
2. **Longitudinal overlap (%) at corner entry**
→ how much one car is alongside the other
3. **Relative heading angle difference**
→ are cars aligned or converging?
4. **Distance at first overlap (not just minimum distance)**
→ when interaction begins
5. **Time spent side-by-side before collision**
6. **Relative position to apex (who is ahead at apex)**

7. Track position classification

→ inside line / outside line / center

8. Closing angle (rate of directional convergence)

 Speed & braking interaction (not just individual stats)

They used average speed—but not *comparative dynamics*.

9. Relative speed at braking onset**10. Difference in braking onset timing (Δt brake start)****11. Peak deceleration difference****12. Time-to-collision (TTC) just before impact****13. Speed differential at moment of turn-in****14. Rate of closing speed (d/dt of relative velocity)**

 Reaction & avoidance behavior

This is key for “avoidable vs unavoidable.”

15. Reaction time to opponent's braking/turning**16. Late evasive steering magnitude****17. Throttle lift timing relative to opponent****18. Brake application AFTER opponent turns in**

→ indicates reactive vs proactive behavior



Cornering & trajectory features

They use a crude “cornering” flag—this is far richer.

19. **Curvature of trajectory (radius of turn)**
 20. **Deviation from typical racing line**
(can be approximated from lap median trajectory)
 21. **Steering change rate (jerk in direction)**
 22. **Corner entry speed relative to typical lap speed**
-



Control inputs & aggression proxies

More behaviorally meaningful than raw averages.

23. **Throttle-brake overlap (aggression indicator)**
 24. **Throttle persistence into corner entry**
(how long throttle is maintained before braking)
 25. **Variance in control inputs (erratic vs smooth driving)**
-



Why these matter (the core issue)

The current paper mostly captures:

“What was each car doing individually?”

But fault depends on:

“How were the two cars interacting *with each other in space and time*?”

That missing layer is why:

- PCA doesn't hurt performance (features are redundant)
 - temporal models don't help
 - generalization collapses
-

Deeper critique (important)

Even more fundamentally:

The authors:

- compute **225 features**
- but almost none encode:
 - **right-of-way**
 - **overlap rules**
 - **racing line entitlement**

These are exactly what stewards use.

So the model is trying to learn a **rule-based judgment system** without giving it **rule-relevant variables**.

One-line takeaway

They engineered *a lot* of features—but not the ones that actually describe **who had the right to the corner and who violated it**.

If you want, I can go one step further and:

- map these features directly to FIA fault rules, or
- show which 5–10 features would likely give the biggest performance gain.

map to FIA fault rules

Good—this is exactly the missing layer in the paper.

Right now, their features describe **what each car did**.

FIA decisions are based on **who had the right to do it**.

Below is a clean mapping from **FIA-style fault principles** → **concrete telemetry features** (many of which they could have computed, but didn't).

1. “Significant overlap” → Who has the right to space?

FIA idea (simplified)

If a car is sufficiently alongside at corner entry, it is entitled to racing room.

What telemetry should capture

- Longitudinal overlap % at turn-in

- **Relative front axle position at corner entry**
- **Time of first overlap (early vs last-second dive)**
- **Overlap duration before apex**

Why it matters

This is often the **single most important rule** in deciding fault.

👉 Missing this means the model cannot distinguish:

- legitimate side-by-side racing
vs
- late divebombs

2. “Leave space” obligation → Did a driver squeeze the other?

FIA idea

If another car is alongside, you must leave at least a car’s width.

Telemetry features

- **Lateral distance between cars at apex**
- **Minimum lateral clearance vs car width**
- **Rate of lateral closing (was space removed?)**
- **Track position (inside/outside line)**

Why it matters

Most “forcing off track” penalties come from this.

👉 The paper has distance—but not **track-relative lateral positioning**, which is critical.

3. “Corner ownership” → Who was ahead at the apex?

FIA idea

The car ahead at the apex typically has priority.

Telemetry features

- Relative longitudinal position at apex
- Time difference to apex
- Speed-adjusted position (projected position at apex)

Why it matters

This determines who “owns” the corner.

👉 Without apex-relative positioning, the model cannot infer priority.

4. “Avoidable contact” → Did the driver react properly?

FIA idea

Drivers must make reasonable efforts to avoid contact.

Telemetry features

- Reaction time to opponent braking/turn-in
- Brake application after opponent commits

- Throttle lift timing
- Evasive steering magnitude

Why it matters

Even if a driver has partial right, failure to react can still assign fault.

👉 Current features don't capture **reaction vs intention**.

⚡ 5. "Late braking / divebombing" → Was the move unrealistic?

FIA idea

Overtaking attempts must be controlled and achievable.

Telemetry features

- Braking onset difference (Δt brake start)
- Deceleration relative to typical corner entry
- Time-to-collision at turn-in
- Closing speed at braking zone

Why it matters

This distinguishes:

- controlled overtakes
- vs
- reckless lunges
-



6. “Maintaining control” → Was the car under control?

FIA idea

A driver must remain in control of the car.

Telemetry features

- Steering instability / oscillation
- Excessive speed vs corner baseline
- Trajectory deviation from typical line
- Sudden yaw/heading change

Why it matters

Loss of control often leads to fault assignment.



7. “Predictability” → Did the driver make a sudden move?

FIA idea

Drivers must avoid erratic or unpredictable behavior.

Telemetry features

- Sudden lateral movement (jerk)
- Change in steering rate
- Late line deviation
- Throttle/brake inconsistency

Why it matters

Unpredictable moves often shift blame.

8. “Right to racing line” → Was the driver entitled to the line?

FIA idea

The leading car can take the optimal racing line unless challenged.

Telemetry features

- Distance from canonical racing line
- Whether deviation occurred due to opponent presence
- Relative position BEFORE corner entry

Why it matters

Determines whether a move was legitimate positioning or blocking.

9. “Sequential responsibility” → Who initiated the conflict?

FIA idea

Fault often depends on who created the situation.

Telemetry features

- Time of first trajectory convergence
- Who deviated first from racing line
- Initiator of closing distance

Why it matters

Causality—not just the final contact—matters.



10. “Sufficient time to react” → Was the move too late?

FIA idea

Drivers must leave enough time/space for others to respond.

Telemetry features

- Time gap between maneuver and collision
- Relative velocity at maneuver start
- Available reaction window

Why it matters

Late moves → higher likelihood of fault.



What this reveals about the manuscript

The authors *did* use:

- speed
- braking
- throttle
- distance

But FIA decisions depend on:

relative positioning + timing + entitlement rules

👉 These are fundamentally **relational and rule-based**, not just statistical summaries.

The core mismatch

What the paper models:

“What patterns in telemetry correlate with penalties?”

What FIA actually does:

“Apply structured rules about overlap, space, and control.”

Bottom-line critique

The authors tried to solve a **rule-based decision problem** using **feature summaries that do not encode the rules**.

That's why:

- performance looks decent in easy settings
 - but collapses in realistic ones
-

One-line takeaway

Without features that explicitly represent **overlap, space, and right-of-way**, the model is fundamentally blind to the logic stewards actually use.

If you want, I can turn this into:

- a figure/table for a manuscript revision, or
- a minimal 8-feature “FIA-aware feature set” that would likely outperform their current pipeline.

both

Great—here are both pieces you asked for:

1. a **clean, publication-ready table/figure mapping telemetry → FIA rules**, and
2. a **minimal, high-impact feature set (8 features)** that would likely outperform their current approach.

 (A) Figure/Table: Telemetry Features Mapped to FIA Fault Rules

You can drop this directly into a manuscript as a conceptual framework.

Table X. Mapping FIA fault attribution principles to telemetry-derived features

FIA Principle	What stewards judge (plain language)	Telemetry-derived features (missing in manuscript)	Why it matters
Significant overlap	Was the overtaking car far enough alongside to deserve space?	Longitudinal overlap % at turn-in; front axle alignment; overlap duration	Determines right to contest corner
Leave space	Did one driver squeeze the other off track?	Lateral distance at apex; minimum clearance vs car width; lateral closing rate	Core criterion for “forcing off track”
Corner ownership	Who was ahead at the apex?	Relative longitudinal position at apex; time-to-apex difference	Establishes priority through corner
Avoidable contact	Did the driver try to avoid the crash?	Reaction time to opponent; brake/throttle response timing; evasive steering	Distinguishes unavoidable vs negligent contact
Late braking / divebomb	Was the move realistic or reckless?	Δ braking onset; closing speed at entry; time-to-collision at turn-in	Identifies overly aggressive overtakes
Maintaining control	Was the car under control?	Trajectory curvature; deviation from typical line; speed vs corner baseline	Loss of control → likely fault
Predictability	Did the driver behave erratically?	Steering jerk; sudden lateral shifts; control instability	Erratic moves often penalized

FIA Principle	What stewards judge (plain language)	Telemetry-derived features (missing in manuscript)	Why it matters
Right to racing line	Was the driver entitled to the line?	Distance from canonical racing line; pre-corner positioning	Distinguishes legal positioning vs blocking
Initiation of conflict	Who created the situation?	Time of first convergence; initiator of trajectory deviation	Establishes causality
Time to react	Was there enough time for the other driver to respond?	Time gap between maneuver and impact; relative velocity at maneuver	Late moves → higher fault probability

 How to describe this in the paper (1–2 sentences)

“FIA fault attribution is governed by relational and rule-based criteria such as overlap, entitlement to space, and avoidability. These criteria can be approximated using telemetry-derived interaction features (Table X), which are largely absent from the current feature set.”

 (B) Minimal “FIA-aware” Feature Set (8 features)

This is the **high-value subset**—if you only fixed one thing in the paper, this would be it.

These features directly encode the *decision logic* stewards use.

 The 8 features

1. Overlap ratio at turn-in

- % of trailing car alongside leading car at corner entry
- (e.g., front axle relative position)

👉 Captures **right to space**

2. Relative position at apex

- Who is ahead at apex (binary or continuous)

👉 Captures **corner ownership**

3. Lateral clearance at apex

- Distance between cars normalized by car width

👉 Captures **leave-space obligation**

4. Δ braking onset (timing difference)

- When each driver begins braking relative to corner entry

👉 Captures **late braking / divebomb behavior**

5. Time-to-collision at turn-in

- Predicted collision time based on relative velocity

👉 Captures **feasibility of maneuver**

6. Reaction delay to opponent maneuver

- Time between opponent's turn/brake and driver's response

👉 Captures **avoidability**

7. Trajectory deviation from racing line

- Difference from typical (median) line through that corner

👉 Captures **control + entitlement to line**

8. Initiator flag (who deviated first)

- Which driver first changed trajectory toward the other

👉 Captures **causality**

🧠 Why these 8 beat 225 generic features

The current paper uses:

- means, variances, slopes (generic summaries)

These 8 features instead encode:

- **who had the right**
- **who violated it**
- **who caused the interaction**

👉 In other words:

They compress the problem into **rule-aligned variables**, not raw statistics.



Expected impact (realistically)

If implemented correctly, you would expect:

- **Stratified CV**: similar or slightly better ($\approx 75\text{--}80\%$)
- **LOSO (real-world)**: meaningful improvement (e.g., $\sim 55\% \rightarrow \sim 60\text{--}65\%$ BAcc)

Not perfect—but **much closer to actual steward reasoning**.



Key insight you can include in review

“The current feature set is high-dimensional but not rule-aligned. A small number of interaction-based features grounded in FIA decision criteria would likely outperform the existing 225-feature statistical representation, particularly under distribution shift.”



Final takeaway

The paper’s main weakness is not the model—it’s that the features do not encode the **rules of racing**, only the **statistics of motion**.

If you want, I can:

- turn this into a **revision section the authors can directly insert**, or
- sketch how to compute each feature step-by-step from FastF1 telemetry.

how to compute each feature

Below is a **practical, telemetry-only recipe** for computing each of the 8 “FIA-aware” features from the kind of data the paper already uses (time, X–Y position, speed, throttle, brake, gear). I’ll assume you have **two synchronized time series** for car A and car B, resampled to a common timeline (as the paper already does).

I’ll keep it implementation-oriented (you can drop this into Python fairly directly).

Preliminaries (you need these first)

0. Synchronize and smooth

- Resample both cars to same timestamps (they already use 600 points)
- Optionally apply light smoothing (Savitzky–Golay or rolling mean) to reduce noise

1. Compute basic geometry (core building blocks)

For each timestep t :

- Position:
 - $(x_A, y_A), (x_B, y_B)$
- Velocity vectors:

- $v_A = (dx_A/dt, dy_A/dt)$
 - $v_B = (dx_B/dt, dy_B/dt)$
 - Relative vector:
 - $r = (x_B - x_A, y_B - y_A)$
-

2. Define track-aligned coordinates (important)

You need to approximate:

- **longitudinal direction** (along track)
- **lateral direction** (across track)

Simple method (no map required):

At each timestep:

- Use **average velocity direction** as track direction:
 - $\hat{t} = \frac{v_A + v_B}{\|v_A + v_B\|}$
- Define lateral:
 - $\hat{n} = (-t_y, t_x)$

Then:

- Longitudinal separation:

$$\Delta s = r \cdot \hat{t}$$
 - Lateral separation:

$$\Delta d = r \cdot \hat{n}$$
-



Now compute the 8 features

1. Overlap ratio at turn-in

Step 1: detect turn-in

Turn-in $\approx$ when steering begins $\rightarrow$ approximate with:

- peak in curvature OR
- start of sustained deceleration

Simple proxy:

- first time where:
 - speed decreasing AND
 - lateral acceleration increasing
-

Step 2: compute overlap

Let:

- L = car length (~5 m)

Then:

$$\text{overlap} = \frac{L - |\Delta s|}{L}$$

Clamp between 0 and 1.

Interpretation:

- 0 $\rightarrow$ no overlap
- 1 $\rightarrow$ fully alongside

2. Relative position at apex

Step 1: detect apex

Apex $\approx$ maximum curvature point:

- compute curvature:

$$\kappa = \frac{|x'y'' - y'x''|}{(x'^2 + y'^2)^{3/2}}$$

- apex = time of **max curvature**

Step 2: compare longitudinal position

$$\text{apex_lead} = \begin{cases} 1 & \text{if } \Delta s > 0 \text{ (B ahead)} \\ 0 & \text{if } \Delta s < 0 \text{ (A ahead)} \end{cases}$$

You can also keep it continuous (Δs value).

3. Lateral clearance at apex

At apex time:

$$\text{clearance} = |\Delta d|$$

Normalize:

$$\text{clearance_norm} = \frac{|\Delta d|}{W}$$

where $W \approx 2$ m (car width)

Interpretation:

- $< 1 \rightarrow$ overlap (danger zone)
- $1 \rightarrow$ safe spacing

4. Δ braking onset

Step 1: detect braking start for each car

Braking onset = first time where:

- brake > threshold (e.g., 0.1)
- OR strong negative acceleration

Step 2: compute difference

$$\Delta t_{\text{brake}} = t_B - t_A$$

Interpretation:

- negative $\rightarrow$ B brakes later (more aggressive)

5. Time-to-collision (TTC) at turn-in

At turn-in:

- Relative velocity:

$$v_{\text{rel}} = v_B - v_A$$

- Project onto separation vector:

$$\text{closing speed} = -\frac{r \cdot v_{\text{rel}}}{\|r\|}$$

Then:

$$\text{TTC} = \frac{\|r\|}{\text{closing speed}}$$

Clamp:

- if closing speed $\leq 0 \rightarrow \text{TTC} = \infty$
-

6. Reaction delay

Step 1: detect opponent action

- e.g., opponent braking or turning:
 - brake > threshold OR curvature increase

Step 2: detect subject response

- first time subject:
 - brakes OR changes direction
-

Step 3: compute delay

$$\text{reaction delay} = t_{\text{response}} - t_{\text{opponent action}}$$

7. Trajectory deviation from racing line

Step 1: estimate “reference line”

For that corner:

- average trajectory across laps (or use median path)
-

Step 2: compute deviation

At each time:

$$\text{deviation}(t) = \text{distance to reference line}$$

Feature:

- mean deviation over window OR
 - max deviation near apex
-

8. Initiator flag (who moved first)

Step 1: detect lateral movement

Compute:

$$\frac{d(\Delta d)}{dt}$$

Step 2: find first significant change

- threshold on lateral velocity

Step 3: assign initiator

- if A changes first → A initiator
- else B

Binary feature:

- +1 = A initiated
 - -1 = B initiated
-

Implementation notes (important)

1. Robustness matters more than precision

- You don't need perfect physics
 - You need **consistent approximations**
-

2. Smoothing is critical

- Raw derivatives are noisy
 - Always smooth before computing curvature, velocity
-

3. Use relative features

Instead of absolute:

- speed → use speed difference
- position → use relative position

This improves generalization.

Minimal pseudo-code structure

```
for incident in dataset:
    A, B = load_telemetry()

    sync(A, B)
    smooth(A, B)

    compute_velocity(A, B)
    compute_relative_geometry(A, B)

    turn_in = detect_turn_in()
    apex = detect_apex()

    features = {
        "overlap": compute_overlap(turn_in),
        "apex_position": compute_apex_position(apex),
        "clearance": compute_clearance(apex),
        "brake_delta": compute_brake_delta(),
        "ttc": compute_ttc(turn_in),
        "reaction_delay": compute_reaction(),
        "line_deviation": compute_deviation(),
        "initiator": compute_initiator()
    }
```

Final takeaway

All of these features are computable from the **same telemetry the paper already uses**—but they require transforming raw signals into **interaction-aware, rule-relevant quantities**.

That's the key leap the manuscript doesn't make.

If you want, I can:

- give you **actual Python code using FastF1**, or
- show which of these features are most statistically stable given small sample sizes.