Unsupervised Learning for Robust Vehicle Impact Detection: A Statistical Replacement for Rigid Physical Thresholds

Hilario Aradiel Castañeda[1,*], Guillermo Antonio Mas-Azahuanche[1], Alfonso Herminio Geronimo Vasquez[2], Humberto Urbano Arteaga-Cortez[1] and Enrique Wilfredo Carpena-Velasquez[3]

[1]*Universidad Nacional del Callao, Callao, Peru*
[2]*Universidad Nacional de Ingeniería, Lima, Peru*
[3]*Universidad Nacional Pedro Ruiz Gallo, Lambayeque, Peru*

Abstract

This study presents an interpretable unsupervised anomaly-detection framework for prioritizing candidate vehicle-impact events from 5,447 triaxial acceleration records acquired with a low-cost MPU6050 sensor under urban driving conditions. The binary impact label available in the historical dataset was excluded from model fitting and reserved exclusively for external validation, thereby preserving the unsupervised nature of the analysis. The proposed pipeline combines Z-score standardization, Principal Component Analysis (PCA), exploration K-Means and DBSCAN clustering, Gaussian Mixture Model (GMM) density estimation, and a convergent anomaly criterion based on simultaneously low GMM log-likelihood and high PCA reconstruction error. Two principal components retained 75.51% of the total variance. K-Means with $k = 2$ achieved a Silhouette score of 0.5843, whereas ARI = 0.2706 and AMI = 0.1161 showed limited correspondence between the discovered clusters and the external impact labels. For DBSCAN, $\epsilon = 0.1$ and min_samples = 10 produced 8 clusters and 1,127 noise points (20.69%), revealing substantial sensitivity to hyperparameter selection. GMM and PCA each identified 273 candidates through percentile-based criteria; their intersection contained 131 statistically robust anomalies, corresponding to 2.405% of the dataset, with a Jaccard index of 0.3157. Because the row-level reference labels and a reproducible magnitude-threshold run could not be recovered from the archived material, external classification metrics cannot be reported and no claim of superiority over a physical-threshold detector is made. Accordingly, PCA+GMM is interpreted as an auditable unsupervised mechanism for prioritizing statistically robust candidate events rather than as a fully validated impact classifier.

Keywords

Unsupervised anomaly detection, Vehicle impact detection, Inertial sensors, MPU6050, DBSCAN

1. Introduction

Vehicle impact detection is a relevant problem in intelligent transportation systems, road safety, and the timely activation of emergency-response mechanisms. Low-cost inertial sensors enable widely available triaxial acceleration measurements; however, the resulting signals also contain vibration, abrupt maneuvers, road-surface irregularities, sensor bias, and mounting effects, such that an acceleration peak does not necessarily correspond to a true impact [1]. Recent studies using in-vehicle sensors have confirmed the potential of machine learning for detecting crashes and anomalous driving events, while also showing that performance strongly depends on dataset quality, labeling strategy, and the traceability of the experimental setup [2, 3].

Reported solutions range from magnitude-based physical rules to supervised, semi-supervised, and multisensor-fusion models. Deterministic thresholds are attractive because of their simplicity and low computational cost, but they require context-specific calibration and may respond to non-

IWSAI 2026: 1st International Workshop on Systems, Applications, Artificial Intelligence and Innovation, September 16–18, 2026, Piura, Peru
*Corresponding author.
haradielc@unac.edu.pe (H. A. Castañeda); gamasa@unac.edu.pe (G. A. Mas-Azahuanche); ageronimov@uni.edu.pe (A. H. G. Vasquez); huarteagac@unac.edu.pe (H. U. Arteaga-Cortez); ecarpena@unprg.edu.pe (E. W. Carpena-Velasquez)

critical disturbances. Supervised approaches, in turn, can provide direct discrimination when reliable ground truth is available, whereas semi-supervised methods aim to reduce dependence on exhaustive labeling [4, 2, 5]. Accordingly, the challenge is not merely to select an algorithm, but to construct a robust representation and a validation protocol suitable for rare events and incomplete experimental information.

Unsupervised anomaly detection is particularly relevant when labels are scarce, expensive to obtain, or unavailable during model fitting. The literature on multivariate signals shows that rarity may be expressed through low probability density, reconstruction error, disruption of inter-variable dependencies, or combinations of complementary criteria [6, 7, 8]. Nevertheless, a statistical anomaly should not be conflated with a physically confirmed event: detector quality can only be established externally when row-level correspondence between model outputs and actual events is preserved. This distinction is essential to avoid circular inference and unverifiable performance claims.

Against this background, the objective of this study is to design and analyze an interpretable unsupervised methodology that uses only the acceleration variables ax, ay, and az to construct the anomaly decision, while keeping the binary 0/1 label entirely outside model fitting. The research question is whether combining probabilistic rarity with structural deviation can isolate a consistent subset of candidate events without imposing a fixed physical decision boundary. To this end, Z-score standardization, PCA, K-Means, and DBSCAN are used for data characterization, GMM is used for density estimation, and PCA reconstruction error provides a second anomaly criterion [9, 10, 11].

The contribution is fourfold: (i) to formalize an unsupervised pipeline in which external labels do not participate in model fitting; (ii) to document the sensitivity of the clustering methods, including five recovered DBSCAN configurations and the operational GMM configuration; (iii) to define a convergent PCA+GMM criterion that reduces two sets of 273 candidates to 131 robust anomalies; and (iv) to explicitly distinguish statistically identifiable findings from external performance metrics that cannot be reconstructed from the archived evidence. The contribution therefore does not lie in claiming superiority over a physical threshold, but in establishing an auditable, interpretable methodology that is reproducible to the extent permitted by the available evidence.

2. Related Work

2.1. Vehicle Impact Detection Using Inertial Sensors

Inertial sensors provide a direct means of characterizing vehicle dynamics, but their interpretation must account for noise, drift, orientation, mounting, sampling frequency, and operating conditions [1]. Mahariba et al. [4] used inertial measurements and machine learning to detect accidents in powered two-wheelers, whereas Hozhabr Pour et al. [2] proposed a multimodal framework for cars using naturalistic data. Both studies demonstrate the value of physical sensing signals while also showing that the availability of real events and reliable labels constrains evaluation.

Recent research has broadened the problem from binary crash detection to risky-driving and vehicular-anomaly identification. Cheng et al. [5] introduced a benchmark and a semi-supervised model designed to reduce dependence on complete labels, whereas Onsu et al. [3] investigated multisensor fusion and deep-learning-based anomaly detection on experimental roads. Collectively, these contributions indicate a trend toward richer sensing systems, while simultaneously increasing the need to document acquisition, synchronization, labeling, and validation conditions.

2.2. Unsupervised Multivariate Anomaly Detection

When anomalous events are infrequent or difficult to label, unsupervised methods can model dominant behavior and assign higher anomaly scores to observations that depart from it. Belay et al. [6] reviewed unsupervised solutions for multivariate sensor time series and emphasized the importance of jointly representing inter-variable dependencies and temporal structure. Although the archived dataset used in this study lacks sufficient temporal metadata to be treated as a fully characterized time series, the three

acceleration axes ax, ay, and az form a multivariate representation in which cross-axis dependencies may contain information that a scalar threshold can overlook.

Contemporary evaluations further caution against summarizing anomaly-detector performance with a single metric. Mejri et al. [7] compared unsupervised methods in terms of Precision, Recall, F1-score, stability, model size, and anomaly type, whereas Correia et al. [8] identified benchmarking, threshold selection, ground-truth quality, and evaluation under real operating conditions as open challenges. These issues are directly relevant to the present work because the loss of row-level correspondence prevents classification performance from being attributed to the 131 convergent anomalies.

2.3. Classical and Interpretable Statistical-Rarity Methods

Among interpretable methods, PCA provides a lower-dimensional orthogonal representation and enables structural deviation to be quantified through reconstruction error [9]. K-Means identifies regimes around centroids, whereas DBSCAN explores dense regions and identifies noise points [10]. These techniques are useful for characterizing structure; however, strong internal separation does not imply correspondence with a physically meaningful label. Silhouette, ARI, and AMI therefore address different questions and must be interpreted separately [12, 13].

GMM represents the data distribution as a weighted combination of Gaussian components and provides a continuous log-likelihood that can be used as a rarity score [11]. Combining GMM with PCA reconstruction error integrates two complementary mechanisms: low probability under the density model and poor representation within the principal subspace. This approach retains interpretability and low dimensionality, but its practical utility depends on validating candidate events against real impacts and verifying that selected parameters remain stable across vehicles and operating conditions.

2.4. Research Gap and Positioning of the Study

The literature therefore reveals three relevant families of approaches: simple physics-based detectors, supervised or semi-supervised models that rely on labels, and unsupervised detectors of multivariate rarity. Physics-based methods are operationally simple but calibration-sensitive; supervised methods can provide direct discrimination when high-quality ground truth is available; and unsupervised methods can explore rare events without using labels during fitting, although they still require rigorous external evaluation [6, 2, 5, 7, 8, 3]. An applied gap remains for interpretable, low-complexity vehicular detectors that clearly distinguish internal structure, statistical anomaly, and physical confirmation of an impact.

The present study addresses this gap by using K-Means and DBSCAN for exploratory characterization and a PCA+GMM consensus rule to prioritize candidate events, while reserving the 0/1 label exclusively for external validation. This design avoids turning clustering into supervised classification and allows the scope of the findings to be stated precisely: the 131 convergent observations are robust statistical anomalies that are candidates for impacts, not confirmed impacts. Table 1 compares the most relevant studies and approaches and highlights the distinctive contribution of the present work.

3. Materials and Methods

3.1. Study Design and Unsupervised Learning Setting

This study adopts an applied, quantitative, exploratory experimental design focused on anomaly detection in multivariate inertial signals. Each triaxial acceleration record constitutes one unit of analysis. StandardScaler, PCA, K-Means, DBSCAN, and GMM are fitted exclusively using the acceleration variables; the binary reference label (0 = non-impact; 1 = impact) is kept separate and is used only after model fitting for external evaluation. Consequently, the decision boundary is not learned from known classes, and the training procedure remains unsupervised.

The methodological workflow comprises historical signal acquisition and cleaning, Z-score standardization, PCA projection, exploratory analysis with K-Means and DBSCAN, density estimation with

Table 1
Summary of relevant approaches and their methodological role in this study.

Reference / approach	Data / sensors	Paradigm and contribution	Limitation / relationship to this study
Mahariba et al. [4], 2022	Inertial measurements from powered two-wheelers	Supervised; automatic crash detection using inertial variables	Requires labels and specific operating conditions; demonstrates the value of IMUs as a physical sensing source.
Hozhabr Pour et al. [2], 2022	Multimodal in-vehicle sensors	Supervised; detection framework combining feature extraction and classification	Availability of real crashes and reliable ground truth constrains evaluation.
Belay et al. [6], 2023	Multivariate IoT time series	Review of unsupervised multivariate anomaly detection	Highlights inter-variable dependencies and benchmarking challenges; supports the unsupervised paradigm.
Cheng et al. [5], 2024	Risky-driving data	Semi-supervised; benchmark designed to reduce dependence on complete labels	Does not specifically address PCA+GMM consensus or traceability of archived experiments.
Mejri et al. [7], 2024	Multiple time-series datasets	Broad evaluation of unsupervised methods and performance protocols	Emphasizes the need for multiple metrics and anomaly types in evaluation.
Correia et al. [8], 2024	Multivariate time series	Taxonomy of online anomaly detection and validation challenges	Identifies threshold selection, ground truth, and benchmarking as open challenges.
Onsu et al. [3], 2025	Vehicular multisensor fusion	Deep learning and anomaly detection on experimental roads	Higher complexity and data dependence; reinforces the need for experimental traceability.
Present study	MPU6050; 5,447 ax, ay, az records	Unsupervised and interpretable; GMM(Q5)+PCA(Q95) consensus with exploratory K-Means/DBSCAN	Prioritizes 131 robust candidates; does not claim classification performance without row-level ground truth.

GMM, and a convergent PCA+GMM decision rule. Figure 1 summarizes this architecture and explicitly separates unsupervised detector construction from subsequent validation using external labels.

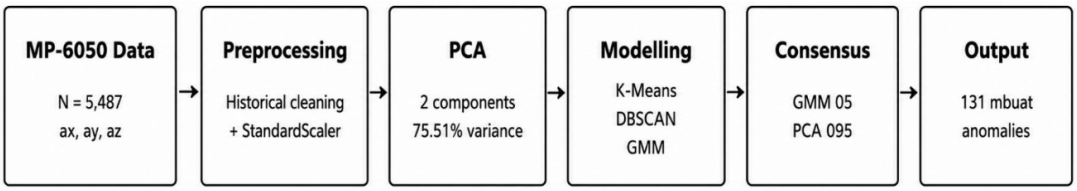

Figure 1: Methodological architecture of the convergent anomaly-detection framework.

3.2. Data Source and Experimental Setup

An audit of the archived experimental material was conducted to reconstruct the origin of the data and the available setup information. Primary evidence identifies the file datos_mpu_limpio.csv, containing 5,447 acceleration records along the ax, ay, and az axes acquired using an MPU6050 sensor installed in a vehicle under urban driving conditions. The original documentation does not identify a driving simulator as the primary experimental platform; accordingly, the dataset is treated as a real-vehicle acquisition. However, the vehicle makes model, and year; exact route; pavement type; sampling frequency and range; sensor location and orientation; and gravity/inclination compensation procedure could not recover reliably. A summary of the recovered dataset evidence and the corresponding methodological treatment is provided in Table 2.

Accordingly, the findings are restricted to the archived dataset and are not presented as evidence of generalization across vehicles or mounting configurations. Missing metadata is reported as reproducibility limitations rather than being completed through assumptions. This restriction is particularly relevant when comparing the method with magnitude thresholds because the physical interpretation of acceleration depends on units, orientation, gravity, mounting, and sampling frequency.

Table 2
Traceability of the dataset and recovered experimental metadata.

Item	Recovered evidence	Methodological treatment
Data file	datos_mpu_limpio.csv; 5,447 records.	Historical primary source; the physical file could not be recovered in the current audit.
Input variables	ax, ay, az.	Feature matrix X for the unsupervised pipeline.
External label	Binary impact label 0/1.	Used only for external validation; N_0/N_1 were not recovered.
Sensor	MPU6050.	Confirmed in the original documentation.
Platform	Sensor installed in a vehicle under urban driving conditions.	No driving simulator is documented as the primary experimental platform.
Vehicle / route / surface	Not recovered.	Vehicle make/model/year, route, and pavement type are reported as limitations.
Sampling frequency / range	Not recovered.	Prevents temporal reproducibility and a fully calibrated physical comparison.
Mounting / orientation / gravity	Not recovered.	Affects physical interpretation and transfer across vehicles.

3.3. External Labels and Impact Prevalence

The documentation confirms the existence of a binary reference variable $y_i \in \{0, 1\}$, used exclusively to evaluate the unsupervised results externally. Let $N = 5,447$ denote the total number of records, N_1 the number of reference impacts, and N_0 the number of non-impact records. Prevalence is defined by equation (1). However, neither the historical manuscript nor the recovered files preserve the N_0/N_1 counts or the row-level labels outside the original CSV; consequently, prevalence is neither imputed nor estimated from the 131 detected anomalies.

$$\pi\_impact = N_1/N, \; with \quad N_0 = N - N_1 \tag{1}$$

This distinction avoids circular inference: the 2.405% selected by the PCA+GMM intersection is the detector selection rate, not the empirical prevalence of impacts. The external label is not used to define percentiles, select anomalies, or refit the model.

3.4. Data Preparation and Dimensionality Reduction

The original experimental version describes the dataset as previously cleaned and standardized, with null values, static deviations, and values considered obvious outliers removed. The exact criterion used for this preprocessing step is not preserved in the audited material; because anomaly detection is sensitive to the prior removal of extreme values, this missing information is treated as a reproducibility limitation. Z-score standardization is then applied to the cleaned dataset using StandardScaler [14, 15], as shown in equation (2).

$$z_{ij} = (x_{ij} - \mu_j)/\sigma_j \tag{2}$$

PCA is subsequently applied. Two components retain 75.51% of the total variance, providing a latent representation and a measure of structural deviation [9]. The projection and squared reconstruction error are expressed in equations (3) and (4), respectively.

$$Z = XW \tag{3}$$

$$e_i = ||x_i - \hat{x_i}||_2^2 \tag{4}$$

Large values of e_i indicate observations that are poorly represented by the principal subspace. PCA is therefore used both for dimensionality reduction and as the structural component of the anomaly detector.

3.5. Gaussian Mixture Model Configuration

Archived execution evidence confirms that the GMM used in the subsequent comparison with the impact variable employed $K = 2$ components and covariance_type = 'tied'. During development, different numbers of components and covariance structures were explored using AIC/BIC. However, the preserved records do not allow a complete and consistent selection procedure to be reconstructed that would justify $K = 2$ with tied covariance as the global BIC optimum. The manuscript therefore reports $K = 2$, tied as the operational configuration actually used in the subsequent analysis, without attributing to it an optimality claim that cannot be demonstrated from the historical record.

$$BIC = -2ln(L) + qln(N) \tag{5}$$

In equation. (5), L denotes the maximized likelihood, q the number of free parameters, and N the number of observations. Once the operational configuration is fixed, the GMM density is given by equation (6) [11].

$$p(x_i) = \sum_{k=1}^{K} \pi_k N(x_i|\mu_k, \sum_k) \tag{6}$$

Probabilistic rarity is quantified using $\log p(x_i)$. In the historical run, the 5th percentile of the log-likelihood was -6.67 and was used as an empirical low-density cutoff. This value is specific to that run and is not interpreted as a universal constant transferable to other vehicles or mounting configurations.

3.6. DBSCAN and Hyperparameter Sensitivity Analysis

DBSCAN is used as an exploratory density-based analysis to determine whether rare events appear as spatial noise or occur within dense regions [10]. The execution log preserves five combinations of ϵ and min_samples: (0.1, 3) produced 45 clusters and 601 noise points; (0.1, 5), 21 clusters and 780 noise points; (0.1, 10), 8 clusters and 1,127 noise points; (0.3, 3), 108 clusters and 233 noise points; and (0.3, 5), 60 clusters and 450 noise points.

The configuration used in the subsequent visual comparison with the impact label was $\epsilon = 0.1$ and min_samples = 10. With $N = 5,447$, its 1,127 noise points represent 20.69% of the dataset. The pronounced variation in both the number of clusters and the amount of noise across configurations confirms DBSCAN's high sensitivity to its hyperparameters. The label y_i was not involved in fitting; it was overlaid only afterward to observe that many impact events were located among points labeled as noise (-1). The exact number of impacts within the noise set is not preserved and is not estimated visually.

3.7. Convergent PCA+GMM Criterion

The primary unsupervised decision rule combines probabilistic rarity and structural deviation. A_GMM is defined as the set of observations whose log-likelihood lies in the lowest 5%, whereas A_PCA is the set whose reconstruction error lies in the highest 5%. The robust anomaly set is obtained by intersecting these two criteria, as expressed in equation (7).

$$A^R = \{i : logp(x_i) \leq Q_{0.05}[log\ p(x)] \wedge e_i \geq Q_{0.95}[e]\} \tag{7}$$

The 5th and 95th percentile thresholds generate initial sets of similar size by construction; therefore, 273 GMM candidates and 273 PCA candidates do not constitute evidence of convergence. Convergence is established only through row-level coincidence: A^R contains 131 observations shared by both criteria.

3.8. External Validation Protocol

External evaluation of the detector requires comparing the PCA+GMM decision with the binary reference label on a record-by-record basis. This cross-comparison yields true positives, false positives, true negatives, and false negatives, from which the performance metrics are derived. Specifically, precision, recall, specificity, and F1-score are calculated according to equations (8)–(11), respectively. These metrics do not participate in training and are reserved exclusively for performance evaluation after unsupervised fitting is complete.

$$Precision = TP/(TP + FP) \tag{8}$$

$$Recall = TP/(TP + FN) \tag{9}$$

$$Specificity = TN/(TN + FP) \tag{10}$$

$$F1 = 2 \cdot Precision \cdot Recall/(Precision + Recall) \tag{11}$$

However, the archived documentation does not preserve row-level correspondence between the external labels and the 5,447 observations. These metrics are therefore presented as part of the validation protocol rather than as study results. Overall accuracy should likewise not be used as the sole metric when the impact class is rare; if a future replication preserves a continuous anomaly score and complete ground truth, PR-AUC would provide an appropriate complementary measure.

3.9. Acceleration-Magnitude Baseline

To formalize comparison with a deterministic approach, a baseline is defined using the vector magnitude of acceleration. For each record, m_i is computed according to equation (12), and a decision is obtained from a threshold T using equation (13).

$$m_i = \sqrt{(a_{x,i}^2 + a_{y,i}^2 + a\_z, i^2)} \tag{12}$$

$$\hat{y}_i(T) = 1 \quad if\ m_i > T; \quad \hat{y}_i(T) = 0\ otherwise \tag{13}$$

The historical archive does not preserve an operational value of T, a baseline run over the 5,447 records, or sufficient metadata to guarantee a homogeneous physical interpretation of acceleration magnitude. The baseline is therefore formally defined but is not presented as an experiment that was actually executed. Any future estimate of T should be obtained from a documented engineering criterion or from an independent calibration subset to prevent information leakage.

3.10. Scope of the PCA+GMM-versus-Baseline Comparison

An empirical comparison must use exactly the same records and the same external labels for both detectors. In addition to classification metrics, the sets $D\_PG \cap D\_T$, $D\_PG - D\_T$, and $D_T - D\_PG$ should be analyzed because they represent common detections and disagreements between the convergent detector and the magnitude baseline. Such an analysis would determine whether the multivariate criterion captures information from structurally atypical events that are not merely large-magnitude peaks.

With the evidence currently available, only the decision principles of the two approaches can be compared, not their quantitative performance. Consequently, no superiority of PCA+GMM over the baseline is claimed until row-level ground truth and a reproducible threshold run are available. The mathematical notation used in Sections 3 and 4 is summarized in Table 3.

Table 3
Mathematical notation and variables used in the validation framework.

Symbol	Definition	Interpretation
X	N × p matrix of standardized variables.	Input to unsupervised model fitting.
y_i	Binary 0/1 reference label.	External validation only.
$N_0,\ N_1$	True counts of non-impact and impact records.	Required to estimate prevalence and perform external validation.
π_impact	N_1/N.	True prevalence; must not be confused with the detector selection rate.
e_i	PCA reconstruction error.	Structural deviation.
$log\ p(x_i)$	GMM log-likelihood.	Probabilistic rarity.
A^R	$A\_GMM \cap A\_PCA$.	131 robust candidate anomalies.
m_i	Vector magnitude of acceleration.	Score used by the deterministic baseline.
T	Magnitude threshold.	Not recovered from the historical experiment.

4. Results

The results are restricted to values that can be verified from the archived experimental material and are explicitly distinguished from quantities that require the original tabular file. This strategy allows recoverable quantitative results—including the DBSCAN counts—to be reported without presenting metrics that cannot be rigorously reconstructed as experimental evidence.

4.1. Dataset Traceability and Reference Impact Prevalence

The analysis was conducted on 5,447 triaxial acceleration records (ax, ay, and az), and the documentation confirms the existence of a binary 0/1 impact label. After reviewing historical versions, execution screenshots, and the available files, neither the N_0/N_1 counts nor the row-level correspondence y_i could be recovered. Consequently, the true prevalence of impacts cannot be calculated and is not inferred from the 131 robust anomalies.

The 2.405% reported below represents only the fraction selected by the PCA+GMM intersection. It is not used as an estimate of prevalence, sensitivity, precision, or coverage of the impact class.

4.2. PCA Results and Reconstruction Criterion

The two-component PCA projection retained 75.51% of the total variance, leaving 24.49% outside the two-dimensional subspace. Figure 2 shows this proportion. In the archived experimental run, the 95th percentile of reconstruction error was 2.11; applying this cutoff identified 273 structurally atypical observations. The value 2.11 is an empirical threshold from that run rather than a universal constant.

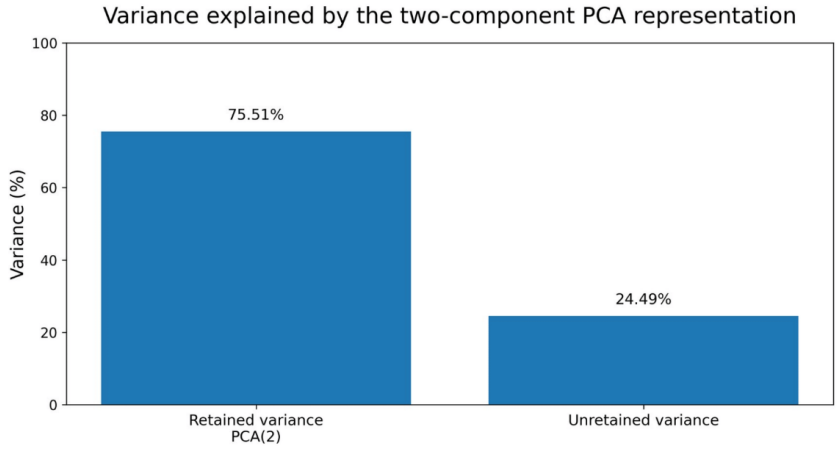

Figure 2: Variance retained by the two-component PCA representation.

In the archived experimental run, the 95th percentile of reconstruction error was 2.11. Applying this cutoff identified 273 records with reconstruction error equal to or greater than this value. This result corresponds to the structural component of the detector: a record is considered a PCA anomaly candidate when it cannot be adequately represented by the principal subspace. The value 2.11 should therefore be interpreted as an empirical threshold from that specific run, not as a universal constant transferable to other vehicles or configurations.

4.3. K-Means Results and Alignment with the Impact Label

K-Means with $k = 2$ represented two dominant motion regimes and achieved Silhouette = 0.5843. Figure 3 summarizes this internal metric together with ARI = 0.2706 and AMI = 0.1161 relative to the impact label. The Silhouette value is consistent with a reasonably well-defined internal structure; however, ARI and AMI indicate limited correspondence between the partition and the impact variable.

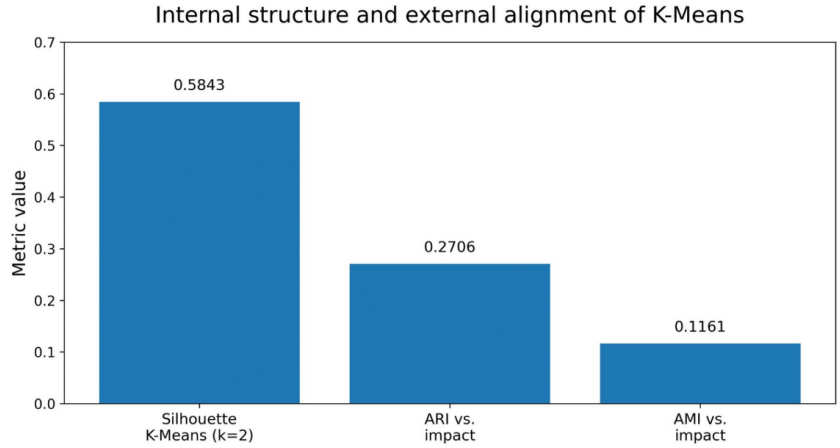

Figure 3: Internal structure and external alignment of K-Means clustering.

Historical graphical evidence shows that the cluster associated with higher-magnitude observations

contains impacts but also many records not labeled as impacts. K-Means therefore describes motion regimes and should not be interpreted as a binary impact classifier.

4.4. DBSCAN Results

The archived execution log makes it possible to recover DBSCAN's sensitivity quantitatively. With $\epsilon = 0.1$ and min_samples = 3, the algorithm produced 45 clusters and 601 noise points (11.03%); with $\epsilon = 0.1$ and min_samples = 5, 21 clusters and 780 noise points (14.32%); with $\epsilon = 0.1$ and min_samples = 10, 8 clusters and 1,127 noise points (20.69%); with $\epsilon = 0.3$ and min_samples = 3, 108 clusters and 233 noise points (4.28%); and with $\epsilon = 0.3$ and min_samples = 5, 60 clusters and 450 noise points (8.26%).

The configuration used in the subsequent comparison with the impact label was $\epsilon = 0.1$ and min_samples = 10. Its result—8 clusters and 1,127 noise points—confirms that DBSCAN is highly sensitive to parameterization. The historical comparison figure shows many impacts overlaid on the noise class (-1), supporting its exploratory value for spatial rarity; however, the exact number of impacts within the noise set is not preserved, and DBSCAN is therefore not reported as a validated classifier.

4.5. Gaussian Mixture Model Results

The operational configuration used in the historical comparison with the impact variable was a GMM with $K = 2$ and covariance_type = 'tied'. Development included an AIC/BIC sweep over different numbers of components and covariance structures, but the archived evidence is insufficient to establish that $K = 2$ with tied covariance corresponds to the global minimum BIC. Any claim of BIC optimality is therefore omitted, and only the configuration actually used in the subsequent analysis is retained. This distinction improves conceptual reproducibility and avoids attributing to the information criterion a selection that cannot be reconstructed unambiguously.

In that run, the 5th percentile of the log-likelihood was -6.67. Using this criterion, 273 low-density observations were identified, corresponding to approximately the lowest 5% of the log-likelihood distribution. Detection was performed without using the impact label during model fitting.

4.6. Convergent PCA+GMM Detection

GMM and PCA each generated 273 candidates by percentile construction. Figure 4 shows that row-level intersection reduced the selection to 131 robust anomalies. Equality 273/273 does not constitute convergence; evidence of consensus arises exclusively from simultaneous membership in both sets.

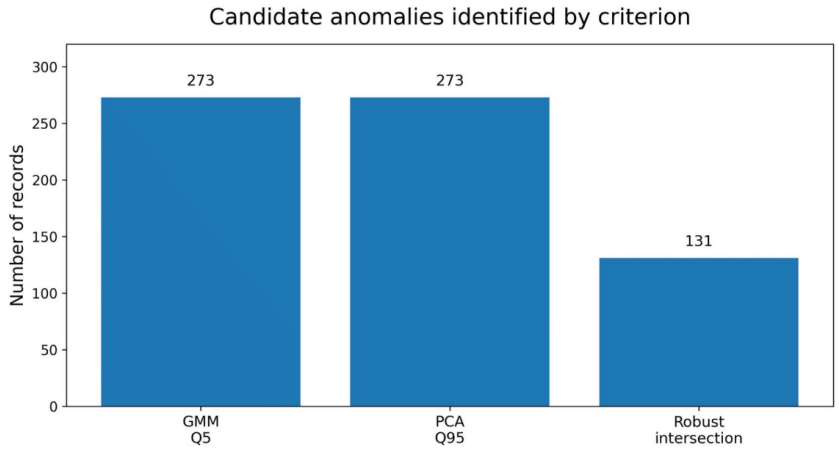

Figure 4: Candidate counts identified by GMM, PCA, and their robust intersection.

Of the 273 candidates identified by each method, 131 were shared, 142 were unique to GMM, and 142 were unique to PCA. Their union contains 415 observations (7.619% of the dataset), whereas the

intersection represents 2.405%. The Jaccard index is 131/415 = 0.3157. Figure 5 summarizes these relationships and shows partial overlap rather than equivalence between the two criteria.

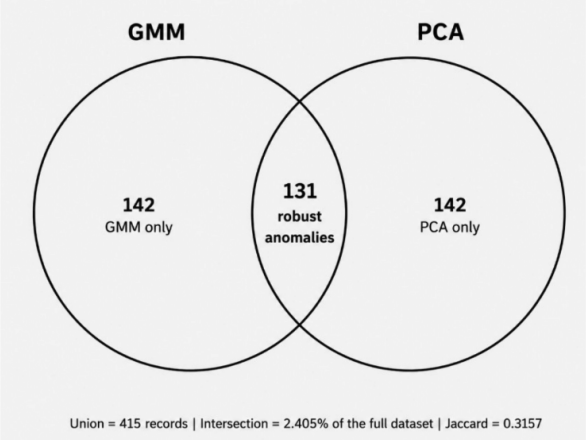

Figure 5: Candidate counts identified by GMM, PCA, and their robust intersection.

Accordingly, the 131 observations are referred to as robust anomalies or impact candidates. The available evidence demonstrates convergent statistical rarity but does not support labeling them as true impacts without row-level external validation.

4.7. Scope of External Validation

The convergent detector selected 131 observations as robust anomalies and left 5,316 observations outside the selected set. These values describe detector output only and must not be interpreted as classification successes or errors. Although the historical documentation confirms the existence of a binary reference label, it does not preserve its row-level correspondence with the 5,447 observations.

Consequently, Precision, Recall, Specificity, F1-score, ROC-AUC, and PR-AUC cannot be identified rigorously from the historical screenshots. The ARI and AMI values for K-Means do not substitute for these metrics because they assess correspondence between partitions rather than the performance of the convergent detector. The 131 observations are therefore retained as statistically robust impact candidates, and claims of high precision, demonstrated false-positive reduction, or superior sensitivity that cannot be verified against the original ground truth are avoided.

4.8. Comparison with Magnitude Baseline

A quantitative magnitude-baseline ran over the same 5,447 records and an operational threshold T could not be recovered. A comprehensive comparison of Precision, Recall, or F1-score between the baseline and PCA+GMM is therefore not possible from the available material. What can be compared is the decision principle: the baseline responds exclusively to scalar magnitude, whereas PCA+GMM combines low probabilistic density with multivariate structural deviation.

In terms of input regimes, a magnitude threshold is designed to respond to large peaks, whereas the convergent criterion may select observations whose rarity arises from the joint relationship among ax, ay, and az even when overall magnitude is not extreme. This difference is a mechanistic hypothesis rather than evidence of superiority. Demonstrating it requires comparing the disagreement subsets D_PG - D_T and D_T - D_PG against y_i.

The recovered quantitative results and the reproducibility status of each component are summarized in Table 4, which explicitly reports the status of external validation and the quantities that cannot be reconstructed without row-level ground truth.

Table 4
Recovered quantitative results and reproducibility status.

Indicator	Recovered value / evidence	Interpretation
Records analyzed	5,447	Dataset size used for modeling.
External label	Binary 0/1; N_0/N_1 not recovered	Ground truth exists, but prevalence cannot be reconstructed.
PCA(2): retained variance	75.51%	24.49% of the variance is not retained.
PCA: Q95 reconstruction error	2.11	Empirical threshold from the historical run.
PCA anomalies	273	Upper 5% of reconstruction error.
K-Means	$k = 2$	Two dominant motion regimes.
Silhouette	0.5843	Reasonably well-defined internal structure.
ARI / AMI vs. impact	0.2706 / 0.1161	Limited alignment between clusters and the impact label.
DBSCAN $\epsilon = 0.1$, min_samples=3	45 clusters; 601 noise points (11.03%)	High fragmentation.
DBSCAN $\epsilon = 0.1$, min_samples=5	21 clusters; 780 noise points (14.32%)	Larger noise set.
DBSCAN $\epsilon = 0.1$, min_samples=10	8 clusters; 1,127 noise points (20.69%)	Configuration used in comparison with impact labels.
DBSCAN $\epsilon = 0.3$, min_samples=3	108 clusters; 233 noise points (4.28%)	Greater fragmentation with less noise.
DBSCAN $\epsilon = 0.3$, min_samples=5	60 clusters; 450 noise points (8.26%)	Marked hyperparameter sensitivity.
Operational GMM	$K = 2$; covariance_type='tied'	Configuration used; no claim of global BIC optimality.
GMM: Q5 log-likelihood	-6.67	Empirical low-density cutoff.
GMM anomalies	273	Lowest 5% of log-likelihood values.
Robust anomalies	131 (2.405%)	$PCA \cap GMM$ intersection; candidates, not confirmed impacts.
GMM-only / PCA-only	142 / 142	Non-overlapping components.
$PCA \cup GMM$ union	415 (7.619%)	Selected by at least one criterion.
PCA-GMM Jaccard index	0.3157	Partial overlap between candidate sets.
External validation	Row-level ground truth not recovered	Precision, Recall, Specificity, F1-score, ROC-AUC, and PR-AUC are not reported; the 131 selections are treated as candidates rather than confirmed impacts.

4.9. Summary of Findings

Four empirical conclusions can be supported within the scope of the archived dataset. First, PCA(2) retains 75.51% of the variance, and K-Means identifies a reasonably well-defined internal structure but with limited alignment with the impact label. Second, DBSCAN is highly sensitive to ϵ and min_samples; the configuration $\epsilon = 0.1$, min_samples = 10 yields 8 clusters and 1,127 noise points (20.69%). Third, GMM and PCA each identify 273 candidates, and their intersection yields 131 robust anomalies with a Jaccard index of 0.3157. Fourth, the archived evidence does not permit calculation of the true prevalence, external classification metrics, or a quantitative comparison with a magnitude threshold. PCA+GMM should therefore be interpreted as an unsupervised detector of statistically robust candidates rather than as a validated impact classifier.

5. Discussion

5.1. Latent Structure and the Meaning of Clustering

The results first distinguish two questions that should not be conflated in impact detection: whether the signal contains internal structure and whether that structure corresponds to the impact class. K-Means with $k = 2$ achieved Silhouette = 0.5843, which is consistent with two reasonably well-defined motion regimes; however, ARI = 0.2706 and AMI = 0.1161 indicate limited correspondence with the binary impact label. The clusters therefore represent dominant patterns of vehicle dynamics but do not, by themselves, constitute impact/non-impact classes [12].

This discrepancy supports treating events of interest as rare observations relative to dominant dynamics rather than expecting them to form a compact, independent cluster. Historical evidence shows that the cluster associated with higher magnitudes contains both impacts and records not labeled as impacts; directly converting the K-Means assignment into a binary decision would therefore produce a conceptually inappropriate boundary for an infrequent class.

5.2. DBSCAN as Evidence of Spatial Rarity and Parameter Sensitivity

DBSCAN provides a complementary view of the data geometry but also exhibits pronounced dependence on its hyperparameters [10]. The five recovered configurations produce between 8 and 108 clusters and between 233 and 1,127 noise points. In particular, $\epsilon = 0.1$ and min_samples = 10—the configuration used in the visual comparison with the external label—produces 8 clusters and 1,127 noise observations, representing 20.69% of the dataset. This variability confirms that the notion of spatial noise is not stable with respect to changes in ϵ and min_samples.

The visual presence of many impacts within label -1 is therefore consistent with the spatial-rarity hypothesis but does not justify interpreting DBSCAN as a validated classifier. The exact number of impacts within the noise set is not preserved; sensitivity, precision, and false-positive rate cannot therefore be calculated for this model. Its role in the present study is exploratory: it shows that some relevant events lie outside dense regions and that any decision based solely on local density would be highly parameter-sensitive.

5.3. Complementarity Between GMM and PCA Reconstruction Error

GMM and PCA capture two distinct forms of rarity. GMM identifies observations located in low-probability-density regions, whereas PCA reconstruction error identifies observations poorly represented by the principal subspace [9, 11]. In the historical run, each criterion selected 273 records by percentile construction; the equality 273/273 therefore does not indicate agreement between the models.

Consensus becomes evident only when the selected records are compared directly: 131 observations belong to both sets, 142 are unique to GMM, and 142 are unique to PCA. Their union contains 415 records, and the Jaccard index is 0.3157. This degree of overlap indicates partial complementarity between density and reconstruction: the criteria capture related, but non-equivalent, regions of rarity. The intersection reduces the selection rate from 7.619% for the union to 2.405% of the full dataset, thereby increasing the statistical selectivity of the filter.

Statistical selectivity, however, is not equivalent to diagnostic accuracy. The 131 observations must remain classified as robust anomalies or impact candidates until row-level external validation becomes available. Likewise, the GMM configuration $K = 2$ with covariance_type = tied is reported because it was operationally used in the historical analysis; the archived evidence does not demonstrate that it corresponds to the global minimum BIC, and no unreproducible optimality claim is made.

5.4. Scope of External Validation and Relationship to the Magnitude Baseline

With respect to detection capability, the available evidence confirms that PCA+GMM selects 131 observations and excludes 5,316, but it cannot determine which of these correspond to true impacts.

There is therefore no sufficient basis for reporting Precision, Recall, Specificity, F1-score, ROC-AUC, or PR-AUC. This limitation is substantive: ARI and AMI characterize the alignment of K-Means with an external label but do not replace record-level validation of the convergent detector. The analysis is consequently restricted to statistical rarity, complementarity between criteria, and consensus selectivity, avoiding unverifiable claims of classification performance.

Comparison with a magnitude-based detector must likewise remain at the level of a mechanistic hypothesis until the original dataset is recovered or the acquisition is repeated. A scalar threshold responds to $m_i = ||a_i||$ and therefore prioritizes magnitude peaks, whereas PCA+GMM jointly considers multivariate relationships and density. This difference explains why the approaches may respond to different input regimes, but it does not establish that one is superior. Without a common run, a documented threshold T, and the same ground truth for both detectors, it is not valid to claim demonstrated false-positive reduction, higher sensitivity, or definitive replacement of the physics-based criterion.

5.5. Generalization and Transition Toward Supervised Validation

Generalization is constrained by dependence on the physical domain. Differences in suspension, mass, MPU6050 mounting and orientation, gravity, sampling frequency, and pavement may alter the distributions of ax, ay, and az [1]. Because several of these metadata could not be recovered, the findings should be interpreted as valid for the archived dataset rather than as a universal calibration across vehicles. Accordingly, Table 5 summarizes the evidence recovered from each analytical component and defines the corresponding limits of interpretation.

Within this context, the 131 robust anomalies may be used to prioritize human review or to help construct a labeled dataset, but they should not be treated as definitive pseudo-labels. Transition to supervised learning requires confirming each event using an external source—such as synchronized video, a timestamped log, or a controlled test—and clearly separating data used for calibration from data reserved for evaluation. Domain adaptation may subsequently be investigated as a strategy for transferring parameters across vehicles, provided its effect is measured explicitly [16].

Table 5
Interpretation of findings and limits of inference.

Component	Recovered evidence	Valid interpretation and limit
K-Means	Silhouette = 0.5843; ARI = 0.2706; AMI = 0.1161.	Motion structure is present, but clusters are not equivalent to impact classes. K-Means should not be used directly as a binary impact classifier.
DBSCAN	$\epsilon = 0.1$, min_samples = 10: 8 clusters and 1,127 noise points (20.69%); other configurations vary substantially.	Provides evidence of spatial rarity and strong hyperparameter sensitivity. Classification performance cannot be inferred without ground truth.
GMM	Operational $K = 2$, tied; Q5 = -6.67; 273 candidates.	Quantifies low-density observations. Global BIC optimality is not claimed from the available evidence.
PCA	Two components retain 75.51%; Q95 reconstruction error = 2.11; 273 candidates.	Quantifies structural deviation; 24.49% of the variance lies outside the two-component representation.
PCA+GMM	131 intersections; Jaccard = 0.3157; selection rate = 2.405%.	More selective convergent filter. The 131 cases are candidates, not confirmed impacts.
External validation	Row-level ground truth not recovered; the detector selects 131 observations.	Precision, Recall, F1-score, Specificity, ROC-AUC, and PR-AUC cannot be identified rigorously; the 131 observations remain candidate events.
Magnitude baseline	m_i formalized; T and the comparative run were not recovered.	Only decision principles can be compared. Superiority of PCA+GMM is not demonstrated.

6. Considerations for Embedded Implementation

6.1. Computational Profile of the Current Configuration

Embedded feasibility must distinguish theoretical algorithmic complexity from hardware-measured performance. The pipeline operates with $p = 3$ input variables, $k = 2$ PCA components, and an operational GMM with $K = 2$ and covariance_type = tied. Per record, standardization requires O(p), PCA projection and reconstruction require O(pk), and GMM evaluation requires approximately O(Kp^2) when the Mahalanobis quadratic form is considered [11]. With $p = 3$, these operations are mathematically compact; however, asymptotic complexity alone does not establish real-time execution, low power consumption, or computational superiority.

The statistical state is also compact: the GMM contains 13 free parameters (one independent mixture weight, six mean parameters, and six independent elements of a shared symmetric covariance matrix), PCA requires six loading coefficients for two components, and StandardScaler stores at least three means and three standard deviations. Effective memory use, however, also includes auxiliary matrices, buffers, numerical precision, code, libraries, and logging mechanisms. Embedded feasibility must therefore be demonstrated through direct measurements of memory, latency, and energy on the target platform.

6.2. Timing Requirements and Decision Logic

Real-time operation cannot yet be claimed because the original MPU6050 sampling frequency was not recovered. A future implementation should measure, for each sample, the time required for acquisition, signal-quality control, standardization, PCA projection, reconstruction-error computation, GMM evaluation, and consensus decision, and compare total latency with the actual sampling period. Latency percentiles, rather than averages alone, should also be reported because a safety-related system requires predictable timing behavior.

The current study produces decisions at the record level, whereas a vehicular application must operate at the event level. Alert logic should therefore define the temporal window, persistence criterion, minimum number of consecutive samples, lockout period, and event-termination rules. These mechanisms should not be calibrated on the final test set because doing so would artificially alter false-alarm and miss rates. Validation must clearly separate calibration, testing, and event-level analysis.

6.3. Sensor Calibration and Vehicle-Specific Adaptation

Sensor calibration must therefore be an explicit component of the system. Before activating the detector, the MPU6050 orientation and mounting should be documented, units and range verified, stationary bias estimated, gravity and inclination handled, and a representative period of normal driving recorded. StandardScaler means and standard deviations, the PCA basis, decision percentiles, and GMM parameters should remain version-controlled; any recalibration for a different vehicle or mounting configuration constitutes a new domain and requires independent validation [16].

6.4. Proposed Operational Architecture

Functionally, the proposed architecture comprises MPU6050 acquisition, signal-integrity control, standardization using calibrated parameters, PCA projection and reconstruction-error computation, GMM log-likelihood evaluation, the convergent decision rule, temporal event aggregation, and alert logging or transmission. The architecture also incorporates two cross-cutting mechanisms, decision traceability and fail-safe handling, to ensure that each detected event can be reconstructed from the model configuration, scores, timestamps, and calibration parameters. Figure 6 illustrates this processing sequence and the relationship between the analytical stages and these operational mechanisms.

The architecture shown in Figure 6 translates the analytical PCA+GMM procedure into a sequential implementation in which data quality is verified before a decision is produced. The system should therefore not force classification when signal integrity is insufficient. Explicit states should be defined

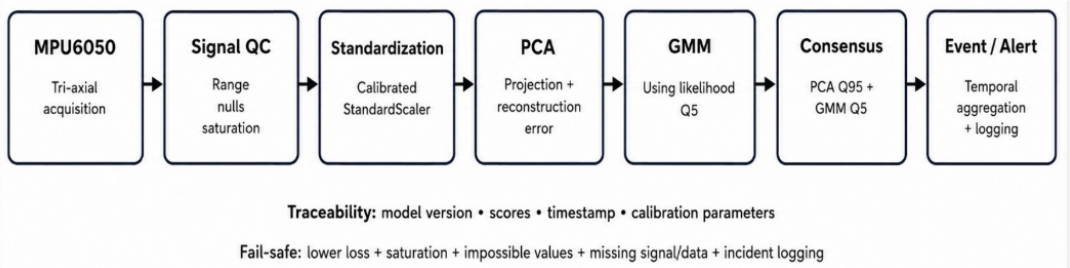

Figure 6: Proposed operational architecture for an auditable embedded implementation.

for power loss, saturation, physically impossible values, missing data, and calibration mismatch. When an integrity failure occurs, the corresponding stage is not evaluated, and the incident is logged together with its timestamp. Likewise, PCA/GMM scores, parameter versions, and calibration status should be retained to support subsequent auditing and event analysis. To move from this functional architecture to a reproducible embedded implementation, Table 6 specifies the minimum indicators that should be measured or reported for each deployment dimension.

Table 6
Minimum indicators and requirements for reproducible embedded validation.

Dimension	Current evidence	Requirement before deployment
Model	$p = 3$; PCA $k = 2$; GMM $K = 2$, tied	Freeze and version the scaler, PCA, GMM, percentile thresholds, and pipeline code.
Complexity	$O(p) + O(pk) + O(Kp^2)$ per record	Measure effective operations and compare them with the device timing budget.
Memory	Compact statistical parameterization; actual memory use not measured	Report RAM/flash use including buffers, libraries, and numerical precision.
Latency	Not measured; original sampling frequency not recovered	Measure mean, p95/p99, and worst-case latency at the actual acquisition rate.
Energy	Not measured	Measure power and energy per inference and per event on the target platform.
Temporal logic	Historical decision made at record level	Define windowing, persistence, event termination, and validate performance at event level.
Calibration	Orientation, gravity treatment, and sensor range not recovered	Document mounting, units, bias, gravity handling, and vehicle-specific recalibration protocol.
Safety and traceability	Fail-safe behavior not experimentally validated	Log failures, scores, timestamps, and software/model versions; use a not-evaluable state when signal integrity is invalid.

7. Limitations and Future Work

7.1. Data and Experimental-Traceability Limitations

The primary limitation is the loss of the original tabular file and the row-level correspondence between the external label and detector output. Although N$= 5,447$, the PCA, K-Means, DBSCAN, and GMM results, and the 131 convergent candidates were recovered, N_0, N_1, and the identity y_i of each observation are unavailable. Consequently, the true prevalence and the external metrics Precision, Recall, Specificity, F1-score, ROC-AUC, and PR-AUC cannot be reconstructed; the study is therefore restricted to inferences about structure and statistical rarity.

Acquisition metadata is also incomplete. Vehicle make, model, and year; route; pavement type;

sampling frequency and range; sensor orientation; and the gravity-compensation procedure could not be recovered reliably. These variables may alter the acceleration distribution and limit reproducibility, physical interpretation of magnitude, and transfer to other vehicles [1].

7.2. Preprocessing and Model-Configuration Limitations

The historical dataset was described as previously cleaned, but the exact criterion used to remove values considered obvious outliers is not preserved. In anomaly detection, cleaning based on extreme values can alter precisely the distribution tail that is intended to be studied. Replication should therefore start with raw data, maintain traceability for every exclusion, and conduct a preprocessing sensitivity analysis.

Configuration uncertainty is another limitation. DBSCAN showed strong sensitivity to ϵ and min_samples; GMM with $K = 2$ and tied covariance corresponds to the recovered operational configuration but cannot be demonstrated to be the global BIC optimum; and PCA (2), although retaining 75.51% of the variance, leaves 24.49% outside the retained subspace. A definitive replication should therefore document hyperparameter search, random seeds, selection criteria, and the stability of anomaly percentiles.

7.3. Validation Limitations and External Scope

The magnitude baseline is formally defined, but no quantitative run over the same 5,447 records and no traceable value of T could be recovered. There is therefore no evidence to claim fewer false positives, higher sensitivity, better F1-score, or operational superiority of PCA+GMM over a physical threshold. In addition, the lack of temporal synchronization with real events prevents sample-level evaluation from being distinguished from event-level evaluation.

External validity is also limited. The evidence comes from a single archived dataset and does not include validation across vehicles, mounting configurations, routes, speeds, or pavement conditions. Experimental measurements of latency, memory, energy, saturation, and behavior under signal loss are unavailable. Transferability and embedded feasibility should therefore be treated as engineering hypotheses that remain to be verified.

7.4. Prioritized Future-Work Agenda

The priority is to recover or collect a traceable experimental dataset that preserves raw data, timestamps, sampling frequency, orientation, mounting, units, vehicle information, route, and synchronized ground truth obtained through video, timestamped logs, or controlled testing. The pipeline should then be reproduced from the raw data while documenting cleaning, StandardScaler, PCA, hyperparameter search, percentile thresholds, and random seeds.

External validation should subsequently calculate prevalence and both sample-level and event-level metrics—Precision, Recall, Specificity, F1-score, ROC-AUC, and PR-AUC where appropriate—using independent calibration and test partitions. The magnitude baseline should be executed on exactly the same evaluation sample, together with lightweight supervised references such as logistic regression, SVM, Random Forest, and gradient-based methods, without treating the 131 anomalies as ground truth before they have been independently confirmed [14, 15].

Once internal validity has been established, the study should be extended to multiple vehicles and mounting configurations to quantify performance degradation and investigate domain-adaptation strategies [16]. Deployment studies should measure latency, memory, energy, calibration stability, robustness to potholes and hard braking, saturation, missing data, and sensor failures. Only after these stages can it be determined whether the convergent criterion provides an operational advantage over a physical threshold.

8. Conclusion

The archived MPU6050 signals exhibit statistical structure that can be characterized using unsupervised methods, but the study does not demonstrate that true impacts form a separable class. K-Means with $k = 2$ achieved Silhouette = 0.5843, whereas ARI = 0.2706 and AMI = 0.1161 showed limited correspondence with the external label. DBSCAN further demonstrated pronounced sensitivity to parameterization; with $\epsilon = 0.1$ and min_samples = 10, it produced 8 clusters and 1,127 noise points (20.69%).

The convergent GMM+PCA criterion formalizes two complementary mechanisms of rarity. GMM and PCA each selected 273 candidates by percentile construction, and their intersection yielded 131 shared observations, corresponding to 2.405% of the dataset, with Jaccard = 0.3157. This result demonstrates partial overlap between low probabilistic density and structural deviation; however, the 131 observations must be retained as robust impact candidates rather than confirmed impacts.

The methodological audit also delineates the scope of the evidence. Loss of row-level ground truth prevents the calculation of true prevalence and external classification metrics; similarly, the absence of a reproducible magnitude-baseline run prevents quantitative superiority over a physical threshold from being established. The study therefore does not attribute precision, sensitivity, false-positive reduction, or operational performance that cannot be verified directly.

The main contribution is thus an interpretable and auditable unsupervised methodology for prioritizing candidate events through PCA+GMM consensus, accompanied by explicit reproducibility and deployment criteria. Transition to an operational vehicular system requires traceable data, event-level validation, common-baseline comparisons, multi-vehicle replication, and hardware measurements of latency, memory, energy, and robustness. Under these conditions, future work can determine whether the observed statistical convergence translates into a practical advantage over conventional physics-based detectors.

Acknowledgments

The authors gratefully acknowledge Universidad Nacional del Callao (UNAC) [National University of Callao] for the academic and institutional support provided for this research.

Declaration on Generative AI

During preparation of this manuscript, generative AI tools were used solely for language editing, formatting adaptation, and consistency checking. The authors reviewed and edited the content as appropriate and take full responsibility for the scientific claims, data interpretation, references, and final version of the manuscript.

References

[1] T. Seel, M. Kok, R. S. McGinnis, Inertial sensors—applications and challenges in a nutshell, Sensors 20 (2020) 6221.

[2] H. Hozhabr Pour, F. Li, L. Wegmeth, C. Trense, R. Doniec, M. Grzegorzek, R. Wismüller, A machine learning framework for automated accident detection based on multimodal sensors in cars, Sensors 22 (2022) 3634.

[3] M. A. Onsu, M. Simsek, M. Fobert, B. Kantarci, Intelligent multi-sensor fusion and anomaly detection in vehicles via deep learning, Internet of Things 31 (2025) 101561.

[4] A. J. Mahariba, A. Uthra, G. B. Rajan, An efficient automatic accident detection system using inertial measurement through machine learning techniques for powered two wheelers, Expert Systems with Applications 192 (2022) 116389.

[5] Q. Cheng, H. Li, Y. Yang, J. Ling, X. Huang, Towards efficient risky driving detection: A benchmark and a semi-supervised model, Sensors 24 (2024) 1386.
[6] M. A. Belay, S. S. Blakseth, A. Rasheed, P. Salvo Rossi, Unsupervised anomaly detection for iot-based multivariate time series: Existing solutions, performance analysis and future directions, Sensors 23 (2023) 2844.
[7] N. Mejri, L. Lopez-Fuentes, K. Roy, P. Chernakov, E. Ghorbel, D. Aouada, Unsupervised anomaly detection in time-series: An extensive evaluation and analysis of state-of-the-art methods, Expert Systems with Applications 256 (2024) 124922.
[8] L. Correia, J.-C. Goos, P. Klein, T. Bäck, A. V. Kononova, Online model-based anomaly detection in multivariate time series: Taxonomy, survey, research challenges and future directions, Engineering Applications of Artificial Intelligence 138 (2024) 109323.
[9] I. T. Jolliffe, J. Cadima, Principal component analysis: a review and recent developments, Philosophical transactions. Series A, Mathematical, physical, and engineering sciences 374 (2016) 20150202.
[10] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al., A density-based algorithm for discovering clusters in large spatial databases with noise, in: kdd, volume 96, 1996, pp. 226–231.
[11] C. M. Bishop, N. M. Nasrabadi, Pattern recognition and machine learning, volume 4, Springer, 2006.
[12] P. J. Rousseeuw, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, Journal of computational and applied mathematics 20 (1987) 53–65.
[13] D. L. Davies, D. W. Bouldin, A cluster separation measure, IEEE transactions on pattern analysis and machine intelligence (1979) 224–227.
[14] A. Géron, Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow, " O'Reilly Media, Inc.", 2022.
[15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., Scikit-learn: Machine learning in python, the Journal of machine Learning research 12 (2011) 2825–2830.
[16] S. J. Pan, Q. Yang, A survey on transfer learning, IEEE Transactions on knowledge and data engineering 22 (2009) 1345–1359.