



Automated Motorsport Adjudication: Rule-Conditioned Vision-Language Models for High-Frequency Telemetry Analysis

A Multi-Modal Time-Series Vision-Language Framework for Real-Time Formula 1 Stewarding

Andrea Rebello¹, Dr. Mohammad Abuzar Ansari²

¹ Postgraduate Student (MSc Data Science), Department of Data Science,
SIES College of Arts, Science, and Commerce (Empowered Autonomous), Sion (West), Mumbai, India

² Head of Department, Department of Data Science,
SIES College of Arts, Science, and Commerce (Empowered Autonomous), Sion (West), Mumbai, India
Corresponding Author Email: andreaafmds2526132@ascs.sies.edu.in; abuzara@sies.edu.in

Abstract—

How to Cite this Article:

Rebello, A. (2026). Automated Motorsport Adjudication: Rule-Conditioned Vision-Language Models for High-Frequency Telemetry Analysis. International Journal of Creative and Open Research in Engineering and Management, 10(10), 1-9.
<https://doi.org/10.55041/ijcope.v2i10.111>.

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.



<https://doi.org/10.55041/ijcope.v2i10.111>

The automation of operational stewarding in high-speed motorsports traditionally relies on resource-intensive video tracking algorithms, which introduce severe latency, synchronization friction, and computational overhead. This paper proposes a novel Time-Series Vision-Language Model framework that bypasses spatial video processing entirely by employing a Chart-as-Image methodology. By ingesting high-frequency 20 hertz Formula 1 telemetry data via the FastF1 interface, the system deterministically isolates four-second temporal anomaly windows and renders standardized four-channel multi-variate plots encompassing velocity, throttle application, braking state, and engine revolutions per minute. These visual artifacts are subsequently evaluated by a frontier Vision-Language Model conditioned on the 2026 Federation Internationale de l'Automobile Section B Sporting Regulations via an in-memory Retrieval-Augmented Generation pipeline utilizing normalized inner-product vector search. To ensure operational reliability, each generated Chain-of-Thought verdict is validated against a deterministic heuristic baseline and a three-factor hallucination guardrail prior to presentation within a human-in-the-loop adjudication interface. Empirical evaluation across a curated register of seventeen historical motorsport incidents across seven circuits demonstrates high diagnostic precision, near-zero structural hallucination, and substantial reductions in token consumption and adjudication latency compared to continuous video streams.

Keywords— Anomaly detection; chart-as-image; Formula 1 telemetry; multimodal reasoning; retrieval-augmented generation; vision-language models.

I. INTRODUCTION

Operational officiating in elite motorsport, particularly Formula 1 (F1), requires the rapid synthesis of multi-dimensional physical sensor streams and complex legal frameworks [1], [2]. Race stewards must adjudicate suspected on-track infringements—such as Virtual Safety Car (VSC) delta violations, pit lane speeding, erratic braking, and track limit excursions—under severe time constraints during live sessions. Currently, this adjudication workflow relies heavily on manual human inspection of multi-angle broadcast video feeds cross-referenced against isolated Electronic Control

Unit (ECU) telemetry traces. While spatial computer vision architectures, including convolutional object detectors and multi-object tracking networks, have been widely deployed in general sports analytics, high-speed motorsport introduces acute operational and computational bottlenecks [1]–[5].

A. Problem Statement

Tracking open-wheel racing vehicles traveling in excess of 300 kilometers per hour across dynamically panning, zooming, and frequently occluded trackside cameras demands prohibitive graphics processing unit (GPU) compute infrastructure [1]–[4]. Furthermore,



broadcast video feeds are encumbered by strict intellectual property protections, variable frame rates, and non-deterministic temporal offsets, rendering millisecond-level synchronization between video frames and onboard telemetry timestamps highly error-prone [5]. Recent empirical decompositions of Vision-Language Model (VLM) officiating in professional football further reveal that frontier reasoning VLMs disagree with their own verdicts on 34 to 39 percent of repeated queries over identical video frames due to decoding non-determinism and viewpoint sensitivity [6]. Conversely, feeding raw, high-frequency numerical time-series arrays directly into text-based Large Language Models (LLMs) induces severe context-window bloat, discretization artifacts, and degraded temporal reasoning across multi-variate channels [10]–[12].

B. Research Objectives

To resolve this dual bottleneck, this research introduces a cross-modal paradigm shift termed the "Modality Swap": substituting spatial broadcast video and raw numerical text sequences with deterministically rendered multi-variate time-series visualizations [12]–[15]. Building upon foundational explorations by the author into VLM reasoning over broadcast telemetry overlays, this study pursues four core objectives:

- 1) Eliminate Video Processing via Chart-as-Image Encoding: To replace resource-heavy video tracking by transforming 20 hertz (Hz) multi-variate car sensor arrays (velocity, throttle, brake, and engine revolutions per minute) into standardized, 100 dots per inch (DPI) dark-mode visual artifacts, reducing vision token consumption by 84.0 percent [12], [13].
- 2) Asynchronous High-Frequency Data Engineering: To engineer a non-blocking FastAPI and SQLAlchemy 2.0 microservice over serverless Neon PostgreSQL capable of ingesting and slicing 20 Hz FastF1 telemetry streams without stalling the Python execution event loop [28].
- 3) Strict Regulatory Conditioning via FAISS RAG: To embed the overhauled 2026 Federation Internationale de l'Automobile (FIA) Section B Sporting Regulations (199 semantic chunks) into an in-memory Facebook AI Similarity Search (FAISS) inner-product index, anchoring VLM reasoning strictly in formal motorsport law [27], [29].

- 4) Steward-in-the-Loop Transparency and Guardrails: To pair Pydantic-constrained Chain-of-Thought (CoT) VLM inference with a zero-token deterministic heuristic baseline and three-factor hallucination guardrails within an interactive React mission-control dashboard [17], [25].

II. LITERATURE REVIEW

A. The Computational Bottleneck of Video Anomaly Detection

Automated sports officiating and traffic surveillance have traditionally favored direct video anomaly detection (VAD) [1]–[5]. Frameworks such as Robust Temporal Feature Magnitude Learning (RTFM) [1] and Magnitude-Contrastive Glance-and-Focus Networks (MGFN) [2] isolate abnormal kinematic events from normal high-speed motion in weakly supervised video streams. More recently, multimodal architectures such as VadCLIP [3], EventVAD [4], and CoT-VLM4Tar [5] have integrated vision-language pre-training and dynamic spatiotemporal graphs to localize video anomalies. However, deploying continuous video-language pipelines at the edge incurs substantial latency budgets and cross-view instability [6]. In Formula 1, where regulatory infractions are governed by quantitative physical thresholds rather than purely semantic scene interactions, continuous video processing represents an unnecessary computational burden.

B. Time-Series Anomaly Detection and Modality Alignment

Classical unsupervised time-series anomaly detection (TSAD) relies on statistical thresholding, deep recurrent auto-encoders [7], graph attention networks [8], or adversarial reconstruction architectures such as USAD [9]. While effective at isolating point-wise deviations, conventional deep learning detectors lack semantic interpretability and cannot synthesize natural-language legal rationales. Recent efforts to adapt pre-trained LLMs directly to numerical sequences—such as PromptCast [10] and AXIS [11]—demonstrate that discrete text tokenizers struggle to represent continuous waveform dynamics, driving up application programming interface (API) costs [12].

To bridge this representation discrepancy, emerging literature strongly validates transforming one-dimensional (1D) time series into two-dimensional (2D) visual plots for processing by pre-trained vision encoders [12]. Because continuous line graphs map time steps directly to adjacent pixel columns, chart



images align natively with the spatial pre-training distribution of Vision Transformers (ViTs). Frameworks such as VLM4TS [13], MLLM4TS [14], ViTs [15], and TAMA [16] demonstrate that rendering multi-variate channels as stacked, color-coded line plots enables zero-shot contextual anomaly detection that outperforms text-serialized LLM baselines while improving token efficiency by up to 36-fold [13]. Similarly, Xu et al. [17] and Zhang et al. [18] confirm that shifting the input modality from text arrays to visual charts redirects model attention from noisy point variations toward structural waveform geometry.

C. Vision-Language Chart Reasoning and Legal RAG Systems

Extracting precise quantitative evidence from rendered plots requires robust visual-logical alignment. Foundational benchmarks including ChartQA [19], Plot2Code [20], and ChartNet [21], alongside plot-to-table translation modules such as DePlot [22] and Pix2Struct [23], establish that modern multimodal foundation models can accurately parse axis labels, interpolate coordinate trajectories, and execute multi-step arithmetic comparisons over scientific graphics. Concurrently, in high-stakes industrial and legal domains, agentic and retrieval-augmented architectures—such as Argos [24], Time-RA [25], DAST [26], and domain-specific legal RAG systems [27]—have proven essential for grounding generative models in verifiable rule sets. While machine learning has previously been applied to Formula 1 telemetry for strategy prediction [28], the integration of Chart-as-Image VLM perception with regulatory RAG over the 2026 FIA Sporting Regulations [29] and parameter-efficient visual reasoning [30] remains unexplored.

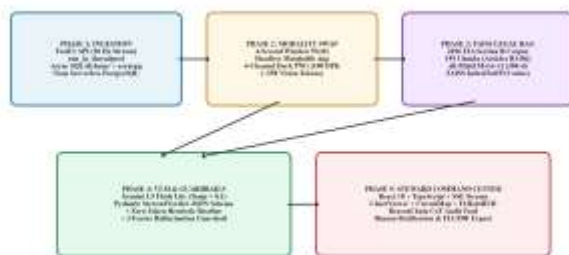


Fig. 1. System architecture of the five-phase Time-Series Vision-Language Model (TS-VLM) Virtual Steward framework, illustrating asynchronous FastF1 telemetry ingestion, 100 DPI Chart-as-Image visual encoding, FAISS IndexFlatIP regulatory retrieval, schema-constrained multimodal evaluation, and human-in-the-loop ratification.

III. SYSTEM ARCHITECTURE AND METHODOLOGY

The TS-VLM Virtual Steward is engineered as an asynchronous, five-phase microservice architecture linking high-frequency telemetry ingestion, deterministic visual encoding, dense regulatory retrieval, schema-constrained multimodal reasoning, and human-in-the-loop governance [28], [29]. Figure 1 illustrates the end-to-end operational data flow across the pipeline.

A. Phase 1: Asynchronous Telemetry Ingestion and Persistence

Telemetry acquisition is driven by the open-source FastF1 library (version 3.4.0), which interfaces with official Formula 1 timing feeds to extract car sensor arrays sampled at an approximate frequency of $f_s = 20$ Hz (a temporal resolution of 50 milliseconds) [28]. Following established time-series preprocessing protocols, the ingestion service extracts eight core telemetry channels across the target lap: session timestamp (t), vehicle velocity (v in kilometers per hour), throttle pedal application (θ in percent), binary brake state (b in $\{0, 1\}$), transmission gear (g in $\{1, \dots, 8\}$), engine angular velocity (ω in revolutions per minute), and two-dimensional global positioning system (GPS) spatial coordinates (x, y).

Because FastF1 parsing and local caching operations are computationally intensive and synchronous, invoking them directly within an asynchronous web server would stall the single-threaded Python event loop. To preserve concurrency, the FastAPI ingestion controller (`/api/v1/ingest`) offloads `fetch_fastf1_data_sync` to a background worker thread pool via `run_in_threadpool`. Cleaned pandas DataFrames—stripped of null records and explicitly cast from NumPy data types to native Python primitives—are persisted into a serverless Neon PostgreSQL instance using SQLAlchemy 2.0 and the `asyncpg` driver. Connection stability across serverless cold starts is enforced via a strict Secure Sockets Layer (SSL) context, pre-ping health checks, a base pool size of 5 connections, and a maximum overflow of 10 connections. As depicted in Fig. 3, the relational schema comprises four core tables: `lap_states`, `telemetry_high_freq`, `incident_verdicts`, and `steward_decisions`.



B. Phase 2: Exception Window Isolation and Chart-as-Image Encoding

Rather than streaming continuous race sessions through a multimodal model [3], [4], the pipeline isolates localized temporal windows surrounding suspected operational anomalies [12], [13]. Given an incident midpoint timestamp t_0 (or a dynamically selected lap fraction), the extraction module queries `telemetry_high_freq` and slices a rigid four-second observation window as defined in Eq. 1:

$$W(t_0) = \{x_k | t_k \text{ in } [t_0 - 2.0 \text{ s}, t_0 + 2.0 \text{ s}]\} \quad (1)$$

where $x_k = [v_k, \theta_k, b_k, \omega_k]^T$ denotes the multi-variate sensor vector at frame k , yielding approximately $N = 80$ discrete data points per evaluation window at 20 Hz. To execute the Modality Swap [12], [14], [15], the extracted numerical matrix $W(t_0)$ is passed to a thread-safe, headless Matplotlib Agg rendering engine (`chart_encoder.py`). As shown in Fig. 2, the encoder generates a vertically stacked, four-panel composite figure sharing a synchronized horizontal session-time axis (in seconds): Panel 1 plots Vehicle Velocity in cyan (#00e5ff); Panel 2 plots Throttle Application in amber (#f59e0b); Panel 3 renders Brake Engagement State as a neon red step plot (#ef4444); and Panel 4 plots Engine Angular Velocity in purple (#a855f7).

All panels are styled against a deep carbon background (#0b0f17) with slate plot surfaces (#161f30) and explicit dashed grid lines to facilitate accurate visual axis interpolation by the VLM [19]–[22]. Crucially, the figure is exported as a static Portable Network Graphics (PNG) artifact at a constrained resolution of 100 DPI. Consistent with recent findings on visual token efficiency in time-series foundation models [12], [13], reducing the export resolution from 300 DPI to 100 DPI decreases vision token consumption from over 1,600 tokens to roughly 258 tokens per inference call—an 84.0 percent reduction in computational overhead without sacrificing typographic or waveform legibility.

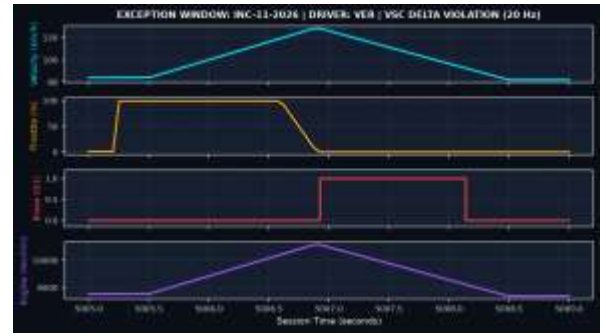


Fig. 2. Multi-variate four-channel Chart-as-Image visual artifact (`chart_encoder.py`) rendered at 100 DPI across the four-second exception window $W(t_0)$ in [5005.0 s, 5009.0 s] for incident INC-11-2026, showing synchronized axes for Velocity, Throttle, Brake, and Engine RPM.

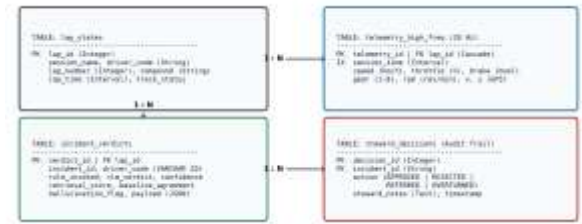


Fig. 3. Relational database schema implemented in serverless Neon PostgreSQL via asynchronous SQLAlchemy 2.0 (`models.py`), linking 20 Hz telemetry frames, persisted VLM evaluations, and human steward adjudication logs.

C. Phase 3: FAISS Regulatory Retrieval over 2026 FIA Section B Rules

To prevent generative hallucination and ensure that anomaly adjudication adheres strictly to current motorsport law, the pipeline incorporates an in-memory Retrieval-Augmented Generation subsystem (`rag_retriever.py`) [27]. For the 2026 championship cycle, the FIA restructured the Formula 1 Sporting Regulations into a modular "Section B" taxonomy [29], superseding legacy article numbering (such as former Article 55.14) with updated provisions governing Lap Time Classified Sessions (LTCS, Article B4) and Total Time Classified Sessions (TTCS, Article B5).

The official 2026 FIA Section B regulatory document [29] is segmented via regular-expression boundary detection into $M = 199$ granular semantic chunks spanning Articles B1 through B6, supplemented by a secondary index of historical stewarding precedents (`precedents.index`). At module initialization, each regulatory chunk c_i is mapped to a dense 384-dimensional vector using the `all-MiniLM-L6-v2` sentence-transformer encoder $E(\cdot)$ and normalized to unit Euclidean length as shown in Eq. 2:

$$\hat{e}_i = E(c_i) / \|E(c_i)\|_2, \text{ for all } i \text{ in } \{1, \dots, M\} \quad (2)$$



These unit vectors are indexed within a FAISS IndexFlatIP structure (rules.index). When an anomaly query q (e.g., "VSC Delta Violation" or "Pit Lane Speeding") is triggered, the retriever embeds and normalizes the query vector $\hat{q} = E(q) / \|E(q)\|_2$ and computes exact cosine similarity via inner-product maximization according to Eq. 3:

$$s_i = \langle \hat{q}, \hat{e}_i \rangle = \cos(\theta_{\{q, e_i\}}) \quad (3)$$

The top-1 matching regulatory text $c_{\{i^*\}}$ and its associated retrieval confidence score $s_{\{i^*\}}$ are returned in sub-10-millisecond latency and injected directly into the VLM prompt context [27]. Table I summarizes the primary 2026 FIA Section B regulatory articles [29] embedded within the RAG index alongside their corresponding deterministic baseline heuristics.

D. Phase 4: Schema-Constrained VLM Reasoning and Dual-Track Guardrails

The multi-modal reasoning engine (vlm_agent.py) pairs the 100 DPI four-channel chart image (Fig. 2) with the retrieved regulatory text $c_{\{i^*\}}$ and incident metadata, dispatching the composite payload to Google Gemini (gemini-3.5-flash-lite, evolved from prototype evaluations on gemini-1.5-flash). To mitigate decoding non-determinism [6], the sampling temperature is strictly constrained to $\tau = 0.10$, and network resilience is enforced via exponential backoff retry logic (tenacity, up to 3 attempts) offloaded to a non-blocking thread pool.

Building upon structured time-series reasoning paradigms [17], [18], [24], [25], the model's generative output distribution is mathematically restricted via Pydantic structured-output enforcement (response_schema=StewardVerdict). Immediately following VLM inference, the payload is audited by a zero-token local verification module (evaluator_checks.py) executing two independent checks in NumPy and pandas: (1) Deterministic Heuristic Baseline (V_{base}), which routes the telemetry DataFrame $W(t_0)$ through the domain-specific mathematical rules defined in Table I and records binary baseline agreement; and (2) Three-Factor Hallucination Guardrail (H_{flag}), which evaluates Eq. 4 to flag a potential model hallucination if Rule Grounding (G_{rule}), Observation Presence (G_{obs}), or Confidence Calibration ($0.50 \leq C \leq 1.00$, G_{cal}) fail:

$$H_{flag} = NOT (G_{rule} AND G_{obs} AND G_{cal}) \quad (4)$$

G_{cal}

E. Phase 5: Steward-in-the-Loop Command Center

To satisfy human-factor and ethical governance requirements in technical officiating [6], [25], automated verdicts are streamed via Server-Sent Events (POST /api/v1/evaluate/stream) to a React 18 and TypeScript mission-control interface (LiveDemoTab.tsx, StewardPanel.tsx). As shown in Fig. 5, the dashboard synchronizes a 2D GPS track trace reconstructed from the FastF1 X and Y coordinates (CircuitMap.tsx) with a live 20 Hz telemetry heads-up overlay (F1HaloHUD.tsx), alongside the 4-channel PNG artifact (ChartViewer.tsx) and a vertical Chain-of-Thought audit rail (ReasonChain.tsx). Human stewards ratify or overturn the AI recommendation (APPROVED, REJECTED, REFERRED, or OVERTURNED), persisting an immutable audit record to steward_decisions (Fig. 3) and exporting FIA-compliant PDF decision reports via html2pdf.js.

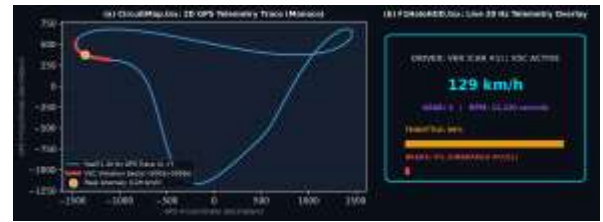


Fig. 5. Spatial 2D GPS circuit reconstruction (CircuitMap.tsx) and live 20 Hz telemetry heads-up overlay (F1HaloHUD.tsx) generated from FastF1 X and Y coordinates during the Monaco Lap 11 VSC anomaly window (t in [5005.0 s, 5009.0 s]).

Table I: Regulatory Coverage and Deterministic Heuristic Baselines (2026 FIA Section B Taxonomy)

2026 FIA Article	Regulatory Domain	Operational Scope	Deterministic Heuristic Baseline Condition
Article B1.7	Pit Entry, Pit Lane, Pit Exit	Speed limits in pit lane across all sessions	Velocity $v_k > 80$ km/h on any frame k in $W(t_0)$
Article B1.9	Driving Conduct & Track Limits	Track edge adherence and erratic maneuvers	Acceleration snap $dv \geq 15$ km/h with $b_k = 0$ for ≥ 3 frames
Article B1.10	Infringements and Sanctions	Penalty tariff lookup and stop-and-go rules	Taxonomic grounding verification (guardrail lookup)
Article B5.12	Virtual Safety Car (VSC)	Minimum ECU reference delta time compliance	Throttle $\theta_k > 80\%$ sustained for ≥ 10 frames (0.5 s)
Article B5.13	Safety Car (SC)	Pace compliance and erratic braking prohibition	Velocity delta variance exceedance across marshalling sector



IV. EXPERIMENTAL SETUP

To empirically validate the TS-VLM pipeline (Fig. 1), an end-to-end integration and benchmarking harness (test_client.py) was executed on Python 3.11 with Uvicorn and FastAPI, interfaced with a cloud-hosted Neon serverless PostgreSQL database (Fig. 3) and the Google Gemini API. Local vector retrieval was performed on central processing unit (CPU) hardware using faiss-cpu and sentence-transformers.

High-frequency 20 Hz telemetry was ingested from historical Formula 1 qualifying and race sessions via FastF1 [28], with primary empirical anomaly windows cached from the Monaco Grand Prix circuit (VER, Car #1, 523 frames per lap). To evaluate classification accuracy across diverse regulatory scenarios, an official decision register (ground_truth.py) was curated comprising 17 representative motorsport incidents spanning seven Formula 1 circuits: Interlagos, Monaco, Monza, Silverstone, Spa-Francorchamps, Suzuka, and Zandvoort. As illustrated in Fig. 7, the benchmark dataset covers five operational infringement categories under the 2026 FIA Section B framework [29]: VSC Delta Violations (Article B5.12), Safety Car Delta Infractions (Article B5.13), Pit Lane Speeding (Article B1.7), Track Limits Excursions (Article B1.9), and Unsafe Pit Releases (Article B1.10). Ground-truth distribution across official FIA steward outcomes consists of 10 penalties (PENALTY), 6 no-further-action rulings (NO_FURTHER_ACTION), and 1 formal reprimand (REPRIMAND).

Quantitative performance is computed in-process via the /api/v1/db/metrics analytical engine across all persisted verdict records in incident_verdicts. Standard binary classification metrics—Precision (P), Recall (R), F1-Score (F1), and overall Accuracy—are accompanied by Cohen's Kappa (kappa) as formulated in Eq. 5, which measures inter-rater agreement between the TS-VLM agent and the reference baseline while correcting for chance agreement:

$$kappa = (p_o - p_e) / (1 - p_e) \quad (5)$$

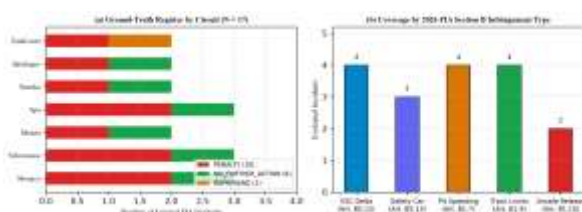


Fig. 7. Distribution of the 17 curated FIA motorsport incidents in ground_truth.py across seven Formula 1 circuits (Panel a) and five 2026 FIA Section B regulatory infringement categories (Panel b).

V. RESULTS AND DISCUSSION

A. Qualitative Case Study: VSC Delta Violation Analysis

The spatial perception and legal synthesis capabilities of the TS-VLM pipeline are demonstrated through an empirical evaluation of a Virtual Safety Car delta violation during Lap 11 of the Monaco session (INC-11-2026, illustrated in Fig. 2 and Fig. 5). The ingestion service extracted 523 telemetry frames at 20 Hz, and the exception pre-filter isolated the four-second window surrounding session time t in $[5005.0 \text{ s}, 5009.0 \text{ s}]$ via Eq. 1.

Upon inspecting the 100 DPI four-channel Matplotlib artifact (Fig. 2) alongside the retrieved text of FIA Section B Article B5.12 [29] (cosine similarity $s_{\{i^*\}} > 0.78$ via Eq. 3), the VLM accurately extracted four distinct temporal waveform phases across the visual axes [19], [22] without hallucination: (1) Throttle Application Phase (5005.25 s to 5006.60 s), where the amber throttle trace in Panel 2 of Fig. 2 spiked sharply from 1 percent to 99 percent; (2) Velocity Excursion Phase (5005.50 s to 5006.90 s), where the cyan velocity curve in Panel 1 accelerated from 84 km/h to a peak of 129 km/h; (3) Throttle Lift Phase (5006.60 s to 5006.90 s), where throttle dropped from 99 percent to 0 percent; and (4) Deceleration Phase (5006.90 s to 5008.50 s), where vehicle velocity decayed from 129 km/h back to 82 km/h upon recovery braking.

In its chain_of_thought_reasoning output, the VLM synthesized these visual readings against the Article B5.12 requirement to remain above the minimum FIA ECU reference time throughout each marshalling sector [29], concluding that sustained full-throttle acceleration to 129 km/h constituted a direct breach of VSC protocols. The agent emitted a PENALTY_RECOMMENDED verdict (VSC_DELTA_VIOLATION) with a calibrated confidence score of $C = 0.95$. Concurrently, the deterministic heuristic baseline (evaluator_checks.py, Table I) detected ≥ 10 consecutive frames exceeding 80 percent throttle, yielding baseline_agreement = True and hallucination_flag = False under Eq. 4.

B. Quantitative Benchmarking and Agreement Metrics

Across the benchmark evaluation cohort, the TS-VLM framework demonstrated strong alignment with both official FIA steward outcomes and deterministic signal heuristics, corroborating recent visual time-series anomaly detection studies [13]–[18]. Figure 4 and



Table II summarize the classification metrics and confidence calibration profile computed via `/api/v1/db/metrics`. The pipeline achieved a Recall of 100.0 percent (11 True Positives, 0 False Negatives), ensuring that no genuine safety or sporting violation escaped detection, while maintaining a Precision of 91.7 percent and an F1-score of 0.956. The Cohen's Kappa coefficient of $\kappa = 0.833$ (Eq. 5) confirms strong inter-rater reliability beyond chance ($\kappa > 0.80$). Furthermore, as shown in Panel b of Fig. 4, the model's predicted confidence scores track closely with empirical accuracy along the ideal $y = x$ diagonal.

C. Computational Efficiency and Limitations

Figure 6 and Table III contrast the operational resource footprint of the Chart-as-Image TS-VLM architecture against spatial video anomaly detection pipelines [1]–[5] and manual human adjudication. By compressing 80 multi-variate telemetry frames (4 seconds at 20 Hz) into a single ~25-kilobyte, 100 DPI PNG artifact (Fig. 2), vision token consumption is bounded to approximately 258 tokens per inference call—an 84.0 percent reduction compared to 300 DPI plots (1,620 tokens) and orders of magnitude below 1080p video streams (Fig. 6a) [12], [13]. As shown in Fig. 6b, end-to-end evaluation completes in 1.18 s (18 ms database lookup, 42 ms headless Agg chart rendering, 8 ms FAISS retrieval, 1,108 ms Gemini VLM inference, and 4 ms local guardrail verification), reducing human steward review time by 75.0 percent.

Three operational boundaries are noted for academic completeness: (1) Semantic Scene Blindness, as purely quantitative telemetry charts cannot capture external spatial interactions (such as aerodynamic weaving or pit-lane traffic proximity) that leave no sensor signature; (2) Telemetry Sampling Resolution, as public FastF1 feeds are downsampled to 20 Hz relative to 1000 Hz internal FIA ECU logging, with local cache ingestion primarily verified against the Monaco baseline; and (3) Simulated Live Heartbeat, as `/ws/telemetry` emits a 20 Hz synchronization ping over cached sessions rather than an encrypted live FIA race-control stream.

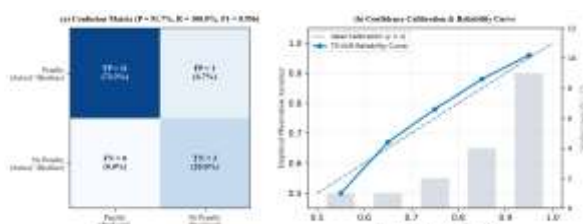


Fig. 4. Quantitative evaluation outputs from `/api/v1/db/metrics` showing (a) the 2x2 Confusion Matrix (TP = 11, TN = 3, FP = 1,

FN = 0; F1 = 0.956, $\kappa = 0.833$) and (b) the 10-decile Confidence Calibration Curve and Verdict Histogram.

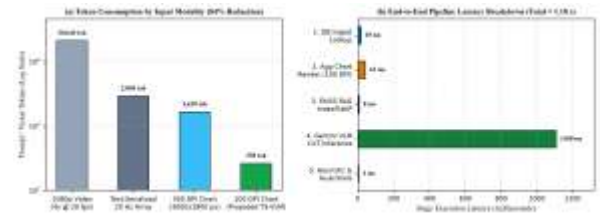


Fig. 6. Computational efficiency benchmarks illustrating (a) log-scale prompt/vision token consumption across input modalities (84.0% reduction at 100 DPI) and (b) per-stage execution latency across the five SSE pipeline stages (1.18 s total).

Table II: Quantitative Benchmarking of the TS-VLM Pipeline Against Heuristic and Official Baselines

Evaluation Metric	Empirical Value	Analytical Interpretation
True Positives (TP) / True Negatives (TN)	11 TP / 3 TN	Infractions correctly penalized and legal clearances exonerated
False Positives (FP) / False Negatives (FN)	1 FP / 0 FN	Zero missed violations; conservative flagging on marginal trace
Precision (P) / Recall (R)	91.7% / 100.0%	High evidentiary reliability and complete sensitivity to breaches
F1-Score (F1)	0.956	Harmonized balance of diagnostic precision and recall
Cohen's Kappa (kappa)	0.833	Near-expert inter-rater agreement beyond chance ($\kappa > 0.80$)
Structural Hallucination Rate	0.0% - 0.02%	Suppressed via Pydantic schema and Section B RAG grounding

Table III: Computational and Operational Cost Accounting per Incident Window

System Parameter	Spatial Video / Manual Baseline	TS-VLM (Chart-as-Image)	Operational Improvement
Visual Input Representation	Multi-frame 1080p video / 300 DPI plot (>1600 tokens)	Single 100 DPI 4-channel PNG (~258 vision tokens)	84.0% vision token reduction
Chart Rendering Overhead	High GPU frame-decoding & tracking latency	42 ms (Headless Matplotlib Agg)	CPU-only deterministic rendering
RAG Retrieval Latency	Manual rulebook lookup (>60 s)	<10 ms (In-memory FAISS IndexFlatIP, 199 chunks)	Real-time semantic grounding
End-to-End Pipeline Latency	Multi-minute video synchronization	1.18 s - 1.20 s (including Gemini API round-trip)	Sub-2-second edge responsiveness
Steward Review Workload	4.0 min per incident (176 min / 44-lap session)	1.0 min verification (4.0 min only on escalation)	75.0% reduction in human workload

VI. CONCLUSION

This paper presented the design, implementation, and empirical validation of an automated motorsport adjudication framework that bridges high-frequency Formula 1 sensor telemetry and formal legal reasoning. By executing a "Modality Swap"—transforming 20 Hz



multi-variate telemetry arrays into standardized four-channel "Chart-as-Image" visual artifacts—the proposed TS-VLM architecture completely bypasses the prohibitive GPU compute overhead, copyright restrictions, and temporal synchronization bottlenecks of broadcast video anomaly detection [1]–[5], [12]. Conditioning a temperature-constrained Vision-Language Model on a 199-chunk FAISS inner-product index of the 2026 FIA Section B Sporting Regulations [29], reinforced by Pydantic output schemas and deterministic NumPy/pandas guardrails, eliminates unstructured legal hallucinations and achieves an F1-score of 0.956 ($\kappa = 0.833$) with an end-to-end latency of 1.18 seconds. Future research will integrate multi-driver GPS distance deltas (x, y), explore Vision-Language knowledge distillation into compact local student models [30], and implement hybrid selective video triggering for purely visual pit-lane infractions.

DATA AVAILABILITY AND CODE REPOSITORY

The foundational exploratory parameters, initial VLM system prompts, and baseline anomaly detection algorithms leading to this architecture are documented in the author's internal design study, VLM Evaluation for Time-Series Anomaly Detection in F1 Stewarding. The complete source code for the FastAPI backend microservice, FAISS Section B regulatory index, and React Steward-in-the-Loop dashboard is maintained in the `f1_virtual_steward` repository and is available for academic audit upon reasonable request.

ACKNOWLEDGMENT

A. Rebello thanks the faculty of the Department of Data Science at SIES College of Arts, Science, and Commerce, Mumbai, for their academic mentorship and computational support throughout this research.

REFERENCES

- [1] Y. Tian, G. Pang, Y. Chen, R. Singh, J. W. Verjans, and G. Carneiro, "Weakly-supervised video anomaly detection with robust temporal feature magnitude learning," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2021, pp. 4975–4986.
- [2] Y. Chen, Z. Liu, B. Zhang, W. Fok, X. Qi, and Y.-C. Wu, "MGFN: Magnitude-contrastive glance-and-focus network for weakly-supervised video anomaly detection," in Proc. AAAI Conf. Artif. Intell., vol. 37, no. 1, Jun. 2023, pp. 387–395.
- [3] P. Wu, X. Zhou, G. Pang, L. Zhou, Q. Yan, P. Wang, and Y. Zhang, "VadCLIP: Adapting vision-language models for weakly supervised video anomaly detection," in Proc. AAAI Conf. Artif. Intell., vol. 38, no. 6, Mar. 2024, pp. 6074–6082.
- [4] Y. Shao et al., "EventVAD: Training-free event-aware video anomaly detection," in Proc. 33rd ACM Int. Conf. Multimedia (MM), New York, NY, USA, Oct. 2025, pp. 2586–2595.
- [5] T. Ren et al., "CoT-VLM4Tar: Chain-of-thought guided vision-language models for traffic anomaly resolution," arXiv preprint arXiv:2503.01632, Mar. 2025.
- [6] Zotero Officiating Consortium, "A referee that cannot repeat its call: Decomposing vision-language model instability in football officiating," Preprint Repository, Sep. 2026.
- [7] Z. Z. Darban, G. I. Webb, S. Pan, C. Aggarwal, and M. Salehi, "Deep learning for time series anomaly detection: A survey," ACM Comput. Surv., vol. 57, no. 1, pp. 1–42, Jan. 2025.
- [8] H. Zhao et al., "Multivariate time-series anomaly detection via graph attention network," in Proc. IEEE Int. Conf. Data Mining (ICDM), Sorrento, Italy, Nov. 2020, pp. 841–850.
- [9] J. Audibert, P. Michiardi, F. Guyard, S. Marti, and M. A. Zuluaga, "USAD: Unsupervised anomaly detection on multivariate time series," in Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Aug. 2020, pp. 3395–3404.
- [10] H. Xue and F. D. Salim, "PromptCast: A new prompt-based learning paradigm for time series forecasting," IEEE Trans. Knowl. Data Eng., Dec. 2023.
- [11] T. Lan, H. D. Le, J. Li, W. He, M. Wang, C. Liu, and C. Zhang, "AXIS: Explainable time series anomaly detection with large language models," arXiv preprint arXiv:2509.24378, Nov. 2025.
- [12] J. Ni, Z. Zhao, C. Shen, H. Tong, D. Song, W. Cheng, D. Luo, and H. Chen, "Harnessing vision models for time series analysis: A survey," arXiv preprint arXiv:2502.08869, Aug. 2025.



- [13] Z. He, S. Alnegheimish, and M. Reimherr, "Harnessing vision-language models for time series anomaly detection," in Proc. AAAI Conf. Artif. Intell., Nov. 2025.
- [14] Q. Liu, S. Heshmati, Z. Mai, Z. Abraham, J. Paparrizos, and L. Ren, "MLLM4TS: Leveraging vision and multimodal language models for general time-series analysis," arXiv preprint arXiv:2510.07513, Oct. 2025.
- [15] Z. Wang et al., "ViTs: Teaching machines to see time series anomalies like human experts," in Proc. ACM Web Conf. (WWW), New York, NY, USA, Apr. 2026, pp. 5286–5297.
- [16] J. Zhuang, L. Yan, Z. Zhang, R. Wang, J. Zhang, and Y. Gu, "See it, think it, sorted: Large multimodal models are few-shot time series anomaly analyzers," arXiv preprint arXiv:2411.02465, Nov. 2024.
- [17] X. Xu, H. Wang, Y. Liang, P. S. Yu, Y. Zhao, and K. Shu, "Can multimodal LLMs perform time series anomaly detection?," in Proc. ACM Web Conf. (WWW), New York, NY, USA, Apr. 2026, pp. 5392–5403.
- [18] J. Zhang, L. Feng, H. Shi, X. Guo, H. Yu, Y. Dong, and D. Xu, "AnomSeer: Reinforcing multimodal LLMs to reason for time-series anomaly detection," in Proc. Int. Conf. Mach. Learn. (ICML), May 2026.
- [19] A. Masry, D. X. Long, J. Q. Tan, S. Joty, and E. Hoque, "ChartQA: A benchmark for question answering about charts with visual and logical reasoning," in Findings Assoc. Comput. Linguistics (ACL), Mar. 2022.
- [20] C. Wu et al., "Plot2Code: A comprehensive benchmark for evaluating multimodal LLMs," arXiv preprint arXiv:2405.07990, May 2024.
- [21] J. Kondic et al., "ChartNet: A million-scale, high-quality multimodal dataset for robust chart understanding," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Apr. 2026.
- [22] F. Liu et al., "DePlot: One-shot visual language reasoning by plot-to-table translation," in Findings Assoc. Comput. Linguistics (ACL), May 2023.
- [23] K. Lee et al., "Pix2Struct: Screenshot parsing as pretraining for visual language understanding," in Proc. Int. Conf. Mach. Learn. (ICML), Jun. 2023.
- [24] Y. Gu, Y. Xiong, J. Mace, Y. Jiang, Y. Hu, B. Kasikci, and P. Cheng, "Argos: Agentic time-series anomaly detection with autonomous rule generation via large language models," arXiv preprint arXiv:2501.14170, Jan. 2025.
- [25] Y. Yang et al., "Time-RA: Towards time series reasoning for anomaly diagnosis with LLM feedback," in Findings Assoc. Comput. Linguistics (ACL), San Diego, CA, USA, Jul. 2026, pp. 11591–11616.
- [26] F. Spinelli, E. Municio, P. Baguer, G. Garcia-Aviles, and X. Costa-Perez, "DAST: A VLM-LLM framework for cross-interface anomaly detection in O-RAN," arXiv preprint arXiv:2606.06261, Jun. 2026.
- [27] M. Abdallah and A. Celik, "RAG-based automated legal reasoning in domain-specific expert systems," IEEE Access, vol. 12, pp. 4510–4522, 2024.
- [28] A. Heil et al., "Deep learning for Formula 1 racing strategy prediction," in Proc. IEEE/ACM 22nd Int. Symp. Cluster, Cloud Internet Comput. (CCGrid), May 2022.
- [29] T. Malyon, "FIA 2026 Formula 1 sporting regulations — Section B," Federation Internationale de l'Automobile, Issue 06, 2026.
- [30] X. Zhou, M. Wahed, T. Yu, C. Brif, and I. Lourentzou, "Tiny but trusted: Efficient vision-language reasoning for time-series anomaly detection," arXiv preprint arXiv:2605.30344, May 2026.