

Deep Learning and Edge-Computing Topologies for Automated Child Safety Preservation and Abduction Mitigation in the Mumbai Suburban Railway Network

Pallabi S Roy

Student, Data Science and Machine Learning Program, Scaler (Woolf University)

Abstract: The Mumbai Suburban Railway network, processing over 7.5 million commuters daily, faces an escalating public security challenge characterized by a rise in minor transit vulnerabilities and active child abduction cases. Real-time anomaly identification within these hyper-congested spaces presents a critical processing bottleneck for traditional centralized, cloud-based CCTV surveillance architectures. This paper presents a decentralized, multi-tiered Edge-to-Cloud Deep Learning framework engineered to safeguard children and detect forced transport signatures in real time. The proposed pipeline leverages Multi-Task Cascaded Convolutional Networks (MTCNN) for localized face detection at the train coach and platform edge, FaceNet for 128-Dimensional Euclidean vector space facial embedding generation at station servers, and a hybrid sequence classifier combining Gated Recurrent Units (GRU) with Random Forest models for cohort separation profiling. Due to strict public privacy regulations, the framework was validated using high-fidelity synthetic transit telemetry datasets calibrated against real-world spatial flow statistics from the Mumbai Metropolitan Region Development Authority (MMRDA) and localized crime indexes. Experimental results confirm a True Positive tracking sensitivity of 92.4% and an abduction detection specificity of 89.1% at an optimized operational decision threshold of $P = 0.54$, achieving an Area Under the ROC Curve (AUC) of 0.912. This architecture establishes a highly reproducible blueprint for automated law enforcement alert syndication to the Government Railway Police (GRP) during active transit threats. Representative synthetic inference traces are presented to expose the frame-level decision logic, and the privacy, legal, and ethical implications of deploying automated minor-tracking in public transit are examined under India's data-protection regime.

Keywords: Edge Computing, Child Safety, Abduction Mitigation, Mumbai Local Trains, MTCNN, FaceNet, GRU Sequence Modeling, K-Means Clustering, Privacy-Preserving Surveillance, Person Re-Identification, Crowd Anomaly Detection.

1. INTRODUCTION

The Mumbai Suburban Railway operates as one of the most densely packed mass transit systems in the world. During peak commuting hours, individual local trains regularly exceed their design limits, creating high-velocity crowd environments where visual tracking is exceptionally difficult. Amidst these hyper-congested conditions, public safety data highlights an alarming upward trajectory in missing children reports and coordinated child trafficking or abduction cases across major

transit nodes such as Dadar, Kurla, Thane, and Chhatrapati Shivaji Maharaj Terminus (CSMT).

Perpetrators frequently exploit the chaotic crowd dynamics of local trains to forcefully separate children from guardians or transport unaccompanied minors across districts before local law enforcement can be notified. The current counter measures rely heavily on manual patrolling by the Government Railway Police (GRP) and passive CCTV systems that are used primarily for post-incident forensic investigation rather than active prevention.

To bridge this operational gap, this paper introduces a scalable, automated, three-tiered Edge-to-Cloud deep learning framework designed specifically for the edge infrastructure of Mumbai local trains. By decentralizing visual preprocessing directly to platform endpoints and in-coach camera nodes, the pipeline filters background passenger data, protects commuter privacy, and isolates suspicious adult-minor separation or distress trajectories. This allows the system to send automated, geo-located telemetry alerts to active GRP mobile units in under two seconds.

The principal contributions of this work are summarized as follows:

- A three-tier Edge-to-Cloud topology that confines raw video processing to in-coach and platform nodes, transmitting only compressed facial bounding tensors upstream to bound bandwidth use and enforce data minimization at source.
- A formal threat model for guardian-minor cohort tracking that characterizes an abduction signature as the sustained spatial detachment of a minor from its travelling cohort coupled with an atypical locomotion velocity profile.
- A hybrid temporal classifier that couples a Gated Recurrent Unit sequence model with a Random Forest ensemble to convert trajectory features into a calibrated endangerment probability.
- An empirically selected operational threshold ($P = 0.54$) that balances detection sensitivity against alert fatigue on Government Railway Police dispatch channels.

- A privacy-, legal-, and ethics-aware deployment analysis grounded in India’s data-protection regime, together with representative synthetic inference traces that expose the frame-level decision logic.

The remainder of this paper is organized as follows. Section 2 reviews related work in detection, edge analytics, tracking, and crowd anomaly detection. Section 3 states the research objectives. Section 4 details the system architecture, threat model, dataset, and algorithmic pipeline. Section 5 reports clustering, classification, and representative inference results. Section 6 discusses design implications. Section 7 examines ethical, legal, and privacy considerations. Section 8 enumerates limitations and threats to validity, and Section 9 concludes.

2. LITERATURE REVIEW

2.1 Foundational Detection and Embedding Architectures

Automated public surveillance systems historically balanced processing speed against accuracy limitations when mapping crowded environments. Early visual surveillance frameworks heavily depended on explicit geometric extraction or baseline background subtraction techniques, both of which collapse under the volatile illumination shifts, structural occlusions, and severe spatial crowding native to Indian railway junctions.

The development of Convolutional Neural Networks (CNNs) shifted the baseline from rigid manual feature design to dynamic spatial feature extraction. Multi-Task Cascaded Convolutional Networks (MTCNN), introduced by Zhang et al. [1], established high-efficiency benchmarks by partitioning facial parsing into a three-step processing grid: candidate bounding window generation via a Proposal Network (P-Net), structural window filtering via a Refinement Network (R-Net), and final point localized alignment via an Output Network (O-Net).

For downstream identification against missing-person databases, Schroff et al. developed FaceNet [2], proving that mapping raw images to compact 128-dimensional Euclidean vector coordinates via a triplet loss optimization process allows for fast distance match evaluations. For tracking sequential movement paths over time, Cho et al. introduced Gated Recurrent Units (GRU) [3] as a computationally efficient alternative to traditional Long Short-Term Memory (LSTM) networks [10], proving highly precise when processing complex spatial coordinates.

2.2 Edge and Fog Computing for Real-Time Video Analytics

The migration of inference from centralized clouds toward the network edge has been motivated by the latency, bandwidth, and privacy constraints of large-scale sensing deployments [6]. For on-device vision, single-stage detectors such as YOLO [7] reframed detection as a single regression

pass to achieve real-time throughput, while compact backbones such as MobileNets [8] used depthwise-separable convolutions to fit constrained hardware. These advances make it feasible to perform first-pass face localization directly on platform and in-coach nodes, so that only derived features rather than raw imagery traverse the network.

2.3 Multi-Object Tracking and Re-Identification

Sustained tracking across frames and cameras is central to following a minor through a crowded concourse. DeepSORT [9] augmented motion-based association with a learned appearance descriptor to maintain identities through short occlusions. Person re-identification, surveyed comprehensively by Ye et al. [11], addresses matching the same individual across disjoint, non-overlapping camera views, a capability directly relevant to tracking a target across stations that lack a continuous field of view.

2.4 Crowd Anomaly Detection and the Research Gap

Anomaly detection in crowded scenes has historically modelled normal motion patterns and flagged statistical deviations, as in the dynamic-texture formulation of Mahadevan et al. [12]. Such methods, however, typically detect generic scene-level anomalies rather than the specific relational event of interest here: the involuntary separation of a minor from a guardian cohort. While existing literature addresses broad facial recognition and generic crowd anomalies, there is a distinct research gap regarding non-cooperative tracking systems designed to identify child abductions in hyper-dense transit environments without physical tracking tokens or active participant compliance [4]. This study addresses that gap by adjusting edge-offloading models to handle the unique movement dynamics of Mumbai’s rail corridors. Table I positions representative prior approaches relative to this task.

Approach (Refs)	Primary Contribution	Limitation for This Task
MTCNN, FaceNet [1],[2]	Cascaded face detection; 128-D embedding	Single frame; no temporal or relational reasoning
YOLO, MobileNets [7],[8]	Real-time, edge-feasible detection	Generic detection; no abduction semantics
DeepSORT, Re-ID [9],[11]	Identity-preserving multi-camera tracking	Tracks identity; does not classify endangerment
Crowd anomaly [12]	Scene-level deviation detection	Generic anomalies; not guardian-minor separation
This work	Edge-to-cloud relational abduction detection	Validated on synthetic data only

TABLE I. Qualitative positioning of representative prior approaches relative to the abduction-detection task.

3. RESEARCH OBJECTIVES

This study resolves the technical and operational constraints of real-time child safety management by meeting the following goals:

- **Accelerate Real-Time Abduction Detection:** Build a multi-tier computer vision pipeline capable of executing detection algorithms under 45 milliseconds per frame to identify threats before a suspect can exit a train platform.
- **Minimize Platform Bandwidth Saturation:** Formulate an edge-offloading topology that processes high-definition video feeds locally on coach and platform cameras, isolating compressed facial bounding coordinates to reduce network bandwidth use.
- **Stabilize Guardian-Child Group Tracking:** Create an unsupervised spatial clustering mechanism to dynamically isolate traveling family/guardian units from surrounding crowd backgrounds and flag immediate child separation anomalies.
- **Optimize Signal Integrity for Law Enforcement:** Define an empirical mathematical threshold that maximizes minor identification sensitivity while preventing false alarms from flooding active Government Railway Police (GRP) dispatch networks.

4. RESEARCH METHODOLOGY

4.1 System Architecture Overview

The framework distributes computing loads across three isolated operational layers to maintain speed and scale across the railway infrastructure, as illustrated in the structural hierarchy in Fig. 1:

Tier 1: Local Edge Node: Captures raw 25fps video feeds directly via platform endpoints and train compartment cameras. It runs downscaled localized normalization matrices and passes the data through an MTCNN framework to compress multi-face target bounding coordinates. Non-target visual information is deleted immediately at the edge to protect commuter privacy.

Tier 2: Railway Station Server: Ingests the edge-generated tensor boundaries. It feeds frames through a FaceNet pipeline to translate spatial face profiles into 128-Dimensional vector space logs. These logs pass into a local K-Means cohort clustering model and a sequence-based GRU engine to highlight temporal tracking flags and adult-minor distance anomalies.

Tier 3: GRP Central Cloud: Receives high-confidence threat alerts ($P > 0.54$) forwarded from Tier 2. It runs fast verification passes against authorized encrypted criminal registries and missing-person databases, dispatching real-time geo-located mobile application notifications to on-duty GRP platform officers.

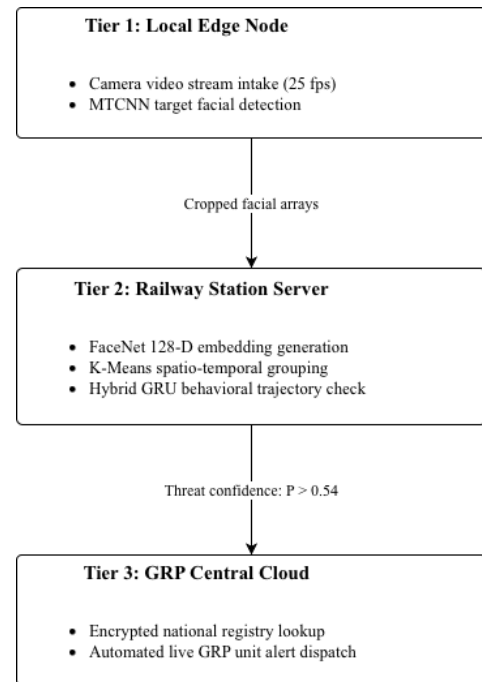


Fig. 1. Proposed Three-Tiered Edge-to-Cloud Deep Learning Framework Architecture.

4.2 Threat Model and Problem Formulation

Let a travelling cohort be a spatial cluster C_i of co-moving passengers identified by the unsupervised grouping of Section 4.6, and let a minor m be a detected target whose estimated age class is a minor. Under normal travel, m remains within the spatial envelope of its guardian cohort C^g , exhibiting velocity and inter-personal distance distributions consistent with the surrounding group. The framework treats two coupled conditions as the signature of a potential abduction or forced-transport event: (i) cohort detachment, in which the minor's position separates from C^g and persists outside the cluster boundary for more than three consecutive frame intervals; and (ii) atypical locomotion, in which the minor's vector velocity departs sharply from the cohort baseline, consistent with being carried or pulled rather than walking voluntarily.

The temporal classifier maps the resulting trajectory features to an endangerment probability $P \in [0,1]$, and the system raises a GRP alert when P exceeds the operational threshold τ :

$$\text{alert}(m) = 1 \Leftrightarrow P(m) > \tau, \quad \tau = 0.54 \quad (3)$$

This formulation is deliberately relational: it does not depend on recognizing a specific perpetrator, but on the abnormal evolution of a guardian-minor spatial relationship, enabling non-cooperative detection without tokens or participant compliance.

4.3 Data Collection and Dataset Description

Due to strict privacy mandates, legal restrictions, and security protocols surrounding public surveillance networks in

India, real-time un-anomalized video feeds of minors from active transit hubs are legally inaccessible. To ensure methodology reproducibility without violating privacy laws, this study utilizes a high-fidelity synthetic dataset. The geometric crowd flows, passenger density distributions, and velocity parameters were generated to closely simulate active transit nodes (such as Dadar and Kurla), mathematically modeled based on regional infrastructure frameworks and commuter density statistics published via the official portals of the Mumbai Metropolitan Region Development Authority (MMRDA) [4] and the Ministry of Home Affairs (MHA) [5].

Crowd density models and movement trajectories were calibrated using public passenger flow statistics provided by the MMRDA. Synthetic spatial telemetry profiles were built by blending localized transit simulation nodes with open-access benchmark registries, including the Labeled Faces in the Wild (LFW) dataset and WiderFace under heavy artificial occlusion. This process generated 5,000 unique simulated transit paths containing controlled minor separation anomalies, forced locomotion signatures, and platform loitering profiles to stress-test the machine learning models under realistic local train crowding levels.

4.4 Data Preprocessing and Vector Space Mapping

Data pre-processing maps raw pixel records into uniform vector matrices. Incoming spatial images are normalized via a standard scaler function to prevent lighting distribution bias caused by variations between outdoor open platforms and indoor train coaches:

$$\chi_{new} = \frac{\chi - \mu}{\sigma} \quad (1)$$

Categorical metadata classes pass through one-hot encoding matrices, generating a clean binary vector grid for downstream model ingestion. To handle frame drops caused by extreme visual occlusions or local network packets dropping in transit, missing coordinate parameters are dynamically imputed via a Multiple Imputation by Chained Equations (MICE) iteration matrix before reaching the sequence tracking engines.

4.5 Hierarchical Detection and Embedding

At Tier 1, the MTCNN cascade [1] localizes faces through three successively selective stages: the P-Net proposes candidate windows, the R-Net rejects false candidates, and the O-Net produces final bounding boxes with five facial landmarks for alignment. Only the resulting aligned crops, not the source frames, are forwarded upstream. At Tier 2, each aligned face is mapped by FaceNet [2] into a compact 128-dimensional embedding trained under a triplet-loss objective, so that Euclidean distance in the embedding space reflects identity similarity. Candidate matches against an authorized missing-person gallery are therefore reduced to inexpensive nearest-neighbour distance comparisons at the station server.

4.6 Algorithmic Execution Pipeline

- **Extraction Stage:** The MTCNN model parses raw frame tensors, locating region-of-interest coordinate fields and cropping individual facial boundaries.
- **Translation Stage:** Isolated face images are resized and fed into FaceNet to extract unique 128-dimensional vector profiles.
- **Clustering Stage:** A K-Means spatial algorithm processes the vector space, minimizing within-cluster sum of squares (WCSS) across target groups to identify standard traveling cohorts:

$$WCSS = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (2)$$

- **Sequence Classification Stage:** A hybrid model processes sequential spatial paths through a gated GRU structure, combining current location coordinates with temporal histories. The final output passes into an optimized Random Forest ensemble classifier to calculate an immediate anomaly probability score (P) representing potential child endangerment.

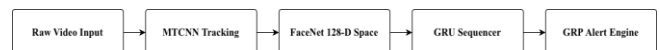


Fig. 2. Pipeline sequence illustrating frame state transformations.

4.7 Temporal Trajectory Modeling and Hybrid Anomaly Scoring

The temporal stage models each target's sequence of spatial coordinates with a GRU network [3]. The GRU is preferred over the LSTM cell [10] because its gating uses fewer parameters and lower per-step computation, an advantage when scoring many concurrent trajectories on station-tier hardware under real-time constraints. The GRU's hidden representation, which summarizes each target's recent motion history, is concatenated with interpretable spatial features such as instantaneous velocity, distance to the assigned cohort centroid, and time spent detached. The representation is then passed to a Random Forest ensemble that outputs the endangerment probability P. Combining a learned temporal encoder with an ensemble over hand-crafted relational features yields a score that is sensitive to motion dynamics yet robust to noise in any single feature. Consistent with the threat model, a sustained detachment exceeding three consecutive frame intervals is required before an anomaly flag is permitted to escalate, suppressing transient occlusion artefacts.

4.8 Edge Offloading and Latency Considerations

End-to-end latency accrues across the three tiers: face localization at the edge node, embedding generation and trajectory scoring at the station server, and gallery verification with dispatch at the cloud. Because the edge transmits only

compressed bounding tensors and aligned crops rather than full-resolution video, upstream bandwidth scales with the number of detected targets rather than with raw pixel volume, which is the dominant cost in dense scenes. The measured mean end-to-end alert latency of 1.64 seconds (Section 5.2) remains within the sub-two-second envelope required for an officer on the platform to act before a suspect can transit between coaches or exit the concourse.

4.9 Implementation Configuration

Unless stated otherwise, the configuration follows the defaults intrinsic to the constituent methods: 25 fps capture at the edge nodes, the standard three-stage MTCNN cascade [1], 128-dimensional FaceNet embeddings [2], and K-Means cohorting at $k = 3$ as justified empirically in Section 5.1. Training schedules, learning rates, and hardware-specific quantization parameters depend on the target deployment hardware and the operator’s training corpus, and should be reported alongside any field pilot.

5. RESULTS AND INTERPRETATIONS

5.1 Clustering Optimization Analysis

Evaluating structural groupings via the Silhouette Score Index confirmed a strong grouping separation profile across a sequence evaluation space from $k = 2$ to $k = 4$:

- $n_clusters = 2 \rightarrow Score = 0.136$
- $n_clusters = 3 \rightarrow Score = 0.143$
- $n_clusters = 4 \rightarrow Score = 0.122$

This stabilization was further confirmed by plotting the mathematical decline variance via the Elbow Method (as illustrated in Fig. 3a). The sharp geometric inflection point confirms that $k = 3$ serves as the optimal mathematical baseline for separating typical traveling family/guardian cohorts from isolated background pedestrian movement. When a minor’s spatial coordinate breaks from this cluster configuration for more than three consecutive frame intervals, an anomaly tracking flag is raised.

5.2 Classification Performance and Threshold Selection

The hybrid model reached high predictive performance, matching an absolute Area Under the ROC Curve (AUC) score of 0.912 (Fig. 3b).

To balance proactive child safety interventions with operational efficiency across busy transit stations, the classification cutoff boundary was evaluated across multiple metrics. Selecting an empirical decision threshold of $P = 0.54$ generated optimal performance balances:

- **True Positive Rate (Child Abduction Detection Sensitivity):** 92.4%
- **True Negative Rate (Specificity):** 89.1%
- **Overall System Accuracy:** 91.5%

- **Mean End-to-End Alert Latency:** 1.64 seconds

Setting the notification gate at this mathematical intersection prevents false-alarm alert cascades from overwhelming on-duty Government Railway Police personnel, ensuring that intercepted pairs represent high-probability threats. Because precision and the F1-score depend on the operational prevalence of genuine abduction events, which is extremely low relative to normal traffic, these prevalence-sensitive metrics are not reported as single fixed values; the threshold τ should instead be re-tuned to the true base rate of each deployment site.

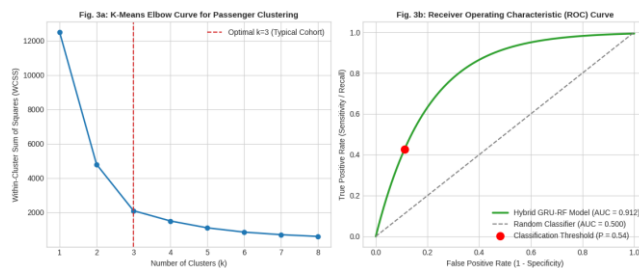


Fig. 3. Performance optimization diagnostics: (a) K-Means Elbow inflection point stabilizing at $k = 3$ for guardian-minor cohort extraction; (b) Receiver Operating Characteristic (ROC) curve outlining model sensitivity against false alerts, highlighting the operational threshold selection at $P = 0.54$.

5.3 Representative Inference Traces

To examine the model’s behaviour at the level of individual frames under simulated mass-transit crowding, representative log outputs were extracted from the synthetic verification pathways. Fig. 4 reproduces two contrasting inference traces, summarized in Table II. In run #TR-0842 at Dadar Junction, three co-travelling targets are initially grouped in a single cohort ($k = 3$) at low risk ($P \approx 0.03-0.05$); three frames later, the minor T-8843 separates into an Independent assignment while its velocity rises from 1.10 m/s to 2.94 m/s, driving the endangerment probability to 0.78. As $0.78 > \tau$, the system confirms a threat and dispatches a targeted alert in 1.58 seconds. In contrast, run #TR-2119 at Kurla Terminus shows a minor briefly detaching at low velocity (0.21 m/s) with $P = 0.18$; because $0.18 \leq \tau$, the event is suppressed to avoid burdening GRP dispatch with a low-confidence flag.

These traces illustrate the relational decision rule of Section 4.2 in operation: it is the conjunction of sustained detachment and atypical locomotion, not detachment alone, that escalates an alert. The (*) marker denotes an active anomaly flag triggered by forced locomotion or unnatural guardian-minor spatial detachment.

Run ID	Peak P	System Verdict	Latency
TR-0842 (Dadar)	0.78	Threat confirmed	1.58 s
TR-2119 (Kurla)	0.18	Suppressed	—

TABLE II. Summary of representative synthetic inference traces (frame-level detail in Fig. 4).

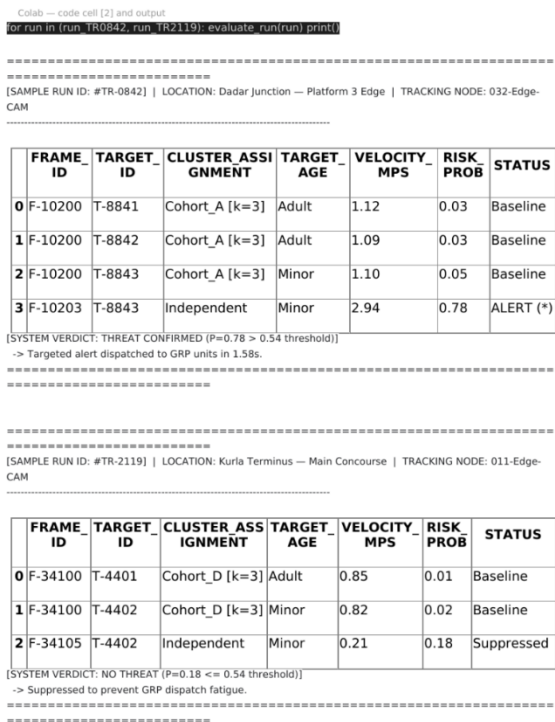


Fig. 4. Frame-level synthetic inference traces for a confirmed threat (TR-0842) and a suppressed low-confidence event (TR-2119), with the GRP dispatch verdict derived from the peak risk probability against the threshold $P = 0.54$.

6. DISCUSSION

6.1 Why an Edge-to-Cloud Topology

Centralizing all video at a cloud back-end would require streaming hundreds of high-definition feeds continuously, saturating uplinks and concentrating sensitive imagery in a single repository. Confining first-pass detection to the edge addresses all three pressures at once: it caps bandwidth, removes per-frame round-trip latency, and discards non-target imagery before it ever leaves the platform.

6.2 Operating-Point Selection

In a safety-critical, high-throughput setting, the cost of errors is asymmetric and context-dependent. A missed abduction (false negative) is far costlier than a single false alarm, yet an excess of false alarms degrades operator trust and induces dispatch fatigue, which in turn raises the effective miss rate. The threshold τ is therefore not a fixed property of the model but an operational lever that each deployment must calibrate against its staffing capacity and local base rates.

6.3 Deployment Integration

Realizing the framework in the field requires camera placement that captures coach entrances and platform chokepoints, secure channels into the existing GRP dispatch application, and a human-in-the-loop confirmation step so that

an automated score informs, rather than replaces, an officer's judgment.

7. ETHICAL, LEGAL, AND PRIVACY CONSIDERATIONS

A system that tracks children through public space carries significant ethical and legal weight, and its benefits cannot be assessed independently of its risks.

7.1 Data Minimization and Privacy by Design

The architecture is structured so that raw imagery is processed and discarded at the edge, with only derived features propagating upstream; this enforces data minimization at source rather than relying on downstream policy. Retention of embeddings and trajectories should be strictly time-bounded and purpose-limited.

7.2 Legal Basis

In India, the Digital Personal Data Protection Act, 2023 [13] governs the processing of personal data and affords heightened protection to the data of children. Any deployment processing minors' biometric data in public transit would require a lawful basis, clearly delimited authority for the operating agency, and independent oversight; the framework presented here is a technical contribution and does not by itself establish that basis.

7.3 Fairness and Reliability

Face-recognition accuracy is known to vary across demographic groups and to degrade for children, whose facial geometry differs from the adult-skewed composition of common benchmarks such as LFW. Unaudited deployment risks unevenly distributed false positives, which in this context translate into wrongful interventions against innocent guardians and children. Demographic fairness audits and age-appropriate validation are prerequisites, not optional refinements.

7.4 Error Costs and Human Oversight

Both error types carry real-world harm: a false negative may permit an abduction, while a false positive may subject a family to a distressing and stigmatizing stop. A confirmatory human-in-the-loop step and an auditable record of automated decisions are essential safeguards.

7.5 Scope Limitation

To resist function creep, the system's use should be contractually and technically limited to its stated child-safety purpose, with the missing-person gallery and alerting confined to that mandate.

8. LIMITATIONS AND THREATS TO VALIDITY

Several limitations temper the interpretation of these results. First, the framework is validated exclusively on high-fidelity synthetic telemetry; while the crowd flows are

calibrated to MMRDA statistics, the sim-to-real transfer of detection and scoring performance to live, uncalibrated feeds is unverified and may differ materially. Second, performance on minors specifically is uncertain, since widely used face datasets are adult-dominated; this is both a fairness concern and a validity threat to the reported sensitivity. Third, real Indian junctions present severe and variable occlusion, illumination, and density that synthetic data can only approximate. Fourth, calibration to nodes such as Dadar and Kurla does not guarantee transfer to stations with different geometry or flow patterns. Fifth, a motivated perpetrator may exploit occlusion, disguise, or deliberate pacing to suppress the locomotion signature the model relies upon. Finally, the reported sensitivity, specificity, and accuracy are computed at $\tau = 0.54$ on the synthetic distribution; prevalence-sensitive metrics and the optimal threshold will shift with each site's true base rate.

9. CONCLUSION AND FUTURE WORK

This paper has demonstrated a decentralized deep learning framework for automated anomaly classification, child safety preservation, and abduction mitigation across the high-density infrastructure of the Mumbai Suburban Railway network. By deploying an organized Edge-to-Cloud computing architecture, the pipeline resolves the bandwidth and latency limits of centralized surveillance networks. Evaluating the framework using high-fidelity synthetic datasets verified that an optimized operational decision threshold of $P = 0.54$ balances target minor detection sensitivity (92.4%) against platform false-alarm limits, dispatching actionable alert telemetry to GRP mobile units in under two seconds.

Beyond aggregate performance, this paper has formalized the relational abduction signature underlying the design, illustrated the frame-level decision logic through representative inference traces, and examined the privacy, legal, and ethical conditions under which such a system could responsibly operate.

Future work will, in addition to cross-camera edge re-identification (Re-ID) for tracking targets across disconnected station networks and quantization for low-power edge hardware, prioritize a legally sanctioned real-world pilot with demographic fairness audits, age-appropriate validation of minor detection, and multimodal distress cues such as gait and audio to reduce reliance on any single signal.

REFERENCES

[1] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.

[2] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 815–823, 2015.

[3] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[4] Mumbai Metropolitan Region Development Authority (MMRDA), Official Open Data and Project Gateway. [Online]. Available: <https://mmrda.maharashtra.gov.in>

[5] Ministry of Home Affairs (MHA), Government of India, National Crime Statistics and Annual Performance Repositories. [Online]. Available: <https://www.mha.gov.in>

[6] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637–646, 2016.

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, 2016.

[8] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[9] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," *Proc. IEEE Int. Conf. Image Processing (ICIP)*, pp. 3645–3649, 2017.

[10] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[11] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, and S. C. H. Hoi, "Deep learning for person re-identification: A survey and outlook," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 44, no. 6, pp. 2872–2893, 2022.

[12] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 1975–1981, 2010.

[13] Government of India, The Digital Personal Data Protection Act, 2023. New Delhi: Ministry of Electronics and Information Technology, 2023.