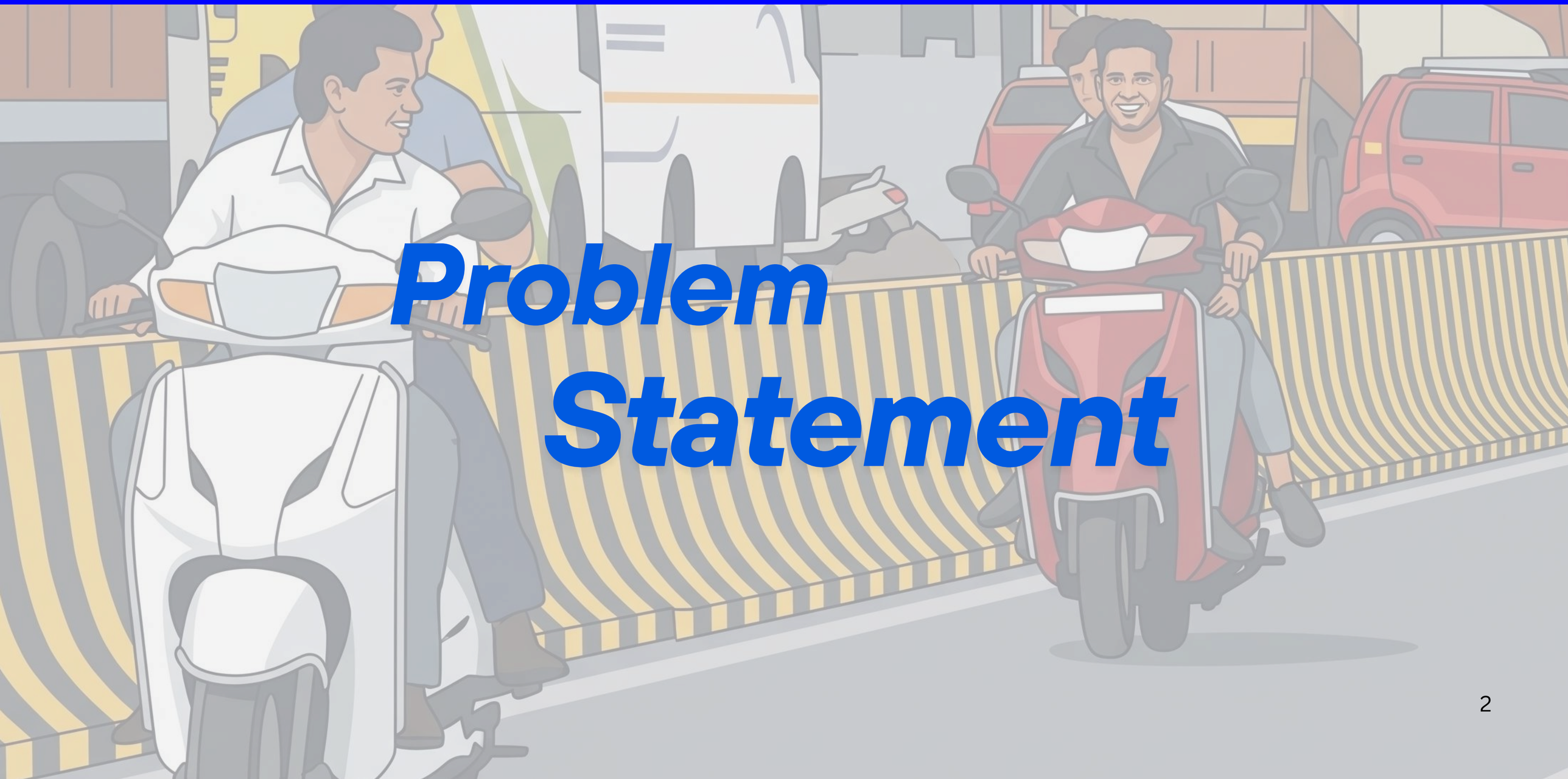
Headgear Detection System

Presented by Advika, Rithwik, Sukrit



Problem Statement

Project vision and mission

Studies show **72%** of all riders per 100 person-minute observation are **not wearing helmets**[1]

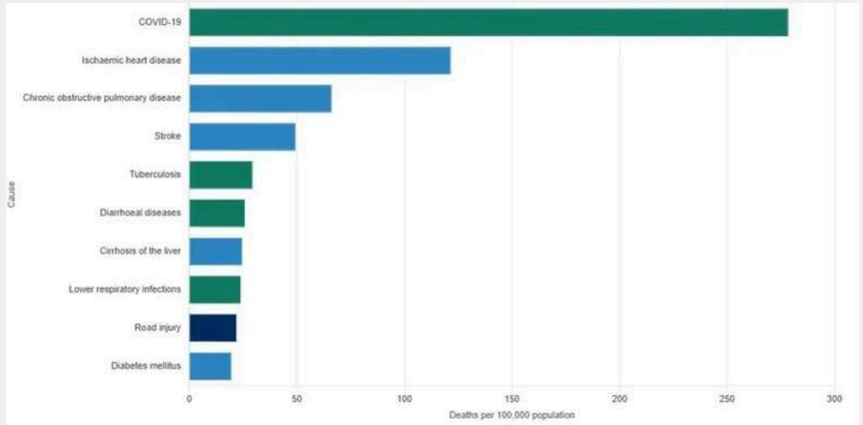
Nationally, Echallan portal reveals that riding without a helmet accounts for approximately **28.2%** of all traffic challans issued.

Road Traffic Accidents are the **9th leading cause of death** in India, Contributing to [1]

1.35 Million Deaths

Rank	State / Union	Total e-Challans	Revenue
1	Tamil Nadu	55,762,916	7,555,816,274
2	Uttar Pradesh	44,003,150	24,951,872,926
3	Kerala	18,835,738	6,909,202,912
4	Haryana	10,390,665	14,651,751,846
5	Delhi	9,022,711	9,022,711
6	Rajasthan	5,855,678	13,934,799,915
7	Odisha	5,411,511	5,000,647,690
8	Bihar	4,341,219	14,038,598,368
9	Himachal Pradesh	3,606,736	3,817,453,286
10	West Bengal	3,344,857	3,184,688,520

The cumulative financial value of these penalties generated a theoretical revenue collection pool of **Rs. 12,631 crore !**

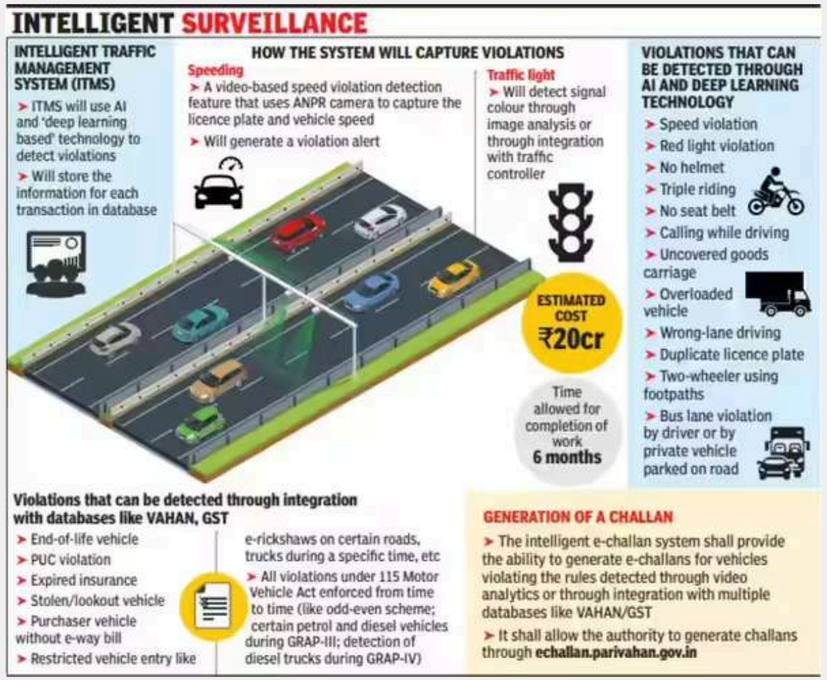


Project vision and mission

Legal BackBone

- **Section 136A** adds a provision requiring state governments to ensure **electronic monitoring** and enforcement of road safety on **national highways,** state highways, **roads**, or in any urban city**.**
- **Rule 167A of the CMVR** complements this. It outlines the placement of electronic enforcement devices in high-risk and high-density areas, and allows the **use of footage** from these devices to issue **penalties for traffic violations**.
- Under **SASCI (Scheme of Special Assistance to States for Capital Investment),** Haryana receives incentives for improving road safety through e-enforcement, Traffic Control Room integration, e-challans, faster challan disposal, and reduced road fatalities.

Rooting from this ITMS is the new way to the future for India, with The national capital currently rolling out an advanced AI-powered ITMS across 1000 traffic signals

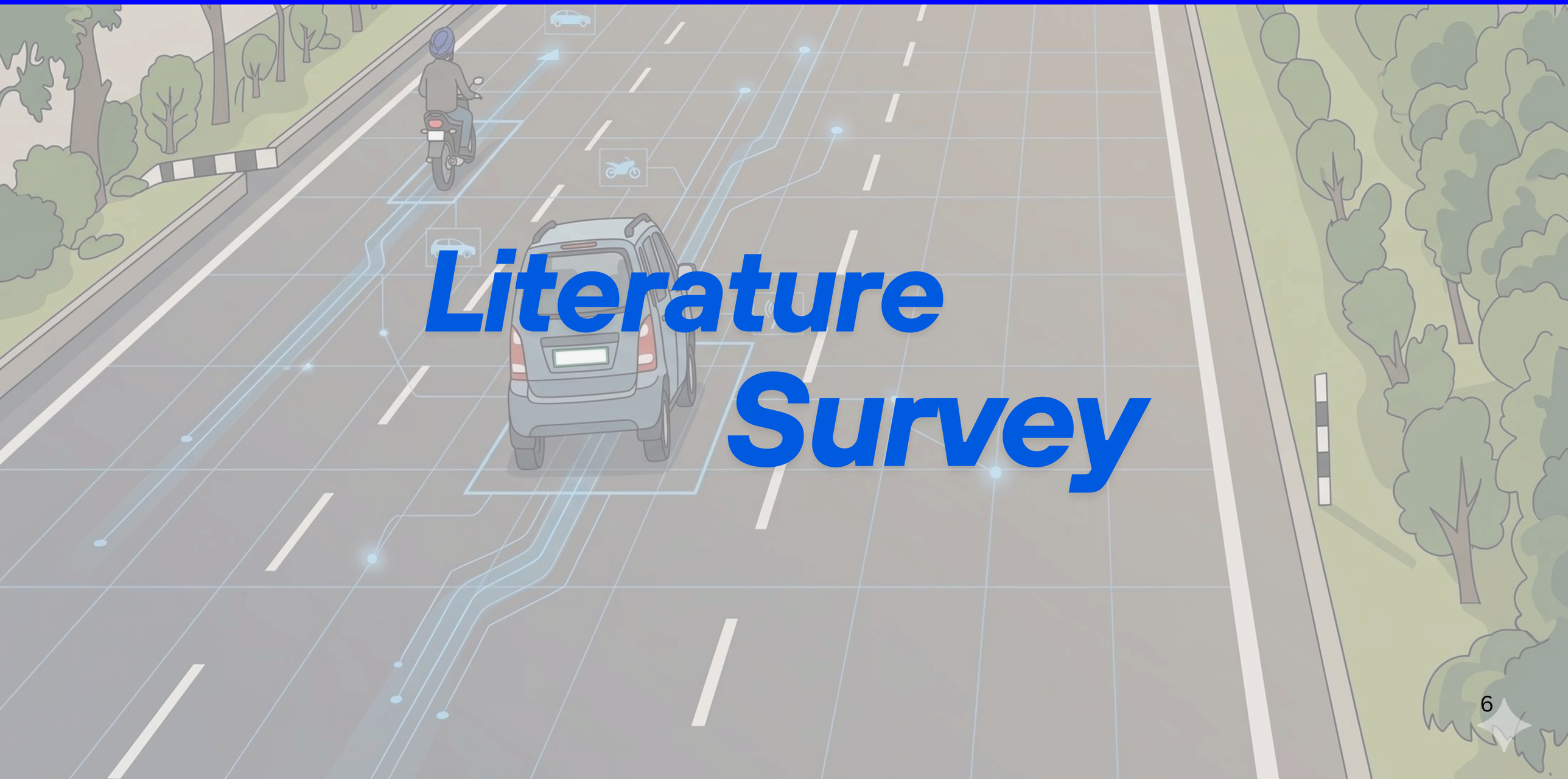


Project vision and mission

Current State

- Real-world deployment of Intelligent Traffic Management System systems in Bengaluru showed that practical performance can differ significantly from theoretical lab testing, since Indian traffic conditions are far more unpredictable and complex.
- While the Bengaluru ITMS achieved nearly 97% accuracy for simple "no-helmet" detection, accuracy for complex violations such as lane indiscipline and dangerous driving dropped to nearly 75–80%, leading to a high risk of false challans.
- These limitations highlight the need for culturally and contextually representative Indian datasets, including handling real-world exceptions such as turbans, hijabs, varied helmet styles, dense traffic, and heavy occlusions.

Model Architecture	Key Application	Precision	Recall	Processing Speed
YOLOv8 + Deep SORT	Holistic Violation & Tracking	98.5%	97.2%	Real-time
LoLTV (ResNet + ViT)	Low-Light Helmet & Signal	98.2%	97.5%	High
GA-YOLOv5	Helmet Detection	95.4%	94.8%	45 FPS
Transformer Models	General Object Detection	93.8%	92.3%	18 FPS
FastViT (MobileNet+ViT)	Vehicle Classification	91.7%	N/A	Real-time
CNN + LSTM	Temporal/Erratic Driving	91.1%	89.2%	22 FPS
DashCop (YOLOv8-x)	Dashcam-based Triple Riding	76.1%*	N/A	Real-time



Literature Survey

Two Dominant Research

A. Generic Binary Classification (The Standard Approach)

Current automated enforcement relies heavily on Convolutional Neural Networks (CNNs) like YOLOv8. However, most research treats this as a strictly binary problem ("Helmet" vs. "No-Helmet").

- Key Works: Barik et al. (2024) and Bamniya et al. (2025).
- The Result: High processing speeds on standard datasets, but catastrophic failures in culturally diverse environments. Sikh riders wearing legally exempt Pagdis are mathematically guaranteed to be wrongfully penalized (False Positives) or misclassified.

B. Data-Level Imbalance Solutions

Other researchers attempt to fix algorithmic biases by manipulating the training data itself.

- Key Work: Balancing Helmet Detection with Synthetic Data (Woo et al., 2025).
- The Result: Introduced "Ratio-Aware Synthetic Data" to boost detection of rare "No-Helmet" events, proving that calibrating dataset distribution is critical.

Component	Standard Binary YOLO (Barik, Bamniya)	DashCop (Rawat et al.)
Feature Extraction	Identifies generic round shapes on heads, misclassifying dupattas as helmets.	Accurately tracks riders, but feature maps for helmets are polluted by turban data.
Classification Logic	Will illegally ticket a Sikh rider as "No-Helmet".	Will classify a Sikh rider as "Helmet", distorting safety analytics.

Two Dominant Research

Cultural Nuance: The Blind Spot of India-Centric AI Models

Taxonomic Flaws & Feature Pollution: The DashCop (Rawat et al., 2025) system attempted to solve the turban issue by instructing annotators to label 'Pagdi' as 'Helmet'. This merges soft fabrics with hard plastics, confusing the neural network and degrading localization accuracy.

Visual Distractors (The Soft Boundary Problem): Standard models misclassify the bulky silhouettes of dupattas, burqas, or gamchas as helmets, causing dangerous False Negatives where severe violations go unpenalized.

Industry Recognition: Very recent research, such as Kandimalla et al. (2026), emphasizes the critical need for "custom-built datasets accommodating cultural and regional variations", confirming this is an unsolved, high-priority industry challenge. current research suffer from vulnerability to sharp angles and occlusions.

Method	Reference	Dataset/Scope	Accuracy	Limitations
DNN-YOLOv8	Barik et al. (2024)	General Traffic	72.8%	Strictly binary ("With Helmet" / "Without Helmet"); fails on cultural exemptions.
YOLOv3 + IoT Ignition	Bamniya et al. (2025)	Urban Environments	96.0%	Lacks cultural context; authors cite "cultural contexts" as unsolved future work.
YOLOv8-x (DashCop)	Rawat et al. (2025)	Indian Dashcam (RideSafe-400)	High (mAP)	Taxonomic flaw: Intentionally groups 'Turban' into the 'Helmet' class to avoid false positives, causing feature pollution.
Ratio-Aware Synthetic Data	Woo et al. (2025)	Highly Imbalanced Data	High	Fixes minority class counts but struggles with localization on soft fabric boundaries.

Two Dominant Research

Bridging Cultural Disparity and Data imbalance Gaps

Taxonomic Flaws

We reject DashCop's feature pollution by **enforcing a strict 3-class system: Helmet, Pagdi, and No-Helmet**. This mathematically isolates cultural exemptions from rule violations.

localization Precision

We adapt Woo et al.'s methodology by applying **Distribution Focal Loss** alongside **Mosaic and Mixup augmentations**. Furthermore, we intentionally include dupattas and hijabs in our "No-Helmet" class to explicitly teach the YOLOv11 model the geometric variance between soft fabrics and hard polycarbonate.

Enhancing Model Robustness

To solve the occlusion and angle vulnerabilities cited in contemporary research, we curated a **highly diverse training dataset** spanning various demographics across the Indian subcontinent. Crucially, we prioritized images capturing riders from sharp, unconventional angles to train the YOLO model to recognize helmets even under severe geometric distortion.

Datasets And Features Preprocessing

Dataset Description And Collection Strategy

Data Collection

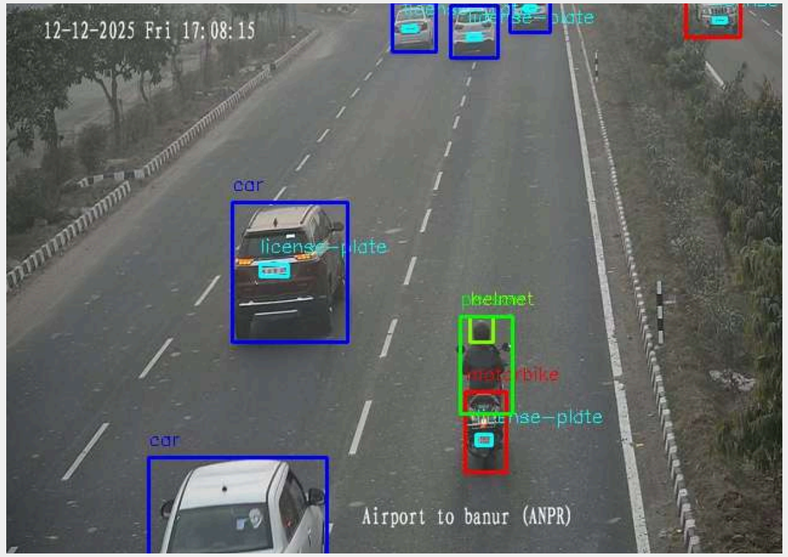
Our initial strategy pivoted around building a native dataset using Plaksha University campus footage to capture real-world top-down surveillance angles. However, we quickly encountered a bottleneck:

- **The Volume Problem**: Human annotation was brutally slow. After manually bounding 100-200 images, we realized the riders were often tiny, blurry pixel clusters, providing insufficient volume to train a neural network.
- **The Scale-Up**: We pivoted to a multi-source approach, scraping established internet datasets and running a local auto-annotation pipeline.

Data Composition

Our final curated dataset is built around a strict 3-class Head ROI (Region of Interest) system:

1. **Helmet** (Safety standard compliant)
2. **No-Helmet** (Rule violators)
3. **Pagdi** (Legally exempt cultural headgear)



Class Imbalance & The Annotation Pipeline

building the dataset = wrestling with infrastructure, algorithmic hallucinations, and severe class imbalances.

Challenge / Feature	Description of the Problem	Actions Taken (Our Solution)
Massive Class Imbalance	Internet datasets had abundant examples of safety helmets but almost zero usable data for *Pagdis*.	**YouTube Mining:** We actively scraped targeted YouTube videos specifically to mine high-quality images of Sikh riders to fix the Pagdi deficit.
Compute & Scalability	Local auto-annotation was too slow for the newly expanded, massive dataset.	**Kaggle Migration**: We migrated our pipeline to Kaggle to leverage free 32GB compute environments, deploying the Vicuna VLM (Vision-Language Model) for rapid auto-annotation.
The "Pagdi" Exception	Our first major Kaggle run resulted in a catastrophic failure: the system labeled *everything* as a Pagdi.	**Logic Patching:** We identified this not as an AI hallucination, but an implementation logic bug in our pipeline. We successfully patched it and re-ran Vicuna.
Cultural Nuance Blindness	Even after patching, Vicuna struggled with Indian cultural nuances, frequently mislabeling draped fabrics (like saris) as Pagdis.	**Human-in-the-Loop Validation:** We manually validated and cleaned Vicuna's mistakes. Though tedious, this yielded a beautifully balanced, high-quality training graph.

Image Pre-Processing & Feature Engineering

For computer vision, features : pixels data & Structural information

Environmental Normalisation

Issue:
datasets - "*too perfect*" - (solid white backgrounds)
Fix:
standardise lighting, contrast, and background noise

Blurry Pixel Cluster

Issue:
Plaksha cams - “tiny, blurry pixel clusters"
Fix:
introduced targeted visual noise in training set.

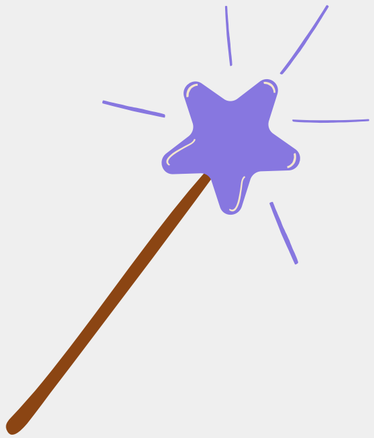
```python
# These filters apply to EVERY image to bridge the sensor/lighting gap
global_augs = [
    A.MotionBlur(blur_limit=(3, 5), p=0.2),
    A.GaussianBlur(blur_limit=(3, 5), p=0.2),
    A.RandomBrightnessContrast(brightness_limit=0.2, contrast_limit=0.2, p=0.4),
    A.HueSaturationValue(hue_shift_limit=10, sat_shift_limit=15, val_shift_limit=10, p=0.3)
]

# These are the surgical filters for specific folder biases
surgical_augs = {
    "studio_bias": [
        # Uses the new range tuples for holes and dimensions
        A.CoarseDropout(num_holes_range=(1, 6), hole_height_range=(10, 48), hole_width_range=(10, 48), p=0.5),
        A.HueSaturationValue(hue_shift_limit=40, sat_shift_limit=40, p=0.6),
        A.ToGray(p=0.2)
    ],
    "low_res_crowd": [
        A.CLAHE(clip_limit=4.0, tile_grid_size=(8, 8), p=0.6),
        A.RandomBrightnessContrast(contrast_limit=0.4, p=0.5)
    ],
    "dashcam_cctv": [
        # Uses std_range (as a fraction) instead of var_limit
        A.GaussNoise(std_range=(0.2, 0.4), p=0.6),
        # Uses quality_range instead of lower/upper
        A.ImageCompression(quality_range=(50, 75), p=0.7)
    ],
    "default": []
}
```

Image Pre-Processing & Feature Engineering

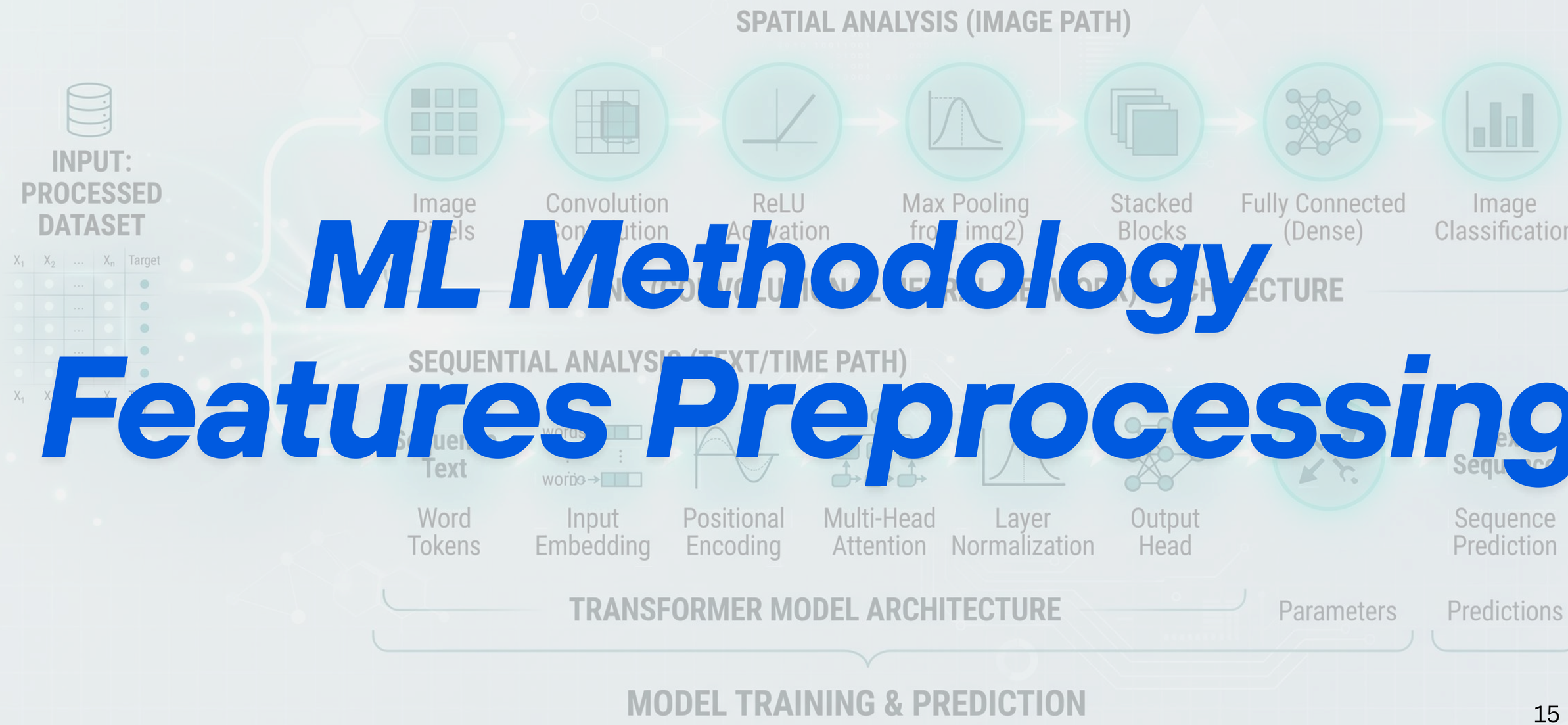
Region Of Interest(ROI) Cropping

Dimensionality Reduction

Instead of processing massive 4K frames filled with irrelevant data (trees, roads, cars), our pipeline mathematically crops the image down to the **specific Head Region**. This drastically **reduces feature dimensionality**, providing the classifier (like **ResNet**) with a concentrated matrix of pixels containing only the rider's head.



```
class_id = CLASS_MAP[final_class]

# The Geometric Heuristic
final_cx, final_cy, final_bw, final_bh = cx, cy, bw, bh

if class_id in [1]:
    new_bh = bh * 0.60
    removed_height = bh * 0.40
    new_cy = cy - (removed_height / 2)

    final_bh = new_bh
    final_cy = new_cy

# Write YOLO Format
f.write(f"{class_id} {final_cx:.6f} {final_cy:.6f} {final_bw:.6f} {final_bh:.6f}\n")
```



ML Methodology
Features Preprocessing

The Architectural Contrast

3 different paradigms of how a computer "sees" : trade-offs in speed, context, and accuracy.

YOLOv11 (Standard Baseline)

- **How it works**: Uses **standard spatial grid convolutions** to divide an image into regions and predict bounding boxes and probabilities simultaneously.
- **Why we used it**: It is the industry standard for **real-time object tracking**, offering an exceptional balance of **high inference speed** and accuracy.

RT-DETR (Transformer Alternative)

- **How it works:** Instead of scanning a grid, it uses "**Self-Attention**" (similar to LLMs) to evaluate the **global context** of the entire image simultaneously.
- **Why we used it:** We hypothesized its global context awareness would be theoretically superior for dense Indian traffic, where overlapping motorcycles and bodies confuse standard convolutional grids.

ResNet-50 (Deep Feature Extractor))

- **How it works:** Utilizes "skip connections" to train **extremely deep neural networks** without the vanishing gradient problem, allowing it to learn highly complex, nuanced textures.
- **Why we used it:** Benchmarked as an "expert attribute classifier" to see if a heavier, decoupled network could better **distinguish the complex textures** of a Pagdi from the smooth shell of a Helmet**.**

How Do the Models Work?

Aspect	YOLOv11 (Spatial Grid)	RT-DETR (Transformer)	ResNet-50 (Residual)
Core Architecture	C3k2 blocks and C2PSA attention mechanisms.	Multi-Scale Deformable Attention.	Deep Residual Blocks with Skip Connections.
Input Logic	Full frame 640x640 grid division.	Full frame sequence flattening.	Cropped ROI (Region of Interest) input.
Feature Extraction Focus	Localized spatial shapes and fast-moving contours.	Global image context and object-to-object relationships.	Deep textural nuances (e.g., fabric folds vs. smooth plastic).

BBox Approach (custom) →

Helmet / No-Helmet : Bounding boxes map the entire face and head to capture the strap and jawline context.

Pagdi: Bounding boxes are strictly drawn around the **headgear only**, intentionally excluding the face to prevent facial features from polluting the fabric data.

Challenges Faced & Algorithmic Resolutions

Challenges

1. Early models struggled with the complex geometries of Indian traffic, frequently **confusing the soft boundaries** of thick beards or scarves with the edges of a Pagdi or Helmet.

2. While RT-DETR and ResNet-50 provided excellent theoretical accuracy, they were **computationally heavy**. Deploying them on edge hardware (like Jetson Nano) for traffic surveillance results in severe frame drops and thermal throttling.

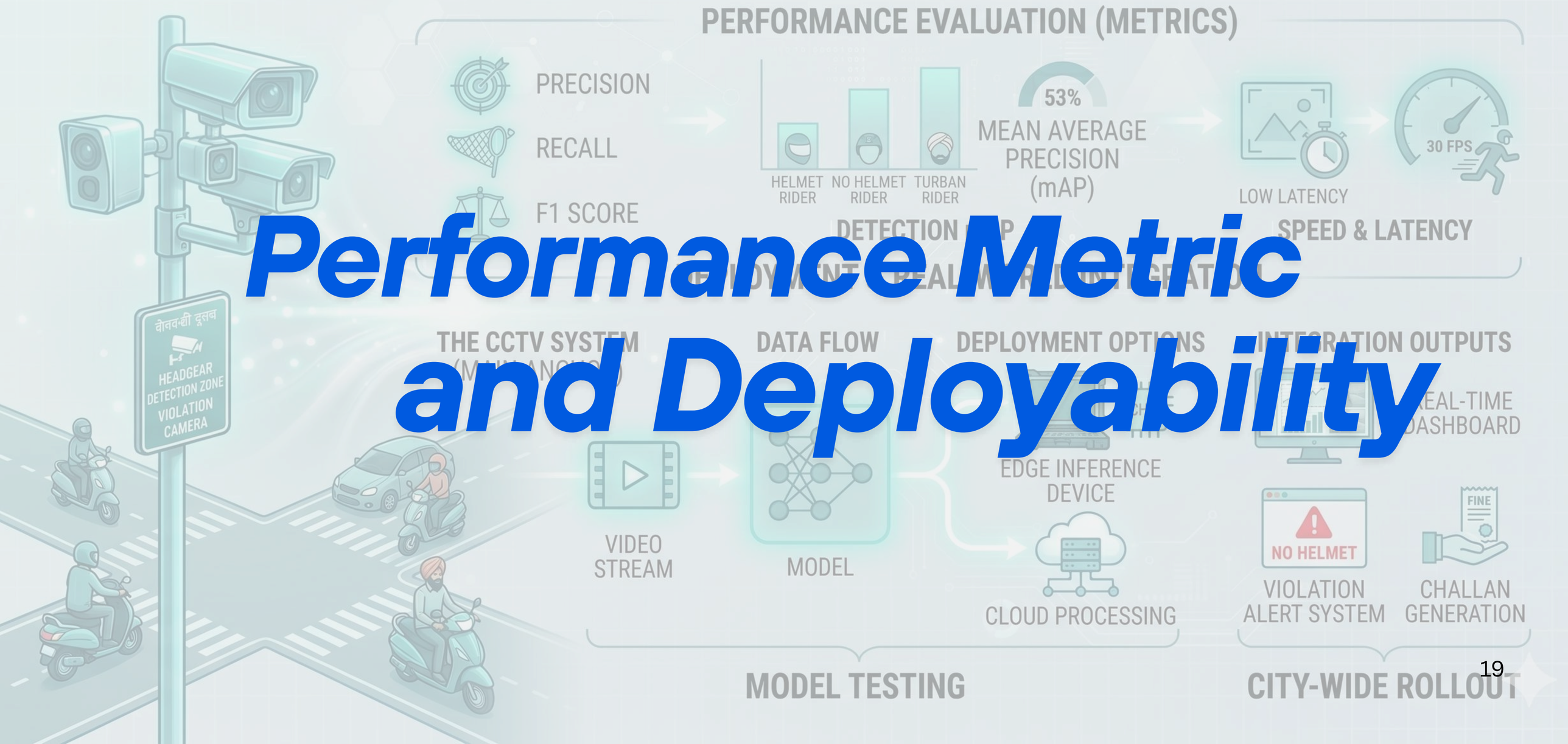
3. Training these heavy architectures on thousands of high-resolution images **maxed out local hardware** capabilities, causing out-of-memory errors and halting progress.

Mitigations

1. We engineered the specific BBox Logic mentioned previously. Isoating the Pagdi annotations away from the face, we forced the neural network to focus only on the geometric folds of the fabric, drastically **reducing false positives.**

2. We ultimately selected **YOLOv11 as our primary deployable model**. It runs flawlessly on live CCTV footage, requires significantly less VRAM, and delivers real-time predictions (>30 FPS) without sacrificing the accuracy needed for traffic enforcement.

3. We migrated our entire training pipeline to **cloud computing environments** (Kaggle), utilizing free 32GB GPU clusters to process the massive, augmented datasets efficiently.



Performance Metric and Deployability

Data Collection

1. MoRTH Parliamentary Reply, Release ID: 2083423. https://www.pib.gov.in/PressReleaseIframePage.aspx?PRID=2083423®=3&lang=2
2.

References

1. Bamniya, H., Garg, R., & Rangnekar, S. (2025). Enhancing road safety through AI-powered helmet detection and connected vehicle systems. ResearchGate.
2. Barik, J. K., Sahoo, S., Mohanty, S., Giri, P., & Barik, R. C. (2024). An autonomous helmet detection model in diversified application areas using DNN-YOLOv8 technique. In 2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS) (pp. 1–7). IEEE. https://doi.org/10.1109/SCEECS61402.2024.10482114
3. Dash, D. K. (2025, December 11). Speeding killed nearly 1.24 lakh people, non-wearing of helmet & seatbelt caused 39% deaths. The Times of India. https://timesofindia.indiatimes.com/india/speeding-killed-nearly-1-24-lakh-people-non-wearing-of-helmet-seatbelt-caused-39-deaths/articleshow/125904755.cms
4. Karim, A., Raza, M. A., Alharthi, Y. Z., Abbas, G., Othmen, S., Hossain, M. S., Nahar, A., & Mercorelli, P. (2024). Visual detection of traffic incident through automatic monitoring of vehicle activities. World Electric Vehicle Journal, 15(9), Article 382. https://doi.org/10.3390/wevj15090382
5. Pandey, Kapil, Singh, S., & Singh, J. (2025). Real-time helmet detection using synthetic dataset and YOLOv11. In Advances in Intelligent Systems and Computing (pp. 36–45). Springer. https://doi.org/10.1007/978-3-032-02831-0_36
6. Press Information Bureau. (2024). Expansion of Intelligent Traffic Management System [Press release]. Ministry of Road Transport & Highways, Government of India. https://www.pib.gov.in/PressReleasePage.aspx?PRID=2078259
7. Rawat, D., Gupta, K., Roy, A., & Sarvadevabhatla, S. R. K. (2025). DashCop: Automated e-ticket generation for two-wheeler traffic violations using dashcam videos. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) (pp. 5387–5397). IEEE. https://doi.org/10.1109/WACV61041.2025.00526
8. Setty, N. K. H., Sukumar, G. M., Majgi, S. M., Goel, A. D., Sharma, P. P., & Anand, M. B. (2020). Prevalence and factors associated with effective helmet use among motorcyclists in Mysuru City of Southern India. Environmental Health and Preventive Medicine, 25(1), Article 47. https://doi.org/10.1186/s12199-020-00888-z
9. Spennemann, D. H. R. (2021). Turbans vs. helmets: A systematic narrative review of the literature on head injuries and impact loci of cranial trauma in several recreational outdoor sports. Sports, 9(12), Article 172. https://doi.org/10.3390/sports9120172
10. Sutar, A. (2025). AI-based integrated traffic violation detection and smart traffic management system: A comprehensive review. International Journal on Science and Technology, 16(2), 669–678. https://doi.org/10.71097/ijsat.v16.i2.6693
11. Udbhavi, B. (2024, March 10). In Bengaluru, AI traffic cameras fail accuracy test. Deccan Herald. https://www.deccanherald.com/india/karnataka/bengaluru/in-bengaluru-ai-traffic-cameras-fail-accuracy-test-2930585
12. Woo, S., Kang, M.-S., Kim, P.-K., Lee, K., Kim, K.-J., & Lee, K. (2025). Balancing helmet detection with synthetic data for class imbalance. In Proceedings of the International Conference on Ubiquitous and Future Networks (ICUFN) (pp. 1–2). IEEE.

Thank You!!