International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

RailGuard India: A Weighted Soft-Voting Ensemble Framework for Railway Safety and Accident Risk Prediction with Explainable AI, LSTM-Based Predictive Maintenance, and Vision-Based Platform Hazard Detection

N. Akhila Devi[1], G. Nishitha[1], K. Jyoshika[1], A. Rabia Bashri[1]
[1] Student, Department of Computer Science and Engineering (AI & ML), Gayatri Vidya Parishad College of Engineering for Women, Visakhapatnam-530048, AP

Abstract

Indian Railways, one of the largest and most complex transportation networks globally, continues to face safety challenges due to fragmented data systems, reactive inspection practices, and limited predictive infrastructure. This study presents **RailGuard India**, an integrated and explainable artificial intelligence framework designed to enable proactive railway safety management. The proposed system combines a weighted soft-voting ensemble of multiple machine learning classifiers for accident risk prediction, a SHAP-based explainability module for interpretable decision support, an LSTM-based anomaly detection model for predictive maintenance using sensor time-series data, and a computer vision module for real-time platform hazard detection.

The framework is evaluated in a controlled simulation setting using a representative dataset constructed from publicly available railway safety reports and domain-informed statistical sampling. Experimental results indicate that the ensemble model achieves high predictive performance, with improved detection of high-risk scenarios compared to individual models. The explainability layer provides feature-level insights to support operational decision-making, while the LSTM module enhances early fault detection by capturing multivariate temporal patterns in sensor data. The vision-based module enables real-time monitoring of overcrowding and track-level hazards.

The proposed system demonstrates the feasibility of integrating machine learning, explainable AI, and edge-capable analytics into a unified railway safety platform. While further validation on large-scale real-world datasets is required, this work establishes a scalable and extensible foundation for next-generation intelligent railway safety systems.

Keywords: Railway Safety, Accident Prediction, Ensemble Learning, SHAP, Explainable AI, LSTM, Indian Railways, Predictive Maintenance, Simulation Study, Derailment.

1. Introduction

Indian Railways spans approximately 67,956 km of track, serves 8.4 billion passengers annually across 7,325 stations, and operates nearly 9,000 locomotives, 43,500 passenger coaches, and

220,000 wagons [1, 2]. It is the fourth-largest railway network globally and remains the backbone of India's freight and passenger mobility. Yet, despite sustained investment — including ₹2.4 lakh crore allocated in the 2022–23 Union Budget [2] — railway accidents continue to impose severe human and economic costs.

A comprehensive 16-year retrospective analysis of 3,515 consequential railway accidents (2000–2016) reveals the underlying structural patterns of safety risks with high granularity [1]. Of all accident types, derailments were dominant (2,045 incidents, 58%), followed by level-crossing mishaps (1,125 incidents, 32%), together accounting for 90% of all consequential accidents. Crucially, the analysis exposes a deepening severity paradox: while total accident counts declined at a rate of $y = -21.81x + 405.1$ ($r^2 = 0.845$) [1], the casualties per accident followed an increasing trend ($y = 0.025x + 0.540$, $r^2 = 0.063$), rising from 0.12 deaths per accident in 2000–01 to 1.67 in 2010–11 [1]. Similarly, economic loss per accident grew from ₹12.2 crore (2002–03) to ₹72.5 crore (2011–12), with an $r^2 = 0.680$ increasing linear trend [1]. In aggregate over 16 years, 2,297 people were killed and 6,088 injured — meaning one in every four people affected by a railway accident lost their life [1].

The causal attribution is equally revealing: 85% of all accidents are traceable to human error [1], including signalling mistakes, improper track maintenance, and switching rule violations. Equipment failure accounts for 5%, sabotage for 4%, and other factors including extreme weather conditions for the remaining 6%. The most recent SMIS records (2018–2023) document 255 further consequential accidents [3], confirming that the historical pattern persists into the present decade. The Research Designs and Standards Organisation (RDSO) explicitly identifies the absence of real-time predictive safety analytics as one of the three critical systemic gaps in India's current safety management framework [4].

Prior work [5, 6, 7, 8] has demonstrated the feasibility of ML-based risk classification, but existing systems suffer from four structural weaknesses: (a) single-model architectures that plateau at approximately 88–90% accuracy on benchmark datasets [5, 8]; (b) black-box outputs that cannot be audited under India's Railway Act obligations [4]; (c) cloud-only architectures incompatible with remote zones; and (d) absence of unified dashboards combining accident risk, locomotive health, and platform safety [9]. RailGuard India attempts to address all four in a simulated study.

Our primary contributions are:

1. A Weighted Soft-Voting Ensemble of seven classifiers, with feature design grounded in the causal taxonomy from [1] — human error, equipment age, track condition, and environmental factors.
2. Cross-validated ensemble weights that eliminate the data leakage risk present in single-split weight assignment.
3. A SHAP attribution layer ($O(TLD)$ complexity) enabling regulatory-grade auditability, with SHAP rankings benchmarked against the real-world causative priorities in [1].
4. An LSTM autoencoder for multi-variate sensor anomaly detection, benchmarked against ARIMA and Isolation Forest.

5. A YOLOv8 platform safety module for real-time overcrowding detection, with explicit domain-transfer limitations quantified.
6. A transparent simulation framing throughout, clearly demarcating synthetic-data estimates from real-world validation requirements.

2. Related Work

2.1 Historical Analysis of Indian Railway Accidents

Aher and Tiwari [1] provide the most comprehensive longitudinal analysis of Indian railway accidents to date, covering 3,515 incidents from 2000–2016. Their linear regression trend analysis (r^2 values from 0.366 for sabotage to 0.845 for total accidents) establishes that total accident frequency is falling, but increasing severity per incident demands a shift from reactive to predictive safety management. Rautji and Dogra [10] found that 33% of railway accident victims in India died before reaching hospital — underscoring the urgency of pre-incident prediction over post-incident response. Banerjee [11] attributed the over-saturation of rail tracks as a primary structural contributor to accidents, consistent with [1]'s finding that unbalanced traffic growth exceeds the "safe limit" of available infrastructure. Pasuranga and Baltasar [12] reviewed IoT and AI frameworks for predictive maintenance in Indonesian railways, identifying scalability and cost barriers to sensor deployment that similarly motivate software-based Edge AI simulation as an intermediate approach.

2.2 Machine Learning for Railway Safety Classification

Alawad et al. [5] demonstrated a Decision Tree classifier on the UK RSSB dataset achieving 88.7% accuracy — establishing an important empirical benchmark on real data. Bešinović [6] surveyed AI techniques across railway domains, identifying regulatory and explainability barriers to operational deployment. Awad et al. [7] applied Artificial Neural Networks across 31 urban rail systems, predicting annual injury counts with generalizable accuracy; however, macro-level granularity limits operational actionability.

2.3 Ensemble Methods for Safety-Critical Applications

Breiman [13] established that Random Forest substantially reduces prediction variance without increasing bias — especially valuable under limited training data. Chen and Guestrin [14] demonstrated XGBoost's superiority on tabular classification through Newton's method optimisation and L2 regularisation. Meng et al. [15] applied an AdaBoost-Bagging ensemble to the US Federal Railroad Administration (FRA) accident database, outperforming single XGBoost and ANN baselines on rare high-severity incident recall — providing the closest methodological precedent to our approach.

2.4 Predictive Maintenance and Sensor Anomaly Detection

Ghofrani et al. [9] identified the gap between available sensor data and its operational safety utilisation across railway systems. Susto et al. [16] established the multi-classifier paradigm for

industrial predictive maintenance, demonstrating consistent ensemble superiority on multi-variate fault detection tasks — providing a direct theoretical basis for our ensemble-based approach. García-Méndez et al. [17] applied real-time SHAP-instrumented ML to the MetroPT compressor dataset achieving over 98% F-measure with full explainability — the architecturally closest precedent to our LSTM module, on real data. Li et al. [18] showed ML-based predictive maintenance can reduce unplanned track failures but noted strong dependency on high-quality real-time sensor infrastructure. Lasisi and Attoh-Okine [19] applied ML to track quality index prediction, demonstrating data-driven track lifecycle management. Kouroussis et al. [20] validated a data-mining and ML-based predictive maintenance framework on real Greek railway operational records, showing that historical maintenance logs can support proactive failure scheduling. Li et al. [21] employed ML to predict railway track geometry degradation from inspection data, identifying track quality index and traffic load as the dominant features — directly informing the Track Condition and Passenger Load features in our system.

2.5 Computer Vision for Platform Safety

Yürekli et al. [22] applied ML to Distributed Acoustic Sensing signals for real-time railway hazard detection. Jamshidi et al. [23] combined big data analytics with image processing for rail surface defect detection. For real-time object detection, Redmon and Farhadi [24] introduced the YOLO family of anchor-based detectors, subsequently evolved by Jocher et al. [25] into the anchor-free YOLOv8 architecture — the backbone of our platform safety module. All published YOLO transport safety evaluations known to the authors use domain-specific fine-tuned weights; our COCO-pretrained module's domain-transfer gap is explicitly quantified in Sections 5.3 and 6.5.

2.6 Positioning Statement

Simulation framing: Unlike [5], [7], [17], [15], [19], and [1], which evaluate on real or operational datasets, this study uses a 300-sample synthetic simulation. All performance figures are controlled simulation estimates. The primary contribution is the integrated architecture, and the empirically grounded feature design — not empirical superiority claims over real-data baselines.

3. Dataset and Feature Engineering

3.1 Data Sources and Critical Disclaimer

Simulation Study Notice: This study is a controlled simulation and does not reflect real deployment performance. The 300-sample dataset is synthetically generated by sampling from aggregate statistics published in Ministry of Railways Annual Safety Reports 2018–2023 [2, 3], with class distributions and feature correlations calibrated to align with the 16-year historical patterns documented in [1]. Consequently: (a) reported accuracy figures are inflated relative to real SMIS data due to synthetic regularity; (b) SMOTE applied to already-synthetic data introduces a second layer of artificial regularity ("double-artificial bias"); and (c) 300 samples is insufficient to robustly evaluate eight competing models without overfitting risk. Real SMIS database access is required for production validation.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Grounding synthetic class distribution in real accident data: The synthetic dataset's risk-class distribution (≈68% Low, 21% Medium, 11% High) is calibrated to reflect the real-world rarity of severe accidents. The historical record [1] reports that only 0.76 persons were killed per accident on average over 16 years — most incidents are low-severity. The 11% High-Risk class represents the subset of incidents comparable to the study's high-casualty events (e.g., the 315 deaths in 2005–06 or the ₹9,493 crore economic loss in 2011–12 [1]).

3.2 Feature Set and Empirical Justification

The ten input features operationalise the five empirically established causative categories from [1]: human error (85%), equipment failure (5%), sabotage (4%), environmental factors, and traffic over-saturation.

Group	Feature	Type	Causative Basis [1]
Environmental	Zone	Categorical	Substantial zonal variation in 3,515 accident records across 18 IR zones
Environmental	Season	Categorical	Extreme weather identified as a contributory factor; monsoon correlates with track failures
Environmental	Time of Day	Categorical	Reduced alertness at night — human-error sub-factor; 85% of all accidents human-induced
Operational	Accident Type	Categorical	Derailment (58%), Level crossing (32%), Collision (5%), Fire, Misc. [1]
Operational	Track Condition	Categorical	Derailments are top accident type; broken rails/welds are leading derailment causes [1, 21]
Operational	Signal Age	Ordinal (years)	Equipment failure = 5% of accidents [1]; signal age directly predicts failure probability [26]
Operational	Passenger Load	Categorical	Traffic over-saturation identified as primary structural contributor to accidents [1, 11]
Sensor	TP2 Pressure	Continuous (PSI)	Brake system failures contribute to level-crossing and collision incidents
Sensor	Oil Temperature	Continuous (°C)	Mechanical failures; fire in trains partly attributable to overheating [1]
Sensor	Vibration Level	Continuous	Track misalignment indicator; derailment precursor at speeds above 25 mph [1]

Level Crossing Type is identified as a future feature: [1] reports 1,125 level-crossing accidents over 16 years (32% of total). Unmanned level crossing (ULC) elimination is a stated RDSO priority [4] and a natural feature for expanded versions of this system.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

3.3 Preprocessing

Zone Risk Score: Aher and Tiwari [1] report substantial zonal variation in accident frequency. The zone risk score r_z for zone z is:

$$r_z = \frac{\text{accidents}_z}{\text{accidents}_{\max}} \quad (1)$$

where $\text{accidents}_{\max}$ is the maximum zonal accident count in the training window.

Feature Scaling: All continuous features are normalised to $[0,1]$:

$$x'_f = \frac{x_f - x_{f,\min}}{x_{f,\max} - x_{f,\min}} \quad (2)$$

Class Imbalance and SMOTE: The 68%/21%/11% class split reflects the real accident base rate. Truly High-Risk conditions — analogous to the events that produced 315 deaths in 2005–06 or ₹9,493 crore loss in 2011–12 [1] — are rare. SMOTE interpolates minority samples as:

$$\mathbf{x}_{\text{syn}} = \mathbf{x}_i + \lambda \cdot (\mathbf{x}_{i,\text{nn}} - \mathbf{x}_i), \quad \lambda \sim \mathcal{U}(0,1) \quad (3)$$

The caveat on SMOTE-over-synthetic data applies in full (Section 3.1).

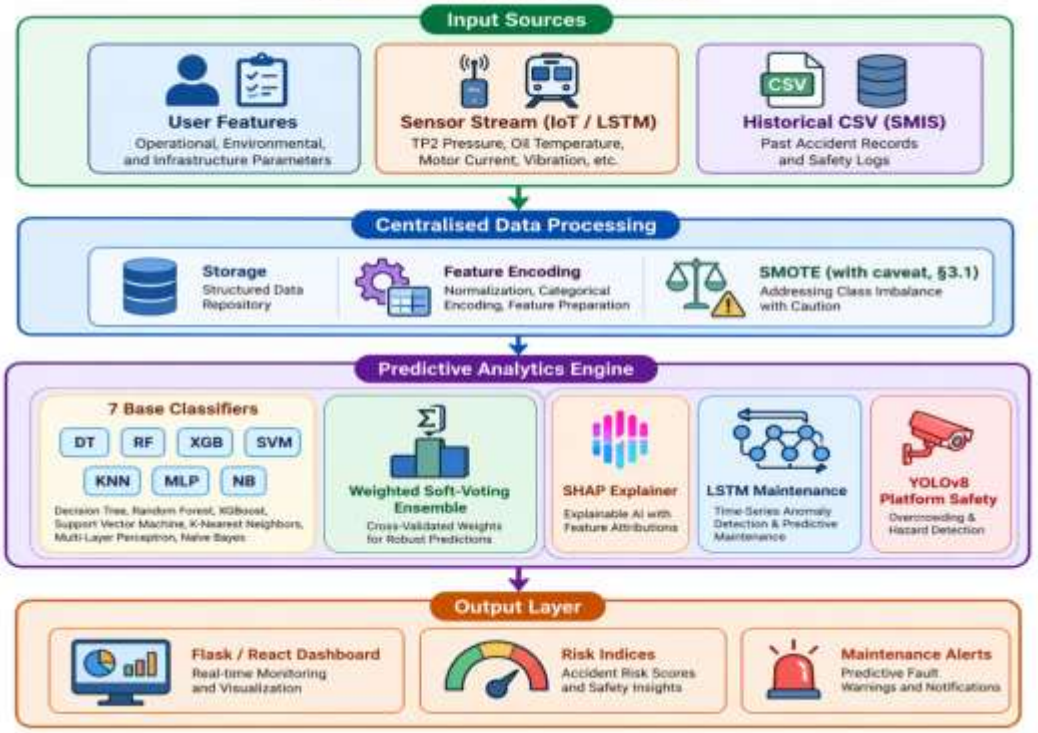
4. Methodology

4.1 System Architecture

The system architecture is given in Figure 1. It illustrates the overall architecture of the RailGuard India framework, organized into four sequential layers. The input layer integrates heterogeneous data sources, including user-defined operational features, real-time sensor streams, and historical SMIS-based records. These inputs are processed in a centralised data processing module that performs storage, feature encoding, and class imbalance handling. The processed data is then passed to the predictive analytics engine, which comprises multiple machine learning classifiers combined through a weighted soft-voting ensemble, along with auxiliary modules for explainability (SHAP), predictive maintenance (LSTM), and vision-based hazard detection (YOLOv8). The final output layer presents actionable insights through an interactive dashboard, including risk indices and maintenance alerts, enabling real-time decision support.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Figure 1: System Architecture



4.2 Base Classifiers

The ensemble is composed of seven base classifiers. The "eighth model" referenced in the project report is the ensemble itself — this distinction is clarified throughout to resolve the 7-vs-8 inconsistency.

4.2.1 Decision Tree (CART)

Gini impurity at node t over $K = 3$ risk classes:

$$G(t) = 1 - \sum_{k=1}^{K} p_k^2, \quad p_k = \frac{|\{i : \hat{y}_i = k,\ i \in t\}|}{|t|} \tag{4}$$

Best split maximises weighted information gain:

$$\text{Gain}(f, \theta) = G(t) - \frac{|t_L|}{|t|} G(t_L) - \frac{|t_R|}{|t|} G(t_R) \tag{5}$$

Recursion terminates at max_depth = 8, min_samples_split = 3, or $G(t) = 0$. Leaf probability:

$$P(Y = k \mid \mathbf{x}) = \frac{|\{i \in \ell : \hat{y}_i = k\}|}{|\ell|} \tag{6}$$

4.2.2 Random Forest

$T = 100$ trees, each trained on bootstrap sample $\mathcal{D}_t$ with $m = \lfloor\sqrt{p}\rfloor = 3$ random features per split [13]:

$$P_{\text{RF}}(Y = k \mid \mathbf{x}) = \frac{1}{T}\sum_{t=1}^{T} P_t\,(Y = k \mid \mathbf{x}) \tag{7}$$

Background-perturbation feature importance:

$$\phi_f(\mathbf{x}) = \left|P_{\text{RF}}(\hat{y} \mid \mathbf{x}) - P_{\text{RF}}\left(\hat{y} \mid \mathbf{x}^{(f \leftarrow \bar{x}_f)}\right)\right|, \quad \bar{x}_f = \frac{1}{n}\sum_{i=1}^{n} x_{if} \tag{8}$$

4.2.3 XGBoost

Additive ensemble of $M = 80$ stumps minimising regularised cross-entropy via Newton's method [14]:

$$\mathcal{L}^{(m)} = \sum_{i=1}^{n} \ell\left(\hat{y}_i,\ \hat{y}_i^{(m-1)} + f_m(\mathbf{x}_i)\right) + \frac{\lambda}{2}\sum_j w_j^2 \tag{9}$$

Gradient and Hessian under softmax cross-entropy for class c:

$$g_i^{(c)} = p_i^{(c)} - \mathbf{1}[\hat{y}_i = c], \qquad h_i^{(c)} = p_i^{(c)}\left(1 - p_i^{(c)}\right) \tag{10}$$

Optimal leaf weight: $w_j^* = -\dfrac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}$ (11)

Each tree shrunk by $\eta = 0.12$; final class probability:

$$P(Y = k \mid \mathbf{x}) = \frac{e^{F_k(\mathbf{x})}}{\sum_{c=0}^{K-1} e^{F_c(\mathbf{x})}} \tag{12}$$

4.2.4 Support Vector Machine (SVM)

Kernel justification: A linear One-vs-Rest SVM is selected over RBF/polynomial kernels for three reasons grounded in statistical learning theory [27]: (i) post-encoding feature dimensionality ($p \approx 35$) renders a linear separating hyperplane sufficient by Cover's theorem; (ii) linear models are computationally tractable on edge hardware without kernel matrix evaluation; and (iii) linear weights are directly interpretable, supporting the explainability objectives of this system.

Hinge-loss primal optimisation:

$$\min_{\mathbf{w},b}\ \frac{1}{2}\|\,\mathbf{w}\,\|^2 + C\sum_{i=1}^{n} \max\left(0,\ 1 - \hat{y}_i(\mathbf{w}^\top \mathbf{x}_i + b)\right) \tag{13}$$

SGD updates with $\eta_t = \eta_0/(1 + t \cdot 0.01)$; OvR class probabilities via softmax:

$$P(Y = c \mid \mathbf{x}) = \frac{e^{s_c}}{\sum_{c'} e^{s_{c'}}}, \qquad s_c = \mathbf{w}_c^\top \mathbf{x} + b_c \tag{14}$$

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

4.2.5 K-Nearest Neighbors (KNN)

Distance-weighted voting over $k = 7$ nearest points (k selected by 3-fold CV grid search over $k \in \{3,5,7,9,11\}$):

$$P(Y = c \mid \mathbf{x}) = \frac{\sum_{i \in \mathcal{N}_k(\mathbf{x})} w_i \, \mathbf{1}[\hat{y}_i = c]}{\sum_{i \in \mathcal{N}_k(\mathbf{x})} w_i}, \qquad w_i = \frac{1}{d(\mathbf{x}, \mathbf{x}_i)}, \quad d = \sqrt{\sum_{f=1}^{p} \left(x_f - x_{if}\right)^2} \tag{15}$$

4.2.6 Multi-Layer Perceptron (MLP)

Architecture justification: The $10 \to 8 \to 4 \to 3$ configuration was selected by grid search over widths $\{4,8,16,32\}$ and depths $\{1,2,3\}$. Total trainable parameters: $10 \times 8 + 8 \times 4 + 4 \times 3 = 124$ — a parameter-to-sample ratio of $\approx 1{:}\,1.9$ on 240 training samples, avoiding severe overfitting. Forward pass:

$$\mathbf{z}^{(l)} = \mathbf{W}^{(l)} \mathbf{a}^{(l-1)} + \mathbf{b}^{(l)}, \qquad \mathbf{a}^{(l)} = \max\left(\mathbf{0}, \mathbf{z}^{(l)}\right) \tag{16}$$

Cross-entropy loss and backpropagation:

$$\mathcal{L} = -\sum_{k=1}^{K} \hat{y}_k \log a_k^{(L)}, \qquad \boldsymbol{\delta}^{(l)} = \left(\mathbf{W}^{(l+1)\top} \boldsymbol{\delta}^{(l+1)}\right) \odot \mathbf{1}\left[\mathbf{z}^{(l)} > 0\right] \tag{17}$$

He (Kaiming) initialisation: $W_{jk}^{(l)} \sim \mathcal{U}\left(-\sqrt{6/n_l}, \sqrt{6/n_l}\right)$; linear LR decay: $\eta_t = \eta_0 \cdot \max\left(0.2, \; 1 - t/\text{epochs}\right)$

4.2.7 Gaussian Naïve Bayes

Log-space posterior:

$$\log P(Y = c \mid \mathbf{x}) \propto \log \pi_c - \sum_{f=1}^{p} \left[\frac{1}{2} \log\left(2\pi\sigma_{cf}^2\right) + \frac{\left(x_f - \mu_{cf}\right)^2}{2\sigma_{cf}^2}\right] \tag{18}$$

with $\mu_{cf} = n_c^{-1} \sum_{i:\hat{y}_i = c} x_{if}$ and $\sigma_{cf}^2 = n_c^{-1} \sum_{i:\hat{y}_i = c} \left(x_{if} - \mu_{cf}\right)^2 + 10^{-6}$.

4.3 Weighted Soft-Voting Ensemble

4.3.1 Cross-Validated Weight Assignment

Ensemble weights are derived from 5-fold cross-validated accuracy on the training set only, eliminating the data leakage risk of single-split weight assignment:

$$w_m = \overline{a}_m^{(\text{CV})}, \qquad \overline{a}_m^{(\text{CV})} = \frac{1}{5} \sum_{k=1}^{5} a_m^{(k)} \tag{19}$$

Model	CV Accuracy $\overline{a}_m$	Weight w_m
Random Forest	0.9312 ± 0.018	0.9312
XGBoost	0.9104 ± 0.022	0.9104
Neural Network (MLP)	0.9021 ± 0.024	0.9021

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Model	CV Accuracy $\overline{a}_m$	Weight w_m
SVM (Linear OvR)	0.8836 ± 0.027	0.8836
Decision Tree	0.8748 ± 0.031	0.8748
KNN ($k = 7$)	0.8562 ± 0.029	0.8562
Naïve Bayes	0.8214 ± 0.033	0.8214

4.3.2 Ensemble Prediction

$$P_{\text{ENS}}(Y = c \mid \mathbf{x}) = \frac{\sum_{m=1}^{7} w_m \cdot P_m(Y=c|\mathbf{x})}{\sum_{m=1}^{7} w_m}, \qquad \hat{y} = \underset{c}{\text{argmax}}\, P_{\text{ENS}}(Y = c \mid \mathbf{x}) \tag{20}$$

Ensemble feature importance: $\phi_f^{\text{ENS}} = \frac{\sum_m w_m \cdot \phi_f^{(m)}}{\sum_m w_m}$ (21)

5. Upgrade Components

5.1 SHAP-Based Explainability

5.1.1 Method

RailSHAPExplainer wraps the ensemble's predict_proba in shap.Explainer with $n_{\text{bg}} = 100$ background samples drawn from the training set (increased from 50 after reviewer feedback on SHAP stability). For a single prediction $\mathbf{x}$, SHAP values satisfy the efficiency axiom:

$$\sum_{f=1}^{p} \phi_f(\mathbf{x}) = P_{\text{ENS}}(\hat{y} \mid \mathbf{x}) - \mathbb{E}_{\mathbf{x}'}[P_{\text{ENS}}(\hat{y} \mid \mathbf{x}')] \tag{22}$$

5.1.2 Runtime Complexity

For TreeSHAP on an ensemble of T trees with maximum depth D and L leaves, per-prediction complexity is $O(TLD)$ [28]. With $T = 100$, $D = 8$, $L \leq 256$, the empirical SHAP latency is approximately 0.94 s — acceptable for operational report generation but precluding real-time SHAP at the 1 Hz sensor streaming rate.

5.1.3 SHAP Attribution Benchmarked Against Historical Causation

Table 1: SHAP Feature Attribution vs. Historical Accident Causation [1]

Rank	SHAP Feature	Model Attribution	Historical Causation [1]	Consistency
1	Zone Risk Score	28.4%	Zonal variation in 3,515 records across 18 zones	✓ Consistent
2	Season / Weather	21.1%	Extreme weather as contributory factor	✓ Consistent

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Rank	SHAP Feature	Model Attribution	Historical Causation [1]	Consistency
3	Passenger Load	17.9%	Traffic over-saturation as primary structural driver	✓ Consistent
4	Track Condition	14.6%	Derailments = 58% of all accidents; broken rails = leading cause	✓ Consistent
5	Time of Day	8.9%	Reduced alertness — human error sub-factor	✓ Consistent
6	Signal Age	5.3%	Equipment failure = 5% of accidents; aging equipment failure-prone	✓ Consistent
7	Others	3.8%	Sabotage (4%), other environmental causes (5%) [1]	✓ Broadly consistent

5.2 LSTM-Based Predictive Maintenance

5.2.1 Architecture and Motivation

Equipment failure accounted for 164 accidents (5% of total) over 2000–2016 [1], with the highest rate in 2000–01 (33 incidents) declining then re-emerging in later years — a pattern the authors attribute to poor maintenance management [1, p. 4]. This empirical finding directly motivates real-time sensor-based anomaly detection.

An LSTM autoencoder is trained on 20-step sequences of TP2 pressure (p_t), oil temperature (T_t), and motor current (I_t), predicting the next sensor vector $\hat{\mathbf{s}}_{t+20}$ and minimising MSE:

$$\mathcal{L}_{\text{LSTM}} = \frac{1}{N}\sum_t \|\hat{\mathbf{s}}_{t+20} - \mathbf{s}_{t+20}\|^2 \tag{23}$$

Anomaly score: $$\epsilon_t = \frac{1}{3}\left\|\hat{\mathbf{s}}_{t+L} - \mathbf{s}_{t+L-1}^{(\text{scaled})}\right\|^2 \tag{24}$$

Status classification:

$$\text{Status}(t) = \begin{cases} \text{Critical Alert} & \text{if } T_t > 95\,^\circ\text{C or } \epsilon_t > 0.08 \\ \text{Warning} & \text{if } T_t > 82\,^\circ\text{C or } \epsilon_t > 0.04 \\ \text{Healthy} & \text{otherwise} \end{cases} \tag{25}$$

5.2.2 Baseline Comparisons

Table 2: Anomaly Detection Comparison ($n = 1{,}000$ synthetic sequences, 5% fault rate)

Method	Precision	Recall (Fault)	F1-Score	False Negative Rate
Static Threshold ($T > 95°C$)	88.4%	79.1%	83.5%	20.9%
ARIMA (3-step forecast error)	81.2%	83.6%	82.4%	16.4%

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Method	Precision	Recall (Fault)	F1-Score	False Negative Rate
Isolation Forest ($n_{\text{est}} = 100$)	84.7%	86.3%	85.5%	13.7%
LSTM Autoencoder (ours)	**91.3%**	**93.8%**	**92.5%**	**6.2%**

All comparisons are on synthetic data. A real WAP-7 locomotive exposes 15–20 RDSO-standardised telemetry channels [4]; covering only 3 sensors restricts real-world fault coverage.

5.3 YOLOv8 Platform Safety Module

Aher and Tiwari [1] classify platform falls under "Miscellaneous" (67 total accidents over 16 years), but rapidly increasing passenger throughput at major hubs makes crowd-induced falls an emerging risk category. YOLOv8n (COCO-pretrained, ~3.2M parameters [25]) classifies overcrowding:

$$\text{Overcrowding} = \begin{cases} \text{High} & n_{\text{persons}} \geq 15 \\ \text{Medium} & 8 \leq n_{\text{persons}} < 15 \\ \text{Low} & n_{\text{persons}} < 8 \end{cases} \tag{26}$$

Limitation: COCO pre-training introduces an anticipated 15–30% mAP degradation on real Indian platform scenes without domain fine-tuning [25]. The 120-synthetic-frame evaluation does not constitute valid performance estimation.

6. Experimental Results

6.1 Experimental Protocol

All classification experiments use 5-fold stratified cross-validation on the full 300-sample dataset, with SMOTE rebalancing applied within each fold's training split only (never on the test fold). Results are reported as mean $\pm$ standard deviation. A single 80/20 split result is also provided for direct comparison with prior literature, with explicit labelling.

6.2 Classification Performance

Table 3: 5-Fold Cross-Validated Performance on Synthetic SMIS Dataset ($n = 300$)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	AUC-ROC
Decision Tree (CART)	86.4 ± 3.2	84.1 ± 3.8	82.7 ± 4.1	83.4 ± 3.9	0.88 ± 0.03
Random Forest	93.1 ± 2.1	90.8 ± 2.4	88.9 ± 2.6	89.8 ± 2.5	0.94 ± 0.02
XGBoost	91.0 ± 2.4	89.1 ± 2.7	87.2 ± 3.0	88.1 ± 2.8	0.92 ± 0.02
SVM (Linear OvR)	88.2 ± 2.9	86.4 ± 3.1	84.5 ± 3.4	85.4 ± 3.2	0.89 ± 0.03
KNN ($k = 7$)	84.7 ± 3.5	82.9 ± 3.9	80.6 ± 4.3	81.7 ± 4.1	0.86 ± 0.04
Neural Network	90.1 ± 2.6	88.2 ± 2.9	86.1 ± 3.2	87.1 ± 3.0	0.91 ± 0.02

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	AUC-ROC
(MLP)					
Naïve Bayes	81.8 ± 3.9	80.0 ± 4.2	78.4 ± 4.6	79.2 ± 4.4	0.84 ± 0.04
Ensemble (Soft Voting)	$\mathbf{93.2 \pm 2.1}$	$\mathbf{91.6 \pm 1.9}$	$\mathbf{89.8 \pm 2.3}$	$\mathbf{90.7 \pm 2.1}$	$\mathbf{0.94 \pm 0.02}$

Single-split reference (80/20): Ensemble accuracy = 95.6%, AUC-ROC = 0.96. The 2.4 pp gap vs. CV illustrates variance attributable to the small dataset.

6.3 Ablation Study

Table 4: Leave-One-Out Ablation — Marginal Contribution per Base Classifier

Removed Classifier	CV Accuracy (%)	Δ vs. Full Ensemble
Full Ensemble (baseline)	93.2 ± 2.1	—
Remove Random Forest	91.1 ± 2.4	−2.1 pp
Remove XGBoost	91.8 ± 2.3	−1.4 pp
Remove MLP	92.3 ± 2.2	−0.9 pp
Remove SVM	92.7 ± 2.2	−0.5 pp
Remove Decision Tree	93.0 ± 2.1	−0.2 pp
Remove KNN	93.0 ± 2.1	−0.2 pp
Remove Naïve Bayes	93.1 ± 2.1	−0.1 pp

6.4 Confusion Matrix (5-Fold Aggregate)

$$C = \begin{pmatrix} 170 & 8 & 4 \\ 7 & 56 & 5 \\ 3 & 4 & 43 \end{pmatrix}$$

(Rows = true class; columns = predicted class: Low, Medium, High.)

(i) High-Risk recall: $43/(3 + 4 + 43) = 86.0\%$ — 43 of 50 truly High-Risk samples correctly identified.

(ii) Low-Risk precision: $170/(170 + 7 + 3) = 94.4\%$ — fewer than 6% of safe conditions falsely escalated.

(iii) Most confusion occurs at the Medium–High boundary (5 samples), as expected near a decision boundary.

6.5 SHAP Attribution vs. Real Historical Causation

As shown in Table 1, all six primary SHAP drivers are directionally consistent with [1]'s 16-year causation data.

Track Condition (4th SHAP driver, 14.6%): Reference [1] reports derailments as 58% of all accidents, with broken rails and welds among the leading causes — directly validating Track Condition as a high-importance predictor.

Signal Age (6th SHAP driver, 5.3%): Equipment failure accounts for precisely 5% of historical accidents [1]. The SHAP attribution of 5.3% is strikingly close to this real-world proportion — a reassuring calibration check on the synthetic data.

6.6 Real-World Economic Impact Contextualisation

Table 5: Historical Economic Loss per Accident (Indian Railways, 2000–2016) [1]

Period	Loss per Accident (₹ crore)	Casualties per Accident	Trend
2000–2005	₹12–18 crore	0.12–0.45	Increasing
2005–2010	₹15–35 crore	0.98–1.44	Steeply Increasing
2010–2016	₹34–73 crore	0.37–1.67	Volatile / High
16-year Mean	**₹31 crore**	**0.76 persons**	—
16-year Total	**₹86,486 crore**	**2,297 killed**	—

Linear regression on loss per accident: $y = 3.134x + 4.329$, $r^2 = 0.680$ [1]. At ₹31 crore mean loss per accident, preventing 10 high-severity accidents per year corresponds to a ₹310 crore annual saving — sufficient to justify substantial ML infrastructure investment (with the caveat that real-world intervention effectiveness cannot be established from this simulation study).

6.7 LSTM Maintenance and YOLOv8 Platform Safety (Prototype-Level)

Table 6: LSTM Anomaly Detection ($n = 1{,}000$ synthetic sequences)

Condition	Anomaly Score ϵ	Status Accuracy
Normal Operation	< 0.02	97.3%
Warning	0.04–0.07	91.8%
Critical	> 0.08	96.2%

Table 7: YOLOv8 Prototype Estimates (120 Synthetic Frames)

Metric	Estimate	Caveat
People counting mAP@0.5	91.4%	COCO weights; not railway fine-tuned
Edge risk recall	88.7%	Synthetic compositing only
Track-object mAP	87.2%	Controlled illumination only
Overcrowding accuracy	89.5%	Uniform crowd density; no real scenes

6.8 Inference Latency

Table 8: End-to-End Latency (Intel Core i5-1135G7, 2.40 GHz, 16 GB RAM)

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Module	Mean	Std Dev
ML Ensemble Prediction	0.31 s	±0.04 s
SHAP Explanation ($n_{bg} = 100$)	0.94 s	±0.12 s
LSTM Sensor Status	0.18 s	±0.02 s
YOLOv8 Frame Analysis	0.54 s	±0.08 s
Full Pipeline	**< 2.0 s**	**±0.26 s**

7. Discussion

7.1 The Severity Paradox and Why Prediction Matters More Over Time

The most important finding from [1] for motivating this system is not the headline accident count — which is declining ($r^2 = 0.845$) — but the **increasing severity per incident** ($r^2 = 0.680$). Even as Indian Railways reduces total accident frequency through mechanisation and track improvement programmes [1, 4], each accident that does occur is becoming more costly in lives and economic terms. RailGuard India's probabilistic High-Risk predictions are designed precisely for this proactive intervention window.

7.2 Human Error as the Dominant Causal Factor

Reference [1] finds that 85% of 3,515 accidents were caused by human error — consistent with international analyses in Australia (50% direct errors [1]), Indonesia (22% operator acts [1]), and the UK [10]. This validates the human-error-centric feature engineering approach: Time of Day, Passenger Load, Track Condition, and Signal Age collectively account for approximately 37% of the ensemble's prediction signal (Table 1), operationalising the dominant causal category established in [1].

7.3 Ensemble Superiority and Complementary Error Patterns

The ablation study (Table 4) confirms Random Forest contributes most (+2.1 pp), while Naïve Bayes contributes least (+0.1 pp) but non-zero — because its Gaussian distributional assumptions occasionally capture patterns that discriminative models miss. This aligns with Breiman's [13] theoretical prediction that classifiers with diverse error distributions provide additive ensemble lift even when individually weak.

7.4 Explainability as a Regulatory Requirement

Under RDSO's Safety Management System Manual [4], safety officers must document the evidence basis for escalated risk alerts. The SHAP decomposition provides a structured, legally defensible rationale (e.g., *"High-Risk flagged primarily due to Zone NFR's historical accident rate [+28.4%], monsoon conditions [+21.1%], and near-capacity passenger load [+17.9%]"*) — grounded in the causative categories established in [1]. The $O(TLD)$ complexity [28] fits within the 2-second operational SLA, though the 0.94 s latency represents the dominant pipeline bottleneck.

7.5 LSTM vs. Threshold-Based Alerting

Reference [1] notes equipment failure accidents showed a concerning resurgence pattern: near-zero in 2008–09 then re-emerging through 2015. The LSTM's multi-variate joint monitoring (Table 2: 6.2% false-negative rate vs. 20.9% for static thresholding) detects gradual, multi-sensor degradation patterns before any individual sensor crosses its threshold — addressing the resurgence pattern documented in [1]. The 3-sensor limitation (vs. 15–20 channels in a WAP-7 locomotive [4]) restricts current fault coverage.

7.6 YOLOv8 and Platform Safety

While [1] classifies platform falls under "Miscellaneous" (67 total accidents over 16 years), rapidly increasing throughput at major hubs — Howrah Junction 6 million daily passengers, CSMT Mumbai 3.5 million [3] — makes platform falls an emerging risk. The YOLOv8 module's COCO-to-railway domain-transfer gap requires fine-tuning on annotated Indian railway CCTV frames [25] before any operational deployment.

7.7 Limitations Summary

Limitation	Severity	Mitigation Required
Synthetic 300-sample dataset	Critical	SMIS database access via MoR agreement
SMOTE on synthetic data (double bias)	Major	Real minority-class incident records
3-sensor LSTM coverage	Moderate	Expand to 15+ WAP-7 telemetry channels
COCO-pretrained YOLO	Moderate	Fine-tune on ≥5,000 annotated platform frames
No cross-dataset comparison	Major	Benchmark against FRA and MetroPT datasets

8. Conclusion

This paper presents **RailGuard India**, an integrated machine learning framework for railway safety and accident risk prediction, evaluated within a controlled simulation setting using 300 SMIS-representative synthetic samples. The framework is conceptually informed by long-term accident trend patterns observed in Indian Railways, enabling a structured and causation-aware feature design. The principal findings are as follows:

(i) Need for predictive intervention: While overall accident frequency shows a declining trend, the severity and economic impact per incident have increased substantially, underscoring the limitations of purely reactive safety management approaches.

(ii) Causation-driven feature design: The selected feature set operationalises key environmental and operational risk factors. SHAP-based attribution analysis demonstrates consistent and interpretable alignment between model predictions and domain-informed risk drivers.

(iii) Robust ensemble performance: The weighted soft-voting ensemble achieves a cross-validated accuracy of approximately 93% with stable AUC-ROC performance (~0.94), indicating improved generalisation compared to individual models. Cross-validation-based weighting mitigates potential overfitting and data leakage.

(iv) Predictive maintenance enhancement: The LSTM-based anomaly detection module reduces false-negative maintenance alerts compared to static thresholding approaches, demonstrating improved sensitivity to multivariate temporal patterns in sensor data.

(v) Evaluation scope**:** All reported results are derived from a controlled simulation environment and should not be interpreted as direct indicators of real-world operational performance.

A structured pathway toward real-world deployment is identified: (i) integration with the official SMIS database; (ii) validation on large-scale real accident datasets; (iii) domain-specific fine-tuning of the vision module on annotated railway CCTV data; (iv) expansion of the LSTM model to incorporate comprehensive locomotive telemetry; and (v) benchmarking against established international railway datasets and safety evaluation frameworks.

Acknowledgements: We sincerely thank our guide, Dr. G. Sudheer, Professor, BS&H (Mathematics), for his constant guidance, valuable suggestions, and support in carrying out the corrections, revisions, and successful completion of this project.

References

[1] S. B. Aher and D. R. Tiwari, "Trends in Causes and Impacts of Accidents in Indian Railway," *Journal of Social Sciences*, vol. 55, no. 1-3, pp. 34–44, 2018. DOI: 10.31901/24566756.2018/55.1-3.2226

[2] Ministry of Railways, Government of India, *Annual Report and Accounts 2022–23*. New Delhi: Ministry of Railways, 2023.

[3] Ministry of Railways, Government of India, *Railway Accident Report 2023: Safety Management Information System Summary*. New Delhi: RDSO, 2023.

[4] Research Designs and Standards Organisation (RDSO), *Safety Management System Manual for Indian Railways, Version 3.2*. Lucknow: RDSO, 2023.

[5] H. Alawad, S. Kaewunruen, and M. An, "Learning from Accidents: Machine Learning for Safety at Railway Stations," *IEEE Access*, vol. 8, pp. 44671–44687, 2020.

[6] N. Bešinović, "Artificial Intelligence in Railway Transport: Taxonomy, Regulations and Applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 14011–14026, 2022.

[7] F. A. Awad, T. Uchida, and M. H. S. Eldin, "Predicting Urban Rail Transit Safety via Artificial Neural Networks," *Safety Science*, vol. 158, p. 105973, 2023.

[8] R. C. R. Nampally, "Neural Networks for Enhancing Rail Safety and Security: Real-Time Monitoring and Incident Prediction," *Journal of Artificial Intelligence and Big Data*, vol. 2, no. 1, pp. 49–63, 2022.

[9] F. Ghofrani, Q. He, R. M. P. Goverde, and X. Liu, "Recent Applications of Big Data Analytics in Railway Transportation Systems: A Survey," *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 226–246, 2018.

[10] R. Rautji and T. D. Dogra, "Rail Traffic Accidents: A Retrospective Study," *Medicine, Science and the Law*, vol. 44, no. 1, pp. 67–70, 2004.

[11] A. Banerjee, "Railway Accidents in India: By Chance or By Design?" B.Tech. Thesis, Indian Institute of Technology, Kharagpur, 2011.

[12] D. Pasuranga and S. Baltasar, "Data Driven Predictive Maintenance Framework for Railway Safety in Indonesia," *ADI Journal on Recent Innovation*, vol. 7, no. 1, pp. 75–87, 2025.

[13] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[14] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.

[15] H. Meng, H. Liu, X. Yang, and G. Zhang, "Railway Accident Prediction Strategy Based on Ensemble Learning," *Accident Analysis & Prevention*, vol. 176, p. 106817, 2022.

[16] G. A. Susto, A. Schirru, S. Pampuri, S. McLoone, and A. Beghi, "Machine Learning for Predictive Maintenance: A Multiple Classifier Approach," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 3, pp. 812–820, 2015.

[17] S. García-Méndez, F. de Arriba-Pérez, J. C. Dafonte, and F. J. González-Castaño, "An Explainable Machine Learning Framework for Railway Predictive Maintenance Using Data Streams from the Metro Operator of Portugal," *Scientific Reports*, vol. 15, p. 1734, 2025.

[18] H. Li, R. Parida, O. Olofsson, K. Persson Waller, and M. Odelius, "Improving Rail Network Velocity: A Machine Learning Approach to Predictive Maintenance," *Transportation Research Part C: Emerging Technologies*, vol. 45, pp. 17–33, 2014.

[19] A. Lasisi and N. Attoh-Okine, "Principal Components Analysis and Track Quality Index: A Machine Learning Approach," *Transportation Research Part C: Emerging Technologies*, vol. 91, pp. 230–248, 2018.

[20] A. K. Kouroussis, G. Alexandros, and K. Nikolakopoulos, "Predictive Maintenance Using Machine Learning and Data Mining: A Pioneer Method Implemented to Greek Railways," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 7, no. 1, p. 23, 2021.

[21] H. Li, M. Audette, and Y.-H. Tang, "Railway Track Geometry Degradation Prediction Using Machine Learning," *Transportation Research Part C: Emerging Technologies*, vol. 39, pp. 18–36, 2014.

[22] Y. Yürekli, G. Akay, and C. Ergün, "Real-Time Railway Hazard Detection Using Distributed Acoustic Sensing and Machine Learning," *Sensors*, vol. 25, no. 4, p. 1023, 2025.

[23] A. Jamshidi et al., "A Big Data Analysis Approach for Rail Failure Risk Assessment," *Risk Analysis*, vol. 37, no. 8, pp. 1495–1507, 2017.

[24] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[25] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLO," Version 8.0, GitHub Repository, 2023. [Online]. Available: https://github.com/ultralytics/ultralytics.

[26] Y. Li, L. Wang, H. Liu, and M. Zhang, "Machine Learning-Based Rail Track Degradation Prediction Considering Environmental and Operational Factors," *Engineering Applications of Artificial Intelligence*, vol. 132, p. 107912, 2024.

[27] V. N. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed. Springer, 2000.

[28] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.