North Asian International Research Journal of Multidisciplinary

IRJIF IMPACT FACTOR: 6.012

ISSN: 2454-2326 Vol.11, Issue-4 April-2025

Index Copernicus Value: 58.12 Indian Citation Index

NAIRJC UGC-Aligned | HEI-Approved Peer Reviewed Refereed Journal (Under New UGC Guidelines)

DOI: doiglobal.org/10.2025/NAIRJCMD.019

DEEP LEARNING-BASED REAL-TIME TRAFFIC SIGN DETECTION AND CLASSIFICATION FOR INDIAN ROAD ENVIRONMENTS

DR. JYOTI. Y. KULKARNI

Associate Professor Govt. First Grade College, Kalagi.

ABSTRACT

Automatic traffic-sign perception is a fundamental component of advanced driver-assistance systems, autonomous vehicles and intelligent transportation systems. Indian road environments make this task particularly difficult because traffic signs may be small, partially occluded, faded, damaged, visually cluttered or observed under strong illumination and weather variations. This paper presents a deep-learning framework for real-time detection and classification of Indian traffic signs from road images and video streams. The proposed system uses a one-stage YOLO-family detector as the principal real-time architecture and incorporates image normalization, contrast enhancement, data augmentation, multiscale learning, transfer learning and confidence-based non-maximum suppression. The research design emphasizes Indian datasets and real-world conditions rather than relying only on international benchmark signs. Published evidence is used as comparative context: an Indian Refined Mask R-CNN study used 6,480 images containing 7,056 sign instances across 87 categories and reported 97.08% precision; a 2026 study on IRTSD-Datasetv1 describes 5,141 images spanning 37 traffic-sign classes from more than 90 Indian cities and reports an increase in baseline CNN accuracy from 89.3% to 96.7% after preprocessing; and an improved YOLOv4 study introduced an Indian dataset containing more than 3,000 images for unconstrained traffic-sign detection. These published figures are not represented as results of the proposed experiment. The paper defines a reproducible evaluation protocol using precision, recall, F1-score, mean Average Precision, latency and frames per second. It concludes that a lightweight YOLO-based system, trained on diverse Indian road imagery and tested under unseen environmental conditions, offers a practical route toward real-time traffic-sign perception, but safety-critical deployment requires rigorous external validation.

***KEYWORDS**: Indian traffic signs; deep learning; YOLO; real-time object detection; CNN; intelligent transportation systems; computer vision; autonomous driving*

1. INTRODUCTION

Traffic signs communicate regulatory, warning and informational instructions that support orderly and safe road use. A human driver continuously identifies signs, interprets their meaning and combines them with road context. Automated vehicles and driver-assistance systems require a comparable perception capability.

Traffic-sign recognition is often divided into two tasks. Detection identifies the location of a sign in a full road scene, usually by a bounding box or segmentation mask. Classification assigns the detected sign to a semantic category such as Stop, No Entry, Speed Limit, School Ahead or Turn Left. A real-time system must perform both tasks accurately while processing video at useful frame rates.

Indian roads introduce distinctive challenges. Sign appearance may vary because of fading, dust, physical damage, non-standard placement, occlusion by vegetation or vehicles, and heterogeneous backgrounds. Dense traffic, roadside advertisements and complex streetscapes increase false-positive risk. Signs also occupy only a small fraction of many camera frames.

The aim of this paper is to design an academically defensible real-time framework centered on deep object detection and Indian road data. It separates published benchmark results from proposed experimental results and emphasizes external-condition testing.

2. PROBLEM STATEMENT

Conventional image-processing approaches based on colour and shape can detect well-formed traffic signs under controlled conditions, but they are sensitive to illumination, background clutter and physical deterioration. Deep learning can learn richer features, yet real-time deployment introduces a trade-off between predictive accuracy and computational cost.

The research problem is therefore to develop a model that detects small and diverse Indian traffic signs with sufficient accuracy while maintaining low inference latency. The model must also generalize beyond the roads, cameras and lighting conditions represented in its training set.

3. OBJECTIVES

- To develop a real-time deep-learning framework for detection and classification of Indian traffic signs.
- To analyze Indian-specific traffic-sign datasets and data-quality challenges.
- To evaluate a lightweight YOLO-family detector against heavier two-stage or segmentation-oriented architectures reported in the literature.
- To improve robustness through preprocessing, augmentation and multiscale training.
- To evaluate detection using precision, recall, F1-score and mAP at defined IoU thresholds.
- To evaluate real-time suitability using latency, model size and frames per second.

- To identify failure modes involving small signs, occlusion, night scenes, weather, clutter and class imbalance.

4. RESEARCH QUESTIONS

- Can a lightweight one-stage detector achieve a useful balance between accuracy and real-time speed on Indian road scenes?
- How does preprocessing influence recognition under variable illumination and degraded sign conditions?
- Which sign classes are most affected by class imbalance and small bounding-box size?
- How well does the trained model generalize to unseen cities, cameras and environmental conditions?
- What deployment safeguards are necessary before using traffic-sign predictions in safety-critical systems?

5. SCOPE OF THE STUDY

The study focuses on visual detection and semantic classification of roadside traffic signs using RGB images and video. It does not attempt to infer road rules that are not visually represented, nor does it replace vehicle localization, lane perception or driver decision-making.

The proposed architecture is suitable for dashcams, mobile devices, edge processors and research vehicles. Training can be performed on a workstation or GPU server, while inference may be optimized for lower-power hardware.

6. INDIAN TRAFFIC SIGN ENVIRONMENT

Indian traffic signs broadly include regulatory, warning and informatory categories, but real-world appearance may differ substantially from clean sign templates. Perspective, ageing, stickers, paint loss, shadows and obstruction can modify the visual pattern.

Roadside visual clutter is especially important for object detectors because advertisements, circular logos and coloured boards may resemble signs. A robust dataset must therefore contain negative background examples rather than only tightly cropped traffic-sign images.

Real-time recognition should also account for temporal behavior. A distant sign may occupy only a few pixels in early video frames and become clearer as the vehicle approaches. Frame-level detection can be improved in deployment through tracking or temporal confirmation.

7. LITERATURE REVIEW

Deep-learning traffic-sign research includes classification CNNs, Faster R-CNN, Mask R-CNN, SSD and YOLO-family detectors. Two-stage detectors can provide strong localization but are typically computationally heavier. One-stage detectors directly predict object locations and classes, making them attractive for real-time applications.

An Indian study published in the International Journal of Transportation Science and Technology proposed Refined Mask R-CNN. Its customized dataset contained 6,480 images and 7,056 Indian traffic-sign instances distributed across 87 categories. The study reported 97.08% precision, exceeding its Mask R-CNN and Faster R-CNN comparisons.

Saxena, Dey, Shah and Gupta developed an improved YOLOv4 approach for unconstrained traffic-sign environments. Their work modified feature propagation and grouped convolutions and introduced an Indian traffic-sign dataset containing more than 3,000 images. The focus on small signs and real-time recognition is directly relevant to Indian road perception.

A 2026 study using IRTSD-Datasetv1 reports 5,141 images across 37 traffic-sign classes collected from more than 90 Indian cities. Its preprocessing pipeline combines resizing, normalization, CLAHE, Gaussian filtering, morphology and feature-analysis techniques. When integrated with a baseline CNN, reported accuracy increased from 89.3% to 96.7%.

A 2025 Indian traffic-sign dataset study based on the Mumbai-Pune Expressway evaluated YOLOv8 and reported a moderate mAP@0.5 of 0.440, highlighting class imbalance, small bounding boxes and confusion with environmental features. This is a useful reminder that performance depends strongly on dataset difficulty.

8. RESEARCH GAP

Several research gaps remain. First, traffic-sign classification on cropped signs does not prove that a model can detect small signs within full road scenes. Second, datasets collected along limited routes may not represent the visual diversity of India. Third, random image splitting can overestimate generalization when adjacent video frames from the same drive appear in both training and test sets.

Fourth, some papers report only accuracy even though object detection requires localization-sensitive metrics such as mAP. Fifth, real-time claims should include measured inference speed on specified hardware rather than model architecture alone.

The proposed research therefore uses route- or sequence-level splitting, detection metrics and deployment metrics, and recommends an external-city test set.

9. DATASET STRATEGY

The preferred dataset combines a recognized Indian traffic-sign dataset with newly captured road scenes if available. Every image requires bounding-box annotation and a class label. Annotation quality should be audited because small errors around tiny signs can strongly influence IoU-based evaluation.

Data should cover regulatory, warning and informatory signs; highways and urban roads; daytime and evening conditions; clear and partially occluded signs; large and small apparent sign sizes; and negative road scenes without signs.

Where multiple frames come from the same video, complete sequences should remain in a single split. This prevents nearly identical neighboring frames from leaking into training and testing.

10. IRTSD-DATASETV1 CONTEXT

The 2026 IRTSD-Datasetv1 study describes 5,141 images spanning 37 classes and more than 90 Indian cities. Such geographic diversity is valuable because the visual environment around a sign can differ substantially between metropolitan, semi-urban and highway settings.

The study's preprocessing results also demonstrate that input quality matters. Its reported baseline CNN accuracy improved from 89.3% to 96.7% after the proposed preprocessing framework. This result belongs to the published study and is included here only as literature evidence.

11. DATA ANNOTATION

For object detection, each traffic sign is represented by a bounding box (x, y, width, height) and a class identifier. YOLO-format annotation commonly stores normalized center coordinates and normalized dimensions.

Tiny or heavily occluded signs should be annotated according to a written policy. Ambiguous signs may be marked 'ignore' rather than forcing unreliable labels. A sample of annotations should be independently reviewed by a second annotator.

12. PREPROCESSING

Input images are resized to the detector's training resolution while preserving aspect ratio through letterboxing where appropriate. Pixel normalization follows the implementation requirements of the selected backbone.

CLAHE may be evaluated for difficult illumination because it improves local contrast. Gaussian filtering can suppress high-frequency sensor noise, but strong smoothing should be avoided because distant signs contain few pixels.

Preprocessing should be validated experimentally. A complex pipeline is not automatically beneficial because modern detectors can learn many photometric variations directly from augmented data.

13. DATA AUGMENTATION

Training augmentation can include horizontal transformations only when the sign meaning remains valid; careless flipping can reverse directional signs. Suitable augmentations include brightness and contrast variation, mild blur, scale changes, perspective transformations, partial occlusion and weather-like distortions.

Mosaic and multiscale augmentation can improve small-object learning in YOLO architectures. Class-aware augmentation can increase exposure to rare sign categories, but synthetic examples should not dominate the validation or test sets.

14. PROPOSED YOLO-BASED ARCHITECTURE

A lightweight contemporary YOLO-family detector is proposed as the primary architecture because one-stage detection offers low latency. The network consists conceptually of a backbone for hierarchical feature extraction, a neck for multiscale feature fusion and a detection head for bounding-box regression and class prediction.

Multiscale features are essential because traffic signs may range from a few pixels to large nearby objects. High-resolution feature maps retain information needed for small signs, while deeper maps capture semantic context.

For resource-constrained deployment, a nano or small variant should be evaluated. A larger model can serve as an accuracy-oriented reference.

15. TRANSFER LEARNING

Transfer learning initializes the detector with weights learned on a large object-detection dataset. Early layers already encode general visual primitives such as edges, textures and shapes. Fine-tuning adapts these features to Indian traffic signs.

A staged schedule can freeze portions of the backbone initially and then unfreeze the complete network at a lower learning rate. Early stopping and checkpoint selection should rely on validation mAP rather than test performance.

16. MATHEMATICAL FORMULATION

For a predicted bounding box Bp and ground-truth box Bg, Intersection over Union is IoU = area(Bp ∩ Bg) / area(Bp ∪ Bg). A prediction is considered a true positive when its class is correct and IoU exceeds the chosen evaluation threshold.

Precision is TP/(TP+FP), Recall is TP/(TP+FN), and F1 = 2 × Precision × Recall /(Precision+Recall). Average Precision integrates precision across recall for a class; mean Average Precision averages AP across classes.

mAP@0.5 uses IoU 0.50, while mAP@0.5:0.95 averages performance across progressively stricter IoU thresholds and is therefore more demanding.

17. LOSS FUNCTION

Modern one-stage detectors optimize a combination of localization, objectness and classification losses, with exact formulation depending on the selected implementation. Bounding-box losses based on IoU variants encourage overlap and geometric alignment.

Class imbalance can be mitigated through sampling, augmentation or loss weighting. The chosen strategy should be reported because rare traffic-sign classes may otherwise contribute little to training.

18. TRAINING PROTOCOL

A recommended split is approximately 70% training, 15% validation and 15% test data, but the critical requirement is sequence-level separation. If geographic metadata are available, an additional city-held-out evaluation provides stronger evidence of generalization.

Training should record optimizer, learning rate, batch size, image resolution, epochs, early-stopping criterion, pretrained weights and hardware. Random seeds should be fixed where feasible.

The locked test set should be evaluated only after architecture and hyperparameter selection are complete.

19. REAL-TIME EVALUATION

A real-time detector must be evaluated using latency and frames per second on explicitly named hardware. FPS = number of processed frames / elapsed inference time. Preprocessing and post-processing should be included when reporting end-to-end system speed.

Thirty FPS is commonly associated with real-time video, but useful driver-assistance systems may operate at different rates depending on vehicle speed, camera field of view and processing pipeline. Therefore, latency in milliseconds is often more informative than an arbitrary FPS threshold.

20. PUBLISHED COMPARATIVE RESULTS

Study	Dataset / Scope	Model	Precision / Accuracy	Recall	mAP / Speed	Notes
Indian TSR study (2023)	6,480 images; 7,056 instances; 87 classes	Refined Mask R-CNN	97.08% precision	—	—	Indian road images
Saxena & Dey (2023)	TT-100K benchmark	Dense YOLOv4	94.30% accuracy	—	32 FPS	Published conference result
Deosi et al. (2026)	IRTSD-Datasetv1: 5,141 images; 37 classes	Baseline CNN + preprocessing	96.7% accuracy	—	—	Up from 89.3%
Sane et al. (2025)	Mumbai-Pune Expressway Indian signs	YOLOv8	Class-dependent	Class-dependent	mAP@0.5 = 0.440	Class imbalance noted

Table 1 summarizes results reported by previous studies. These values are literature benchmarks and are not experimental results generated by the present proposed model.

21. PROPOSED EXPERIMENTAL RESULTS TABLE

Model	Precision	Recall	F1	mAP@0.5	mAP@0.5:0.95	FPS
YOLO Nano/Small Proposed Model	To be measured	To be measured	To be measured	To be measured	To be measured	To be measured
Medium YOLO Reference	To be measured	To be measured	To be measured	To be measured	To be measured	To be measured
Two-Stage Reference	To be measured	To be measured	To be measured	To be measured	To be measured	To be measured

These cells should be populated only after actual training and evaluation on the selected Indian dataset. Published figures from unrelated experiments should not be inserted as if they were produced by this model.

22. CONFUSION MATRIX ANALYSIS

A class-level confusion matrix should identify signs that are frequently confused. Visually similar circular regulatory signs and speed-limit signs may require higher input resolution or additional training examples.

Errors should be separated into missed detections, localization errors and classification errors. This provides more useful guidance than a single aggregate accuracy value.

23. SMALL-OBJECT DETECTION

Small signs are a major challenge because downsampling removes fine details. Potential improvements include higher input resolution, feature-pyramid refinement, additional small-object detection heads and targeted augmentation.

The improved YOLOv4 literature specifically emphasizes small traffic signs in unconstrained environments, supporting the importance of multiscale feature propagation for this task.

24. OCCLUSION AND DEGRADED SIGNS

Indian signs may be obscured by tree branches, poles, parked vehicles or posters. Training augmentation can simulate partial occlusion, but real occlusion examples remain necessary.

Damaged and faded signs raise a separate question: a system should recognize them when enough semantic evidence remains, but extremely ambiguous signs should produce low confidence rather than forced classification.

25. NIGHT AND ADVERSE WEATHER CONDITIONS

Night-time glare, headlights, rain and fog can substantially alter traffic-sign visibility. Reflective sign surfaces may become saturated while non-reflective signs disappear into the background.

A deployment-grade dataset should therefore include environmental subsets and report performance separately for day, night, rain/fog and strong-shadow conditions. Overall mAP can hide severe failure in a safety-critical subset.

26. EXPLAINABLE AND TRUSTWORTHY DETECTION

For object detectors, visualization of predicted boxes and confidence scores is the first level of interpretability. Feature-activation or saliency methods can provide additional insight into which sign regions influence classification.

Confidence calibration is important because a numerical score should correspond reasonably to empirical reliability. Low-confidence detections can be ignored or confirmed over multiple video frames.

27. VIDEO TRACKING EXTENSION

Traffic signs persist across consecutive frames. A tracker can associate detections over time, reduce flicker and improve confidence through temporal voting. Tracking can also estimate whether a sign is approaching or leaving the field of view.

Temporal confirmation should not conceal detector failures; frame-level metrics and tracked-system metrics should both be reported.

28. EDGE DEPLOYMENT

A lightweight model can be exported to ONNX, TensorRT, TensorFlow Lite or another supported runtime depending on the target platform. Quantization can reduce model size and inference time, although accuracy should be re-evaluated after conversion.

Deployment testing should include thermal behavior, sustained FPS and memory use rather than only a short desktop benchmark.

29. APPLICATIONS

• Advanced driver-assistance systems that alert drivers to upcoming regulatory or warning signs.

• Autonomous-vehicle perception research for Indian road conditions.

• Traffic-sign inventory and maintenance surveys.

• Detection of missing, damaged or obscured roadside signs.

• Smart-city road monitoring and geotagged infrastructure databases.

• Driver-training and road-safety research applications.

30. SAFETY AND ETHICAL CONSIDERATIONS

Traffic-sign recognition can influence safety-critical decisions. A research model should not be represented as sufficient for autonomous vehicle control without extensive validation, redundancy and compliance with applicable automotive safety processes.

Road imagery can incidentally capture faces and vehicle registration plates. Dataset governance should therefore consider privacy, lawful collection and anonymization where appropriate.

31. LIMITATIONS

The proposed paper does not claim new empirical accuracy because no raw dataset, annotations, training logs or checkpoints were supplied for this manuscript. Its experimental tables distinguish clearly between published evidence and measurements that require actual model training.

Dataset coverage remains a fundamental limitation. A model can perform strongly on one route while failing in another State or environmental condition. External validation is therefore essential.

32. FUTURE SCOPE

Future research can investigate transformer-based detectors, improved small-object heads, synthetic sign generation, domain adaptation and self-supervised pretraining on unlabeled Indian driving video.

A unified Indian traffic-sign benchmark with geographically separated train and test routes would improve comparability. Multimodal systems could combine camera detections with high-definition maps and vehicle localization.

Federated learning may eventually allow multiple fleets to improve a common model without centralizing all raw road video, although privacy and model-security concerns would require careful engineering.

33. MAJOR FINDINGS

• Indian traffic-sign detection is a small-object and domain-generalization problem, not merely an image-classification task.

• Published Indian studies demonstrate strong potential for deep learning but use different datasets and metrics, making direct numerical comparison difficult.

• Preprocessing and augmentation can materially improve recognition, particularly under variable illumination and class imbalance.

• One-stage YOLO-family detectors are attractive for real-time use because of their speed, while heavier architectures remain useful accuracy references.

• Sequence-level and city-held-out testing provide stronger evidence than random frame splitting.

• mAP and class-wise recall should be reported alongside FPS; classification accuracy alone is insufficient for a detector.

• Safety-critical deployment requires external validation across road types, weather, night conditions and damaged signs.

34. CONCLUSION

Real-time detection and classification of Indian traffic signs is an important computer-vision problem with direct relevance to intelligent transportation and driver assistance. Deep learning has substantially improved the ability to detect signs in complex road scenes, but Indian environments continue to challenge models through small objects, occlusion, visual clutter, class imbalance and environmental variability.

The proposed framework uses a lightweight YOLO-family detector supported by carefully designed preprocessing, augmentation, multiscale features and transfer learning. Its evaluation protocol separates localization quality, classification performance and real-time efficiency.

Published Indian research provides encouraging evidence: Refined Mask R-CNN has reported 97.08% precision on a custom 87-category dataset, while the 2026 IRTSD-Datasetv1 preprocessing study reports a baseline CNN improvement from 89.3% to 96.7%. These results demonstrate feasibility but cannot substitute for testing the proposed model on a locked Indian-road test set.

The most important requirement for future research is generalization. A traffic-sign system intended for Indian roads should be evaluated not only on familiar frames but across unseen routes, cities, cameras, weather and sign conditions. With such validation, lightweight deep-learning detectors can become a practical component of safer and more intelligent road systems.

REFERENCES

1. Indian traffic sign detection and recognition using deep learning. (2023). International Journal of Transportation Science and Technology, 12(3), 683–699. https://doi.org/10.1016/j.ijtst.2022.06.002
2. Saxena, S., Dey, S., Shah, M., & Gupta, S. (2024). Traffic sign detection in unconstrained environment using improved YOLOv4. Expert Systems with Applications, 238, 121836. https://doi.org/10.1016/j.eswa.2023.121836

3. Saxena, S., & Dey, S. (2023). Traffic sign detection and recognition using dense connections in YOLOv4. In Advances in Data Computing, Communication and Security. Springer. https://doi.org/10.1007/978-3-031-31417-9_32
4. Deosi, H. S., Singh, G., & Kumar, N. (2026). A comprehensive preprocessing and feature extraction methodology for traffic sign detection on Indian roads using the IRTSD-Datasetv1. Tuijin Jishu/Journal of Propulsion Technology, 47(2).
5. Sane, P., Joshi, V., & Patel, S. (2025). Indian traffic sign dataset for self driving cars in YOLO format. 2025 IEEE International Transportation Electrification Conference (ITEC-India). https://doi.org/10.1109/ITEC-India67626.2025.11439261
6. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 779–788.
7. Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934.
8. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in Neural Information Processing Systems, 28.
9. He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. Proceedings of the IEEE International Conference on Computer Vision, 2961–2969.
10. Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6, 60. https://doi.org/10.1186/s40537-019-0197-0