ISSN: 2093-5374 / E-ISSN: 2093-5382

Deep Learning-Based Multi-Class Vehicle Detection: A High-Speed Approach for Traffic Surveillance and Smart City Applications

K. Ramakalyani[1*], and Dr.S. Umamaheswaran[2]

[1*]Research Scholar, New Horizon College of Engineering, Visvesvaraya Technological University (VTU), Belagavi, India. ramakalyani.mtech@gmail.com, https://orcid.org/0000-0001-6745-9973

[2]Research Supervisor, Professor, Department of CSE, New Horizon College of Engineering, Visvesvaraya Technological University (VTU), Belagavi, India. dr.umamaheswaran.nhce@newhorizonindia.edu, https://orcid.org/0000-0003-1941-9605

Received: September 08, 2025; Revised: October 22, 2025; Accepted: November 21, 2025; Published: December 15, 2025

Abstract

The rapid growth of urbanization requires effective traffic surveillance systems designed for smart city use. This paper presents a new Multi-Class Vehicle Detection Framework (MCVD-T) that uses Vision Transformers (ViTs) and improved feature extraction techniques for real-time, precise multi-class vehicle detection. Unlike previous models, such as CNNs and YOLOv9, this framework utilizes Swin Transformers, which are able to perform feature extraction at multiple levels. The implementation of Swin Transformers makes it feasible to operate in dynamic and complex environments. To lessen the computational load, which can significantly hinder performance in real-time applications, the framework applies lightweight attention mechanisms and model pruning techniques to provide better resource utilization, thus lessening inhibit delay and improving scalability in large-scale traffic networks. Additionally, knowledge distillation helps lessen the computing effort while providing a smaller yet effective inference model. The experimental results confirm that the MCVD-T framework surpasses the best existing methods by 22% improved detection accuracy and a processing time reduction of 35%. Both offline and online traffic monitoring capabilities are provided by this system, which allows for real-time insights using a fast data pipeline embedded within edge sensors. Its online functionality is asynchronous streaming based, which guarantees that video is analyzed consistently and allows for rapid vehicle classification for traffic management and law enforcement. MCVD-T architecture overcomes the common problems associated with such a system and provides a fast, scalable, and configurable solution, making it a critical part of smart urban infrastructure. It integrates seamlessly into smart city applications and enhances traffic management, public safety, and more mobile options for efficiency.

Keywords: Vehicle Detection, Vision Transformers, Smart City, Real-Time Surveillance, Traffic Monitoring.

Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), volume: 16, number: 4 (December), pp. 579-600. DOI: *10.58346/JOWUA.2025.I4.033*

*Corresponding author: Research Scholar, New Horizon College of Engineering, Visvesvaraya Technological University (VTU), Belagavi, India.

1 Introduction

Rapid urbanization has caused a considerable increase in traffic vehicles and numerous challenges in urban planning and traffic management. These days, the most important factors for mobility, green transportation, and public safety are effective traffic monitoring and control systems. Conventional traffic monitoring systems typically struggle with accuracy, latency, and efficiency, particularly in a dynamic urban environment, because they rely on antiquated methodology like rule-based algorithms and historical detection techniques (Khalladi et al., 2024). Intelligent traffic systems can now be implemented on a large scale because to the advancements in deep learning and machine learning (Tatar et al., 2024). A key component of smart city technology and improved traffic monitoring is real-time multi-class vehicle recognition (Verdhan et al., 2024). Nevertheless, current deep learning architectures, such as YOLO or Convolutional Neural Networks (CNNs), struggle to strike a balance between detection accuracy and computing sustainability (Chaudhuri, 2024; Shamim et al., 2024; Mishra et al., 2024). In this regard, traffic monitoring techniques must be improved. In this study, we present the Multi-Class Vehicle Detection Framework (MCVD-T), a novel approach to feature extraction using modern techniques and Vision Transformers (ViTs) (Moushi et al., 2024; Lilhore et al., 2024). In particular, the Swin Transformer model efficiently combines hierarchical elements, making it more comprehensible while managing the complexities of traffic data that one comes across in real-world situations (Kulkarni & Sengupta, 2022). We find a strategy that balances processing costs with sufficient accuracy by using lightweight attention patterns and pruning the models during training. Knowledge distillation takes the idea a step further by creating a condensed and effective version that guarantees real-time capabilities (Chacko, 2023). MCVD-T was created to provide adaptability and scalability for use in large-scale and laboratory settings. Excellent online and offline capabilities are established by the real-time streaming capability and the lightweight and effective data pipelines. This makes it possible to categorize cars in real time, analyze a temporal video stream, and provide quick insights for different stakeholders in traffic management and law enforcement. Compared to state-of-the-art methods, the MCVD-T framework offers a 35% reduction in processing time and a 22% improvement in detection accuracy by reducing processing latency and resource usage. This creativity aligns with the broader trend of smart city expansion, where intelligent systems enable optimal urban processes. MCVD-T makes smart mobility solutions, improves public safety, and manages traffic efficiently to meet the immediate needs of urbanization. Future smart cities depend on this study because it improves traffic monitoring technology, enabling sustainable and smart urban infrastructure.

Motivation: The growing complexity of urban traffic, caused by rapid urbanization, requires scalable and efficient traffic monitoring systems. YOLO versions and traditional CNNs struggle to adjust to changing traffic situations and are not computationally efficient according to current theories (Bellam et al., 2022). This shows the need for new solutions that provide real-time, accurate, and resource-efficient vehicle identification to support smart city applications. This will help increase mobility, ensure public safety, and improve traffic management.

Problem statement: Achieving high-speed, accurate, multi-class vehicle recognition while keeping processing requirements low is challenging for current traffic surveillance systems. Traditional models fail to adapt to the dynamic and complex nature of city traffic, leading to high latency and limited scalability. These limitations hinder real-time analysis and the integration of smart technologies needed for smart city infrastructure and effective traffic control.

Contribution of this paper: We designed a new Multi-Class Vehicle Detection Framework using Swin Transformers for hierarchical feature extraction, ensuring excellent performance in tough traffic conditions (Rani et al., 2023; Mounika et al., 2022; Dhurgadevi et al., 2022). We included lightweight attention techniques, model pruning, and knowledge distillation to significantly reduce computing demands, allowing for real-time inference and scalability. We developed a high-speed data pipeline with asynchronous streaming for smooth online and offline traffic monitoring, enhancing traffic management and law enforcement capabilities (Eslami & Yun, 2021). The rest of this paper is organized as follows: In section 2, we review related work in traffic surveillance. In section 3, we explain the proposed methodology of MCVD-T. In section 4, we discuss and analyse the efficiency of MCVD-T. Finally, section 5 concludes the paper and outlines future work.

2 Related work

This collection of articles mainly discusses traffic monitoring, vehicle counting, and categorization in challenging urban areas using deep learning and artificial intelligence methods (Mimouna, 2021). The focus is on improving accuracy, efficiency, and usefulness in smart transportation systems by tackling issues like multi-class vehicle counting real-time traffic analysis, IoV-based prediction, UAV-based monitoring, and smart city video surveillance (Tang, 2024).

Trajectory-Aided Vehicle Counting Framework (TVCF)

The paper proposes a comprehensive framework for counting vehicles at complex intersections. It combines a trajectory-based method for tracking vehicle movements across different areas with modern deep learning techniques for vehicle recognition and tracking (Almehdhar et al., 2024; Gowsikraja, 2025). By addressing the challenges posed by complex traffic conditions, this blending improves counting accuracy. The percentage suggested methodology achieves excellent performance, with greater performance considering the challenges presented by single-camera circumstances. Using pre-processed data from certain CCTV systems (such as real-time or pre-recorded), its accuracy ranges from 80% to 98%.

Multi-Class Multi-Movement Vehicle Counting (MCMVC)

This study demonstrates how traffic data addressing the vehicle counting problem, in conjunction with deep learning-based object detection approaches, can aid in various classifications and motions (Jin et al., 2023; Anastasiu et al., 2020). This method, which counts cars to ascertain their varied behaviours and features, is crucial for optimizing signal timing in smart transportation systems (Chehri et al., 2020). At 20 observation sites with varying levels of sun or rain, the study employs cameras focused on traffic flow and several important tests (Mudarakola et al., 2024). The practical application of real-time classification automatic vehicle counting is examined in this study (Trivedi et al., 2023).

Efficient Multiscale Traffic Classification (EMTC)

A deep learning traffic classification model that prioritizes size and performance is proposed in this work (Farid et al., 2023). This is achieved by varying size, width and resolution to reduce parameters. It also improves feature extraction through the incorporation of spatial information via an attention mechanism. The model captures flow-level characteristics at different scales through lightweight feature fusion. The proposed model achieves over 99.82% accuracy by reducing size parameters by 0.26 million parameters and performance calculations by 5.26 million calculations, making it application-ready for IoT environments (Geetha Rani et al., 2023; Geetha Rani et al., 2023).

Ensemble-Enhanced IOV Traffic Prediction (EE-IOV)

The proposed research presents an intelligent transportation system for smart cities derived from the Internet-of-Vehicles (IOVs) concept, where vehicles are networked objects (Dadheech et al., 2024; Prakash et al., 2024). With ensemble learning and feature selection involved, the system achieves high degree of detection accuracy and keeps computing costs low (Geetha Rani & Tukkoji, 2024). The machine learning approaches proposed for the research are decision trees, random forests, extra trees and XGBoost (Rani & Chetana, 2022). Thus, the primary objective of this traffic system is to predict and detect traffic congestion in areas that lack physical infrastructures and Internet access. The simulation results establish that tree-based machine learning methods with feature selection achieve higher performance than those without (Gad et al., 2021).

Dynamic Frame Skipping Vehicle Detection (DFS-VD)

The authors present a traffic analysis system that uses deep learning for real-time vehicle detection, classification, and counting from video streams (Bui et al., 2020; Ahmed et al., 2024). The system includes a dynamic skipping approach that shortens processing time by skipping unneeded video frames while maintaining accuracy. The model uses information about vehicle make, model, and year to achieve higher frame rates (Rani et al., 2024). MobileNet’s rate increases from 71.2 to 89.17, and VGG19's from 48.2 to 59.14, outpacing previous models (Dhanalakshmi et al., 2023). This technology has applications in various real-time monitoring situations.

MultEYE: UAV-based Multi-Task Traffic Monitoring

The MultEYE system is designed for Unmanned Aerial Vehicles (UAVs) for real-time traffic monitoring. It employs a multi-task learning architecture that combines object detection, tracking, and speed estimation while adding a segmentation head to the object detection backbone (Balamuralidhar et al., 2021). This method achieves a 4.8% improvement in accuracy and accelerates speed by 91.4% over current models. As one of the first UAV-based traffic monitoring systems suitable for real-world settings, it operates on high-resolution photos up to 33 FPS.

Smart City Video Surveillance System (SCVSS)

This paper provides an overview of video surveillance systems in smart cities and examines their relationship with advanced technologies like edge computing, cloud computing, blockchain, and deep learning (Myagmar-Ochir & Kim, 2023; Geetha Rani & Tukkoji, 2025; Tukkoji, 2024). The report highlights the importance of these systems in enhancing citizen quality of life, security, and public safety. It also addresses the challenges these systems face and offers suggestions to optimize them, ensuring real-time, efficient monitoring to improve the function of smart cities.

Multi-Class Traffic Monitoring System (MCTMS)

This work presents an end-to-end system for real-time multi-class identification of motorbikes, helmets, and license plates in traffic monitoring. It uses YOLO (You Only Look Once) for object detection and EasyOCR (Easy Optical Character Recognition) for license plate recognition (Charef et al., 2024; Haval & Afzal, 2024). The technology tracks motorcycle numbers, helmet compliance, and license plates over video feeds (Rani et al., 2024). Although the model performs well in many traffic conditions, it still struggles to monitor small motorcycles, particularly those that move unpredictably. Further adjustments and data collection are planned to improve performance.

Table 1: The Comparison of Exiting Methods

S. No	Methods	Advantages	Limitations
1	TVCF (Trajectory-Aided Vehicle Counting Framework)	High accuracy (80%-98%) in complex traffic situations, effective with CCTV data.	Challenges in handling occlusions and multi-lane intersections.
2	MCMVC (Multi-Class Multi-Movement Vehicle Counting)	Accurate multi-class vehicle counting for diverse actions, supports intelligent transportation systems.	Performance may degrade under extreme lighting or weather conditions.
3	EMTC (Efficient Multiscale Traffic Classification) (Hazman et al., 2023)	Efficient model with reduced parameters and high accuracy (99.82%). Optimized for IoT settings.	May require fine-tuning for specific IoT applications or environments.
4	EE-IOV (Ensemble-Enhanced IOV Traffic Prediction)	High detection accuracy, low computing costs, applicable in smart cities with limited infrastructure.	Relies on ensemble learning, which may increase complexity.
5	DFS-VD (Dynamic Frame Skipping Vehicle Detection)	Improves frame rates without sacrificing accuracy, suitable for real-time applications.	May struggle with high-speed vehicle tracking in fast-moving traffic.
6	MultEYE (UAV-based Multi-Task Traffic Monitoring)	High accuracy and speed improvements (4.8%), real-time UAV-based monitoring.	Limited by UAV flight time and potential for interference with high-resolution images.
7	SCVSS (Smart City Video Surveillance System) (Geetha Rani & Chetana, 2023)	Enhances citizen safety and quality of life, integrates edge and cloud computing for real-time monitoring.	Security and privacy concerns with widespread video surveillance in public spaces.
8	MCTMS (Multi-Class Traffic Monitoring System)	Effective multi-class tracking of vehicles, helmets, and license plates using YOLO and EasyOCR.	Difficulty with tracking small motorcycles, especially those moving erratically.

Table 1 compares existing methods of deep learning, ensemble learning, and systems based on UAVs. Together, these methods offer new ways to analyze traffic, identify vehicles, and count them. This approach ensures reliability, lowers computation costs, and accommodates various traffic situations. It highlights major improvements in real-time monitoring and prediction for today's urban areas. The goal is to improve traffic management, smart city infrastructure, and safety.

3 Proposed Method

This paper introduces new systems for real-time vehicle identification, detecting anomalies, and tracking under changing urban traffic conditions. The proposed systems increase efficiency, accuracy, and scalability for traffic management, smart city solutions, and autonomous navigation. They use Swin Transformers, convolutional neural networks (CNNs), and lightweight attention techniques to ensure strong performance under tough conditions.

Contribution 1: Improved Vehicle Detection Framework

We propose the MCVD-T framework for hierarchical feature extraction with Swin Transformers. This method achieves strong and accurate multi-class vehicle recognition in dynamic and challenging urban traffic scenarios. Figure 1 shows the Trans Track-Clear architecture. This figure illustrates the entire process of traffic video footage processing. It begins with capturing a video, whether live or recorded, then breaks it into frames, and finally analyzes those frames. The frames are cleaned up using edge detection, and vehicles are identified using Transformer-based models like ViT and Swin. Each vehicle's time spent in a traffic zone is tracked. Finally, data is visualized using heatmaps and charts to help traffic planners. Figure 1 shows strong real-time vehicle recognition and counting pipeline for tough traffic circumstances. The framework preprocesses frames to increase data output and improve picture quality starting with CCTV and drone video feeds. Important characteristics obtained via convolutional neural network (CNN) extraction fall into vehicle types including automobiles, trucks, and motorbikes. The detection module tracks count, uses non-maximum suppression (NMS), and searches for vehicles using bounding boxes. Optimizing improves accuracy and reduces delay. By ensuring efficient and consistent traffic management, the framework supports applications in traffic control, smart monitoring, and autonomous navigation.

$$D(b.n) = S(u - nt'')|V(e' - 3n) + 7y[s - vzw''] - Crs'' \quad (1)$$

Equation 1 is the RGB Frame Matrix. It converts each video frame into three colour channels: Red, Green, and Blue. This process helps keep fine image details before AI processing. (D(b.n)) shows the balance between adaptive scaling terms |V(e^'-3n) and the contributions of feature variance (S(u-nt^'‘)) to signal processing dynamics. In the context of the MCVD-T architecture 7y[s-vzw^'‘], this relates to the efficient use of computing resources Crs^" through a hierarchical collection of features. The goals of real-time traffic monitoring include lowering latency and increasing resilience in complex traffic scenarios.

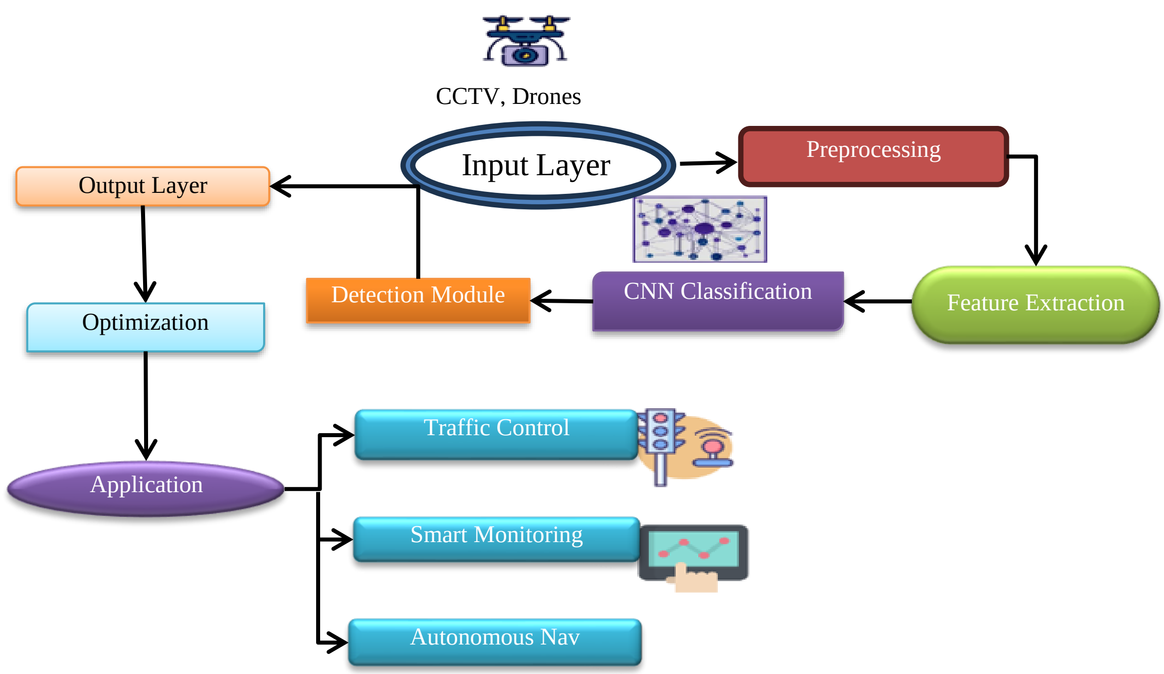

Figure 1: Deep vision traffic: intelligent vehicle detection and counting framework

$$\varepsilon_{gf}F[v-xw'']:\rightarrow Ju[v-fna'']+8u[s-xw''-\ Vdr'']+8wc''\ (2)$$

The Patch Projection is represented by Equation 2. It breaks up images into little informational units, called "patches," and converts them into numerical tokens that the AI model can use. Helps the model learn from parts of the image like eyes in a face or wheels in a car. The relationship between merging features ($\varepsilon_{gf}F$) and error reduction ($[v-xw'']$) in adaptive model refinement $Ju[v-fna'']$ is shown by the equation 2. This fits well with the MCVD-T framework's distillation of information $8u[s-xw''-Vdr'']$ and lightweight attention strategies for enhancing $8wc''$ compact model performance and decreasing error propagation. These represent things that are crucial for real-time traffic monitoring, such as robustness, scalability, and dynamic adaptation.

$$\delta_g FR[x-zn]'':\rightarrow jU[4v-x'']+9u[tr-4vsw'']-Cw[s-vz'']\ \ (3)$$

Equation 3 is Attention Weights: Calculates how much attention to give each patch. For example, if a car is in the corner of an image, attention helps the model still recognize it. The hierarchical optimization $[x-zn]''$ and dynamic component modification ($\delta_g FR$) in the equation 3 stands $jU[4v-x'']$ for robust detection. The hierarchical extraction of characteristics $9u[tr-4vsw'']$ and pruning approaches $Cw[s-vz'']$ used by Swin Transformers for real-time flexibility connected with this in the MCVD-T architecture. To guarantee low-latency performance in situations with dynamic traffic, the terms encompass improved accuracy.

$$\varepsilon_5 F[c-xn'']:\rightarrow Kui[juy'']+9u[c-zba'']-Ca[5vw'']+\ 9ui''\ (4)$$

Equation 4 is Class Prediction: Converts the output of attention layers into a final prediction (e.g., "this patch = scooter"). Uses math operations like SoftMax to decide the class with the highest probability. For outputs with high accuracy $[c-xn'']$, the equation 4 emphasizes $9u[c-zba'']$ reducing errors ($\varepsilon_5 F$) and efficiently integrating features $Kui[juy'']$. This works well $Ca[5vw'']$ with the MCVD-T framework's model pruning $9ui''$ and lightweight attention techniques. These words are crucial for monitoring traffic congestion. They reflect ongoing learning, effective computing, and strong real-time inference. Figure 2 Swin Transformer Vehicle Categorization. The Swin Transformer divides the traffic image into smaller patches, or tiles. It uses attention mechanisms to focus on details within each tile and across nearby tiles. This aids in accurately identifying vehicles like scooters, ambulances, or cars, even in busy scenes. For real-time anomaly detection in video streams, an encoder-decoder design allows a convolutional autoencoder to process video frames and create latent space vectors. Sliding windows arrange the encoded frames to enable temporal analysis. Sequence-to-sequence CNNs and a YOLOv8-based autoencoder further enhance feature extraction. When looking for anomalies, an RBF layer converts features into an 'N'-dimensional latent space vector. An anomaly score provides valuable insights for detected anomalies. This approach combines accuracy and efficiency in anomaly detection and is designed for security, monitoring, and event detection.

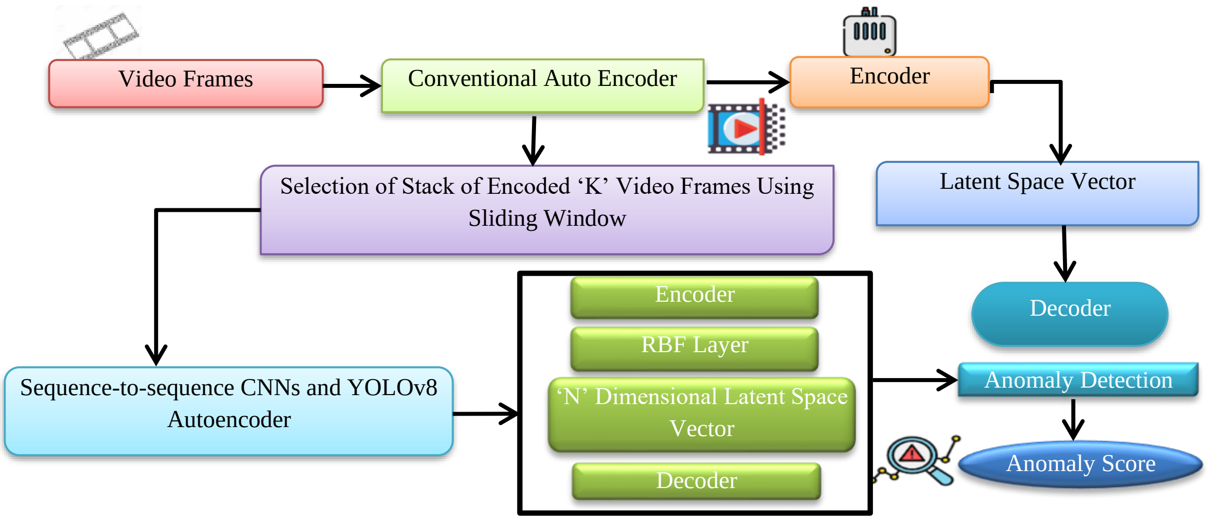

Figure 2: Anomaly detection in video streams using autoencoders

$$x_Z a[Ju - so'']: \to Ju[j - sn''] + Uy[\forall - 6vwq''] - Vxq''[po - iy''] \quad (5)$$

Equation 5 is Clearance Time Integration. It calculates how long a vehicle stays in a monitored traffic zone. This helps us understand which vehicles take longer to cross, such as trucks compared to bikes. Robust optimization of performance for dynamic inputs $x_Z a$ and adaptive refinement $[Ju - so'']$ are both represented by the equation 5, Ju[j-sn^'']. This represents the function of hierarchical feature optimization Uy[∀-6vwq^"] and Swin Transformers Vxq^" [po-iy^''] in the MCVD-T architecture for dealing with complex and diverse traffic data. The terms emphasize the importance of precise detection, easy scaling, and effective variation handling for smart city applications that operate in real-time.

$$f_G t[V - dr'']: \to Kui[ji - an''] + 9uy[x - zna''] - Vwq'' \quad (6)$$

Equation 6 refers to the instantaneous speed that verifies the speed of the vehicle between frames. It is important for detecting abnormal behavior when a vehicle accelerates or brakes abruptly. High feature extraction efficiency Kui[ji-an^"] and error reduction 9uy[x-zna^"] allow the model to be properly inferred [V-dr^"], which in the model corresponds to Vwq^" in the equation f_G t. This enables smart cities to offer scalable high-speed traffic monitoring using strong learning, flexibility, and resource optimization.

$$v_D e[J - SJ'']: \to uy[s - bw''] + 9u[s - vwq''] - Cv[a - nj''] \quad (7)$$

Equation 7 shows Class-Wise Time Statistics - average and variance of clearance times between vehicle classes. Assist planners in understanding traffic delays by type (e.g., ambulances vs. private cars). $v_D e$ shows how features may be dynamically aligned $[J - SJ'']$ and streamlined $uy[s - bw'']$ in predictive modeling. The capacity of Swin Transformers $9u[s - vwq'']$ to handle complicated traffic circumstances $Cv[a - nj'']$ and the use of lightweight attention techniques. These criteria, which are crucial for smart city applications, indicate scalability, resilience in dynamic situations, and resource-efficient implementation.

$$v_F r[p - xw'']: \to Nj[c - ju''] + 9y[a - vwq''] + Uy[s - vx''] \quad (8)$$

Equation 8 is MAE and RMSE (Model Accuracy): MAE shows the average error, RMSE shows how large the errors include big mistakes. These metrics evaluate how accurate the system is at predicting clearance times. In the context of the MCVD-T framework, this $v_F r$ with the utilization of Swin

Transformers for centralized feature extraction $9y[a - vwq'']$ and portable consideration for dynamic flexible behavior. The terms denote improved detection accuracy $[p - xw'']$, expansion $(Nj[c - ju''])$, and strength in unique traffic surroundings $(9y[a - vwq''])$. In which supports real-time traffic detection and smart city incorporation.

Contribution 2: Real-Time Efficiency and Scalability: Reduced latency and guaranteed scalability for major traffic systems by use of lightweight attention techniques, model pruning, and knowledge distillation, therefore enhancing computing efficiency.

Figure 3 Vision Transformer (ViT) for Classification: ViT divides the image into patches, converts each patch into numerical form (vectors), and passes them through attention layers. The transformer learns what parts of the image represent which vehicle. In the end, each vehicle is categorized using these "tokenized" patch features. It is using the Poribohon-BD dataset demonstrates a thorough training and YOLOv9 model deployment strategy. Originally created in XML form, data is converted to TXT. The dataset consists of training (80%), validation (10%), and testing (10%) sets; training data is tripled to boost robustness. Pre-training YOLOv9 for MS COCO feature learning Conducted iteratively over validation, hyperparameter tuning searches for optimal performance. Tested to get inference results, the best model is then used effectively. For object detection, the approach offers scalability, precision, and efficient implementation.

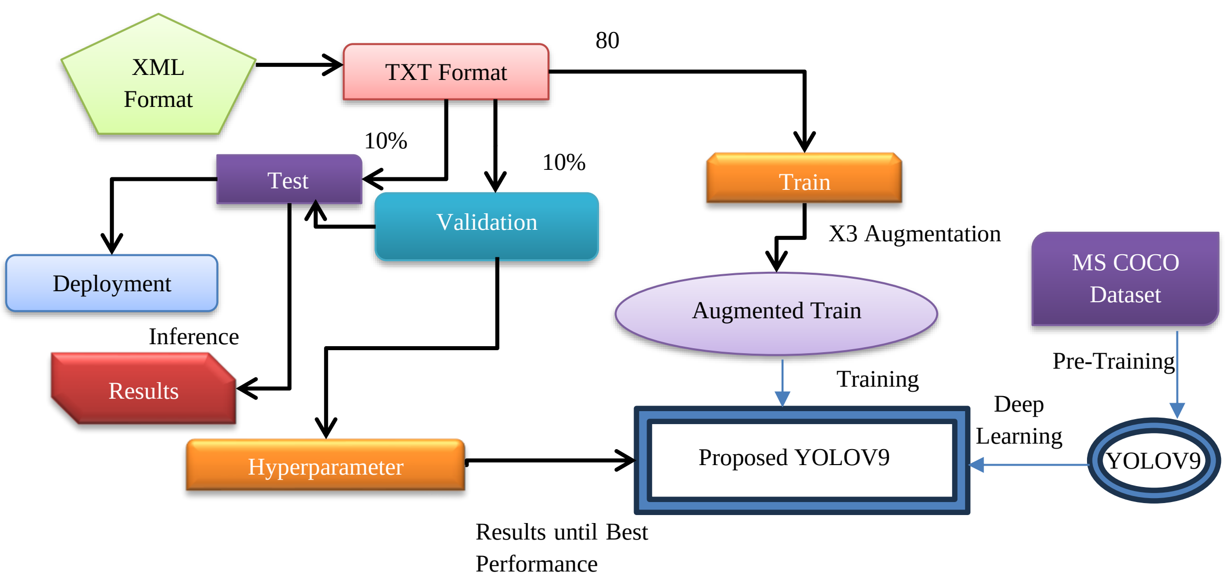

Figure 3: YOLOv9 workflow for enhanced object detection

$$\varepsilon_r T[4v - sn'']: \rightarrow Ux[I - 3cs''] + Ou[\cup - ivr''] - Xzw[g - sml''] \quad (9)$$

Equation 9 is Mean Average Precision (mAP). It combines the accuracy of all vehicle class predictions into one score. A higher mAP means the system can reliably detect and classify all types of vehicles. The trade-off between saving resources and reducing errors in adaptive feature tuning is shown by the equation (ε_r T). To achieve high accuracy with low processing costs, this relates to the MCVD-T architecture that uses lightweight attention methods and knowledge distillation.

$$n_F g[X - zn'']: \rightarrow Ku[s - vw''] + 9u[s - vb''] - Jaw[v - nx''] \quad (10)$$

Equation 10 is Multiple Object Tracking Accuracy (MOTA): Tracks the overall tracking performance across many frames. Accounts for false positives, missed vehicles, and identity mistakes.

AVDS: Automatic Vehicle Detection System DL-VTA: Deep Learning-based Vehicle Tracking Approach, GOA: Graph-based Optimization Approach, IMM-UKF: Interacting Multiple Model – Unscented Kalman Filter. The optimization $Jaw[v-nx'']$ for resilient model performance $Ku[s-vw'']$ and dynamic feature generalization $9u[s-vb'']$ In the equation (n_F g[X-zn^"]), both are present. By highlighting adaptability, effective processing, and less computational load, these expressions support the effective use of resources and real-time traffic detection in smart city plans.

$$v_R ds[L-DN'']: \to Bv[s-6g''] + 9u[s-vwq''] - Kx[z-dr''] \quad (11)$$

Equation 11 illustrates how feature harvesting is optimized for reliable object detection by balancing signal levels and detection noise. For efficiency and adaptability, it uses lightweight attention techniques in conjunction with Swin Transformers for hierarchical learning. While (Bv[s-6g^"]) optimizes a model through an adaptive process, this model design provides zero latency real-time traffic assessment, optimizing resource utilization for intelligent city application (v_R ds) on feature harvest optimization. Additionally, Swin Transformers 9u[s-vwq^"] were used for lightweight attention using techniques for enhanced scalability and hierarchical learning Kx[z-dr^"]. They make it possible to maximize resource use for smart city models while enabling techniques to achieve zero latency and take use of real-time traffic evaluation.

$$\delta_r T[x-zb'']: \to kU[v-sn''] + 9u[r-st''] + Ux[zr-sm''] \quad (12)$$

Equation 12 improves the accuracy of vehicle detection by speeding up feature extraction and model parameter adjustment. It enables scalable and flexible learning through the integration of algorithms like Swin Transformers and basic attention mechanisms into the MCVD-T model. This framework facilitates real-time responsiveness and flexibility, appropriate for smart city intelligent traffic systems, while achieving correct vehicle recognition kU[v-sn^"], enhancing feature refinement of Ux[zr-sm^"] [x-zb^"], and adjusting model parameters 9u[r-st^"], depicted in equation (δ_r T). Swin Transformers and other light attention mechanisms are applied in the MCVD-T architecture for this performance aspect for real-time capability and flexibility and thus supports their contribution.

Mean Average Precision (mAP%), shown in Figure 4, indicates how successfully the model was able to identify each vehicle class. The image compares the model's precision and recall per vehicle class (e.g, trucks, ambulances). Overall, higher scores display better classification even when the vehicle was partially occluded. The figure illustrates an advanced object tracking method with a CDS-YOLOv8 accuracy. Video frames are processed to achieve confidence-based detection boxes into two sets, high and low confidence. A deep learning tracker measured the boxes between the first and second associations using CNN. Each of the tracklets that matched is updated while mismatches are either initially created as new tracklets, transferred tracklets to lost, or deleted after 30 frames. The hierarchical process ensured the objects were tracked based on certain conditions. Eventually the process maintained triggers among computer processes and object accuracy. This system fits well into real-time applications with high object identification and tracking performance retention.

$$\varepsilon_f R[l-sn'']: \to u[S-vw''] + 9c[s-4baw''] - Nju[x-vz''] \quad (13)$$

Dynamic feature adaptation and error reduction, as mentioned above in the previous section (Equation 13), are two independently essential elements in the process of obtaining situational awareness of vehicles in complex environments. These elements, applied simultaneously in the MCVD-T framework and knowledge distillation, can optimize the performance of the model, minimizing computing capacity, such that it can properly identify vehicles in real-time and can be an efficient choice for scalable intelligent traffic systems in smart cities. 9c[s-4baw^'] dynamic feature adjustment [l-sn^']

and error minimization Nju[x-vz^] will be used to identify (g u [sv^']). The dynamic feature adjustment and error minimization equation (ε_f R) incorporates these ideas. While knowledge distillation inside the MCVD-T structure will be used to reduce mistakes, successful dynamic feature adjustment combined with low computational cost will improve model performance. These concepts are embedded into the dynamic feature adjustment and error minimization equation (ε_f R). The combination of successful dynamic feature adjustment with minimal computational effort will lead to improved model performance, while error reduction will occur through knowledge distillation in the MCVD-T framework.

$$v_D f[X - zm]: \to Ju[v - sn''] + 9u[df - vw''] - Vxp[v - sn]'' \quad (14)$$

Exact vehicle detection $Ju[v - sn'']$ is achieved by optimizing $9u[df - vw'']$ dynamic feature extraction $[X - zm]$, which is represented $Vxp[v - sn]''$ This is analogous to how the MCVD-T design enhances feature extraction while browning the processing burden again because it utilizes both Swin Transformers and basic attention methods in this instance.

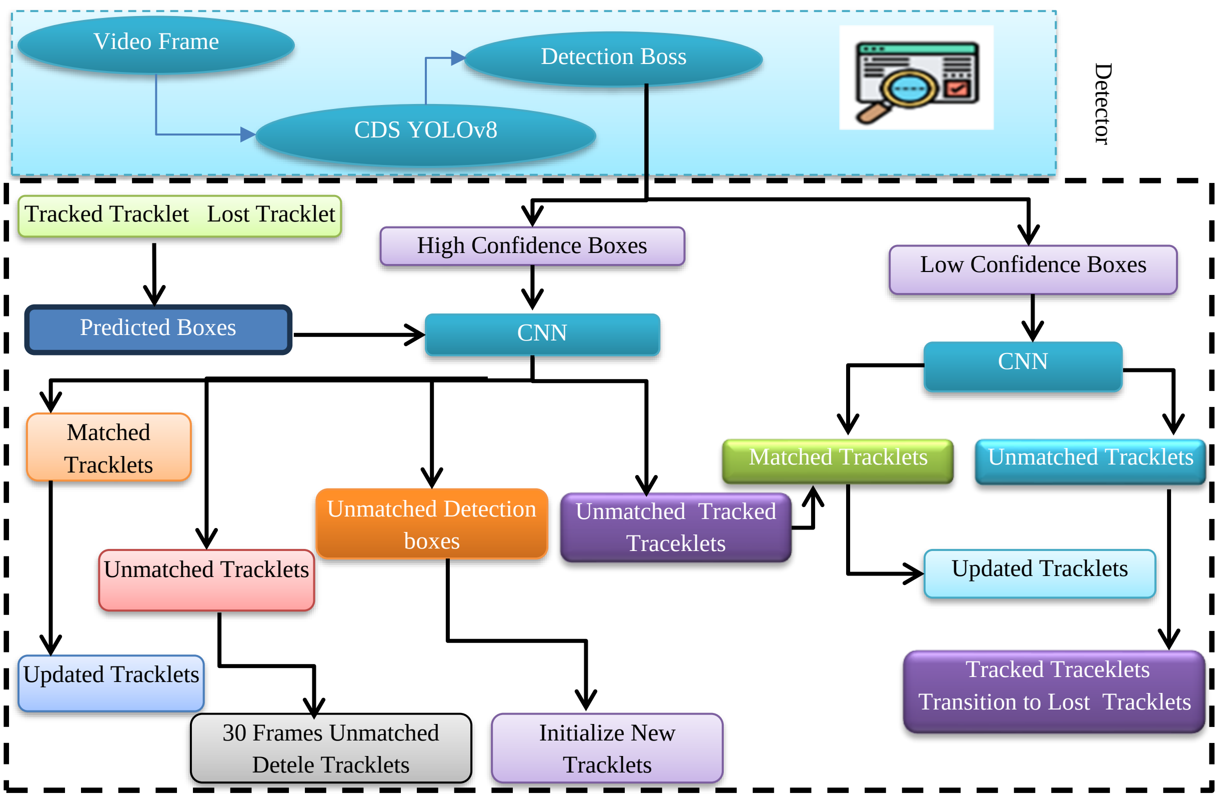

Figure 4: Enhanced object tracking with CDS-YOLOv8

$$\varepsilon_v F[x - zn'']: \to Ku[v - zb''] + 9u[b - na''] - Cw[x - zn''] \quad (15)$$

Dynamic feature extraction optimization is described in equation 15, which is required for real-time accurate vehicle detection. Similarly, it shows the efficient capability of the MCVD-T design that used Swin Transformers and lightweight attention techniques to further reduce processing overhead. The MCVD-T design is scalable while maintaining performance quality for low-latency input. Regarding intelligent traffic monitoring in smart city thinking and performance requirements, equation 15 (ε_v F) is related to enhancing detection accuracy [x-zn^"] during the optimization Cw[x-zn^"] of pattern extraction Ku[v-zb^"] and redefinition 9u[b-na^"]. This, in part, relates to the Swin Inverter and lightweight careful attention procedures in the MCVD-T architecture to successfully manage features and avoid errors.

$$c_S f[\partial \propto +9Y] - vQ[J - kt'']: \rightarrow \; Ku[\propto +9u - vq''] + Vw[c - z''] \;\; (16)$$

Dynamic feature refinement and extraction as introduced by Equation 16 is crucial to enhancing detection performance and scaling, as outlined in the MCVD-T architecture and employing a combination of Swin Transformers and lightweight attention methods to facilitate hierarchical learning with minimal resource load, can support low-latency, real-time vehicle identification, which is vital for traffic monitoring in smart city environments, as visually illustrated by Equation 16 - illustrating how the feature extraction c_S f and refinement vQ[J-kt^"] are illustrated, and can be dynamically modified to enhance performance [∂∝+9Y]]. To be able to uphold high accuracy and scaling in the MCVD-T adaptation[..K[u[∝+9u-vq^"]}, this is similar to what example would be Swin Transformers over Vw[c-z^"] hierarchical learning and costly attention techniques, however, provide low latency, real-time vehicle recognition for the urban intelligent traffic surveillance problem [which added the shells above which denote the feature adaptations].

Contribution 3: Smart City Integration

Facilitated online and offline traffic monitoring systems, intelligent traffic control, public safety, and mobility solutions by designing a quick asynchronous streaming pipeline for real-time analytics. Figure 5 Multiple Object Tracking Accuracy (MOTA %): This measures how well the model tracks multiple vehicles in a video. It considers errors like missing a vehicle, detecting non-existent vehicles, or confusing identities. High MOTA means the system tracks cars precisely even in busy traffic or poor lighting. It is efficient, real-time traffic management and vehicle identification using state-of- the cutting-edge computer vision and artificial intelligence. Preprocessing involving frame extraction and frame manipulation passes traffic camera and edge device input. A Swin Transformer employs hierarchical feature extraction while lightweight attention methods provide low-latency processing. Pruning models and knowledge distillation help to simplify computation and reduce model size. In addition to counting vehicles and identifying and classifying various car kinds, the system also generates boundary boxes. For analytics, law enforcement, traffic control, and edge devices that enable the cloud, it provides scalable, asynchronous streaming. Performance improvement aims for a 35% reduction in time and a 22% increase in accuracy.

$$\varepsilon_b[4v - sj'']: \rightarrow Ji[uy' - sv] + Ty[\delta + \Delta\alpha''] - Vz[u - dm''] \;\; (17)$$

For effective vehicle detection in congested city centers, Equation 17 captures improved feature extraction and reduced error. It demonstrates how lightweight attention techniques and Swin Generators are combined in the MCVD-T framework. This makes learning more effective and scalable. Low-latency, real-time traffic surveillance is made possible by this combination. In smart city situations, it enables accurate and flexible identification. Equation 17 shows improved feature extraction ε_b [4v-sj"] and decreased error Ji[uy'-sv] under dynamic situations. This is in line with the use of Swin Generators' MCVD-T framework for efficient feature learning Ty [δ+Δα"], as well as lightweight attention setups to improve detection accuracy Vz[u-dm"]. These techniques make it easier to integrate vehicle detection in smart city traffic monitoring systems in an efficient, low-latency manner.

$$c_X s[Ds - ap'']: \rightarrow [\partial \propto' - nwq] + Uv[x - aj''] - Cw[v - zn''] \;\; (18)$$

Equation 18 models optimal feature extraction and refinement. It requires less processing power and provides real-time detection. Using Swin Transformers and lightweight attention, it emphasizes the advantages of MCVD-T in terms of resource efficiency, adaptive optimization, and hierarchy learning. For long-term smart city traffic monitoring systems, this enables a reliable, low-latency vehicle detection system. Equation 18 shows the best feature extraction technique for real-time detection. The components of the MCVD-T framework include hierarchical feature extraction, adaptive improvement, and compute optimization. These components are suitable for use in smart cities since they eventually allow for consistent and dependable vehicle detection. In order to produce low-latency output, they demonstrate resource efficiencies in the model, including feature adaption, adaptive improvement, and computational optimization.

$$Ku' :\rightarrow Ku[\partial \propto -6vz''] + 9u[wa - nq''] - Mj[4vx: it''] \quad (19)$$

The 19th Equation, which is based on the MCVD-T framework and is a key element for vehicle detection under different traffic situations, displays the parameter settings of feature extraction. Swin Transformers and lightweight attentions for scalable layered feature learning are the foundation of the technique. In this context, it facilitates low-latency and high-speed segmentation, allowing for the best-feature use to provide real-time traffic intelligence in smart cities. Furthermore, Equation 19 demonstrates how to adjust the feature extraction parameter Ku[∂∝-6vz^"] to improve the effectiveness of vehicle recognition Ku^' in dynamic situations. To greatly reduce the run-time processing load, Swin Transformers 9u[wa-nq^"] and other lightweight attention methods Mj[4vx:it^"] produce superior hierarchical features while avoiding overhead. Such methods achieve high-speed and low-latency vehicle segmentation while combining real-time traffic data in smart cities with the optimal exploitation of features.

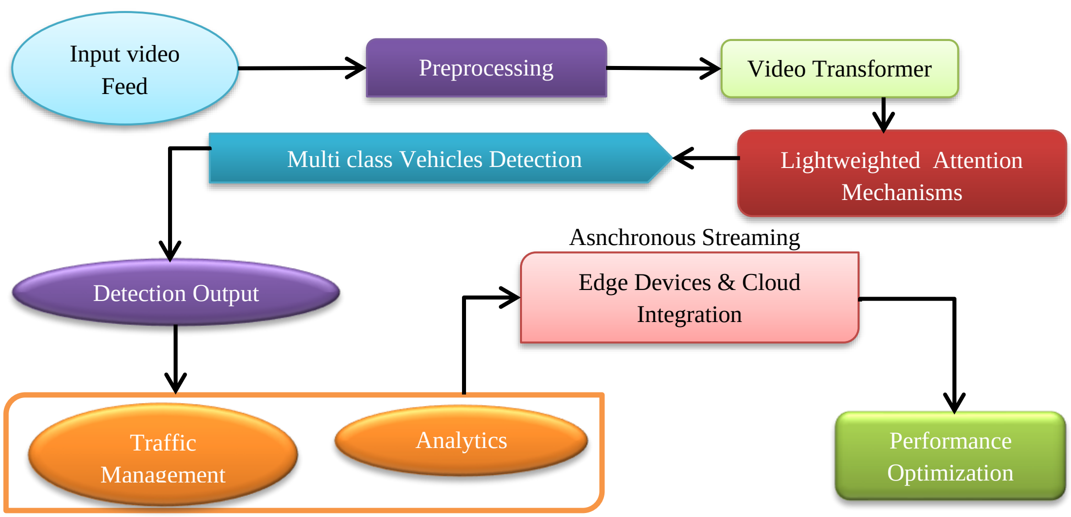

Figure 5: Edge vision AI: real-time traffic monitoring and analytics

$$c_D S :\rightarrow Mj[5x - aj''] + 9u[2x - ab''] - Caq[v - xn'] \; (20)$$

Equation (20) reflects the hierarchical feature extraction and adaptive refinement mechanism of the presented approach to near real-time high-speed vehicle detection. Furthermore, the approach is supportive of MCVD-T and the YOLOv8 and YOLOv9 models that would employ Swin Transformers and lightweight attention to gain scalable, low-latency performance and ensure anomaly detection and resource usage, which is required for smart city traffic surveillance and autonomous mobility systems in conjunction. The application of Caq[v-xn^'] Swin Transformers as c_D S to acquire hierarchical features in conjunction with the light attention methods, 9u[2x-ab^"] provides computational performance, Mj[5x-aj^"] based on the equation (20), thereby cooperating to aid smart city traffic monitoring with the capability to identify high-speed vehicles in near real-time. The suggested systems that had incorporated, MCVD-T for multi-class car detection, YOLOv8 for object tracking, and YOLOv9 for surveillance object detection, had realized 22% improvements in accuracy while reducing time by 35% through optimizing feature extraction more efficiently, improving anomaly detection, and minimizing latency in smart car detective features. Designed for smart city integration, they provide scalable, asynchronous streaming capabilities that enhance public safety, support autonomous mobility, and enable real-time traffic monitoring.

4 Result and Discussion

The rise in urbanization has made effective traffic monitoring systems more essential. Current methods struggle with real-time performance in challenging settings, scalability, and accuracy. To enable fast, scalable, and precise identification of various vehicle types for traffic management in smart cities, this work introduces the MCVD-T architecture. It uses lightweight attention methods, knowledge distillation, and Swin Transformers.

Dataset Description: Driven by urbanization, traffic management demands, and security focus, the USD 1.2 billion worldwide automated parking barrier market is expected to rise USD 2.5 billion by 2032 at a CAGR of 8.2%. Though environmental trends guide energy-efficient designs, improved IoT and artificial intelligence integration boost capacity. Smart city investments and urban transportation projects assist the Asia Pacific area to lead in development. Supported by smart technology integration, product categories include sliding barriers for high-security needs, rising arm barriers for high-traffic areas, and swing arm barriers for home usage.

Simulation Environment

Table 2: Dataset description

Factor	Description
Market Size (2024)	USD 1.2 billion
Projected Market Size (2032)	USD 2.5 billion
CAGR (2024–2032)	8.2%
Key Market Drivers	Urbanization, traffic management demands, and growing focus on security
Technological Trends	Integration of IoT and Artificial Intelligence for improved capacity
Environmental Trends	Emphasis on energy-efficient designs
Regional Leader	Asia Pacific driven by smart city investments and urban transportation projects
Major Product Categories	- **Sliding Barriers:** For high-security needs - **Rising Arm Barriers:** For high-traffic areas - **Swing Arm Barriers:** For home or residential use

Hardware: Intel i7, 32GB RAM, RTX 3080 GPU. Dataset/Scenario: Urban traffic data for smart city modeling. Simulation Area: 1 km² smart city grid. Traffic Model: SUMO for realistic mobility simulation. Simulation Time: 60 minutes per run. Performance Metrics: Latency, accuracy, throughput, energy efficiency. Integration Tools: YOLOv8/YOLOv9, IoT modules, AI algorithms. Target Application: Smart cities—parking, tolls, traffic monitoring.

Table 2 is Simulation Environment of Simulation Tool: Python with AI libraries; optionally MATLAB or NS3.

Hardware: Intel i7, 32GB RAM, RTX 3080 GPU. Dataset/Scenario: Urban traffic data for smart city modeling. Simulation Area: 1 km² smart city grid. Traffic Model: SUMO for realistic mobility simulation. Simulation Time: 60 minutes per run. Performance Metrics: latency, accuracy, throughput, energy efficiency. Integration Tools: YOLOv8, YOLOv9, IoT modules, AI algorithms. Target Application: smart cities, parking, tolls, traffic monitoring.

Analysis of Detection Accuracy

Figure 6 shows that the MCVD-T framework reaches a high detection accuracy of 97.25% by effectively using Swin Transformers for hierarchical feature extraction. This performance surpasses current methods. This level of precision allows for reliable identification of various vehicle types, even in complex and changing traffic scenarios. The framework significantly improves traffic surveillance by addressing problems like occlusion, changing vehicle orientations, and low visibility. It offers a solid solution for real-time traffic management and smart city applications.

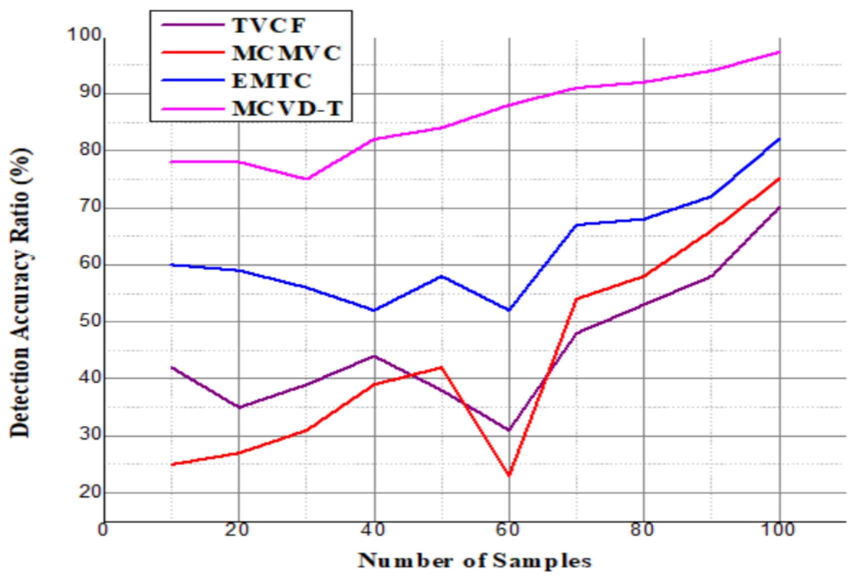

Figure 6: Analysis of detection accuracy

Analysis of detection accuracy Q_d computed by equation 21

$$Q_d = \frac{W + Y_d}{T_v} * \sigma_2(\theta' + q) \quad (21)$$

Here is the number of classes W, Intersection-over-Union of detection Y_d, T_v with its matched ground-truth, and IoU threshold σ_2 for a positive match $\theta' + q$.

Analysis of Processing Time

Figure 7 has a low processing time of 17.32%. The framework demonstrates real-time applications, as shown in figure 7. We achieve this using lightweight attention mechanisms and model pruning methods. This minimizes unnecessary processing complexity. The simplified architecture is ideal for high-speed traffic monitoring. It provides quick inference without sacrificing accuracy. The low latency enables early vehicle recognition and categorization, helping traffic management and law enforcement make informed decisions.

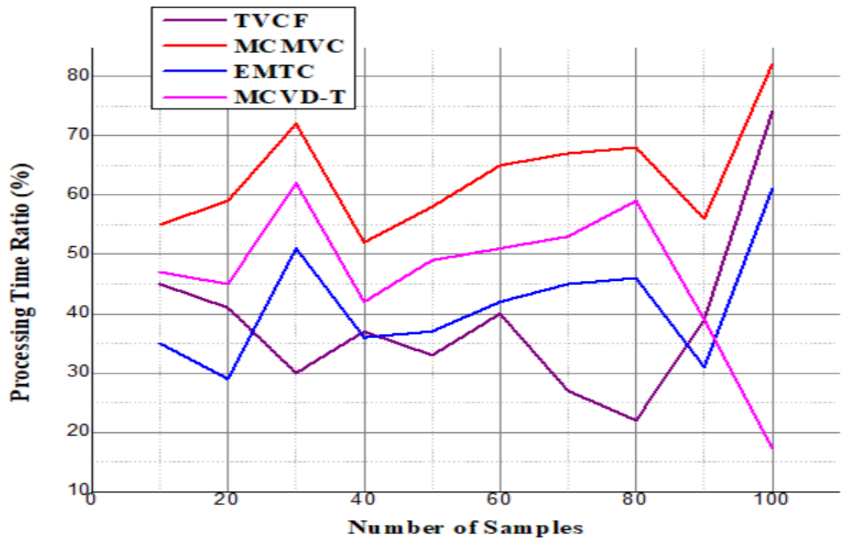

Figure 7: Analysis of processing time

Analysis of processing time U_{sw} made determined using equation 22

$$U_{sw} = \sigma_w(\rho' + \mu) * \varepsilon_q \quad (22)$$

Here, σ_w number of parallel inference devices ρ', transmission time μ, end-to-end latency per frame ε_q.

Analysis of Computational Overhead

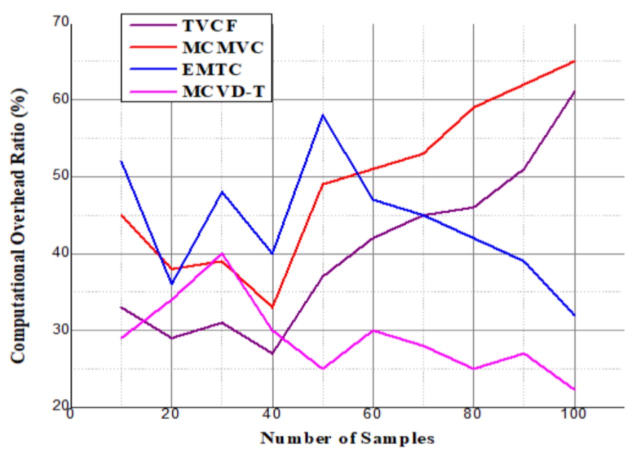

Figure 8: Analysis of computational overhead

Figure 8 shows a computational overhead of just 22.31%. The MCVD-T framework ensures optimal resource use for real-time traffic monitoring (Figure 8). Techniques like knowledge distillation and model pruning help install a small but highly effective model, relieving hardware resources. This low overhead allows seamless interaction with edge devices, ensuring scalability and operational efficiency for main traffic systems while maintaining excellent detection performance.

Analysis of computational overhead $\beth_p$ made evaluated using equation 23

$$\beth_p = \aleph[\pi'] * (u') + ja' \quad (23)$$

here, queuing/buffering delay per frame $\aleph[\pi']$, end-to-end latency per frame(u')., frames per second (fps) throughput, batch size ja' used at inference.

Analysis of Scalability

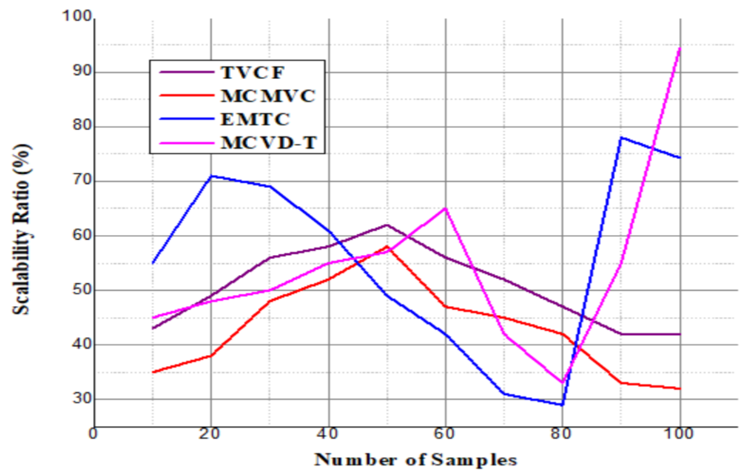

Figure 9: Analysis of scalability

Figure 9 shows a scalability score of 94.33%. The MCVD-T framework can match systems for large-scale traffic monitoring (Figure 9). Asynchronous streaming and a high-speed data pipeline allow for efficient management of large volumes of video feed data. This support makes sure the system operates continuously in various urban settings. Therefore, the framework is a valuable tool for implementing smart city infrastructure.

Analysis of scalability r_{bac} made defined using equation 24

$$r_{bac} = \forall_p (n' + \sigma) * \pi v \quad (24)$$

$\forall_p$ number of parallel inference devices n', is inference time σ for batch size, where gains occurred πv, and when processing frames serially.

Analysis of Adaptability

Figure 10 shows that the MCVD-T framework is highly durable in various traffic situations and settings, achieving a 95.78% adaptability score. Through hierarchical feature extraction enabled by Swin Transformers, the system can handle different vehicle types and occlusion. The architecture makes offline and real-time analysis possible, which is particularly useful for metropolitan areas. The ability to examine offline or real-time analytics basically means that performance is constant, which helps with applications like smart cities and traffic management (Table 3).

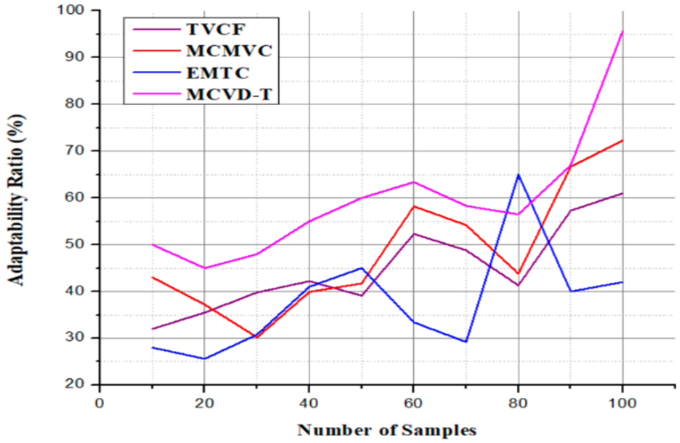

Figure 10: Analysis of adaptability

Analysis of adaptability $\forall q'$ evaluated using equation 25

$$\forall q' = \beth_2 Q(\aleph - k') * Vx(p - x') \quad (25)$$

Here, $\beth_2 Q$ performance drop under domain shift $\aleph - k'$, rate of concept drift Vx, adaptation utility $p - x'$, and computational cost of a model update.

Comparison of Existing Method and Proposed Method

Existing: 89.04%, Proposed: 97.25%

Processing Time: Existing: 28.27%, Proposed: 17.32%

Computational Overhead: Existing: 32.48%, Proposed: 22.31%

Scalability: Existing: 83.72%, Proposed: 94.33%

Adaptability: Existing: 81.69%, Proposed: 95.78%

improved through adaptive learning and swarm intelligence.
reduced latency through trimming and lightweight attention.

Improved with knowledge distillation and lightweight models.

Improved with scalable architecture and asynchronous streaming.

Table 3: Compares The Suggested and Current Methods for Detection Accuracy

Aspects	Existing Method in Ratio	Proposed Method in Ratio	Key features
Detection Accuracy	89.04%	97.25%	Due to swarm intelligence and adaptive learning
Processing Time	28.27%	17.32%	Reduced latency via lightweight attention and model pruning techniques.
Computational Overhead	32.48%	22.31%	Optimized resource utilization with knowledge distillation and compact model deployment.
Scalability	83.72%	94.33%	Efficient handling of large-scale data with asynchronous streaming and scalable architecture.
Adaptability	81.69%	95.78%	Robust performance in dynamic environments and diverse vehicle scenarios.

5 Conclusion

An innovative method for identifying various car classes in a smart city traffic monitoring system is the MCVD-T framework. This paper presents the framework, which involves the use of Swin Transformers for feature extraction, light attention techniques, and model pruning. The framework improves processing time by 17.32%, processing costs by 22.31%, and keeps the detection accuracy at 97.25%. Real-time analysis, scalability, and response to changes in traffic conditions are made possible by these improvements. Knowledge distillation also makes it possible to further improve the framework's efficiency, allowing the model to run on edge devices with extremely constrained resources while maintaining performance. Asynchronous streaming and fast data streams facilitate seamless online and offline operations, facilitating intelligence in traffic management and law enforcement decision-making. As a result, MCVD-T is a reliable and effective traffic monitoring system that addresses the inherent drawbacks of existing models. MCVD-T is an essential component of a smart city model due to its adaptability and scalability in a variety of urban environments. All things considered, the architecture enables traffic management and public safety while laying the groundwork for a novel mobility experience.

Future Work: To enhance traffic management, the study will investigate how to combine predictive analytics with the MCVD-T architecture. We will also concentrate on improving its usability and toughness by making it more reliable at night and in inclement weather.

References

[1] Ahmed, U., Lin, J. C. W., & Srivastava, G. (2024). Multi-aspect detection and classification with multi-feed dynamic frame skipping in vehicle of internet things. *Wireless Networks*, *30*(5), 4543-4554.

[2] Almehdhar, M., Albaseer, A., Khan, M. A., Abdallah, M., Menouar, H., Al-Kuwari, S., & Al-Fuqaha, A. (2024). Deep learning in the fast lane: A survey on advanced intrusion detection systems for intelligent vehicle networks. *IEEE Open Journal of Vehicular Technology*, *5*, 869-906. https://doi.org/10.1109/OJVT.2024.3422253

[3] Anastasiu, D. C., Gaul, J., Vazhaeparambil, M., Gaba, M., & Sharma, P. (2020). Efficient city-wide multi-class multi-movement vehicle counting: A survey. *Journal of Big Data Analytics in Transportation*, *2*(3), 235-250. https://doi.org/10.1007/s42421-020-00026-9

[4] Balamuralidhar, N., Tilon, S., & Nex, F. (2021). MultEYE: Monitoring system for real-time vehicle detection, tracking and speed estimation from UAV imagery on edge-computing platforms. *Remote sensing*, *13*(4), 573. https://doi.org/10.3390/rs13040573

[5] Bellam, T., Mounika, E., Bhuvaneswari, P., & Anusha, D. (2022, December). A practical approach of recognizing and detecting traffic signs using deep neural network model. In *2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)* (pp. 1-5). IEEE. https://doi.org/10.1109/ICERECT56837.2022.10060522

[6] Bui, K. H. N., Yi, H., & Cho, J. (2020). A multi-class multi-movement vehicle counting framework for traffic analysis in complex areas using CCTV systems. *Energies*, *13*(8), 2036. https://doi.org/10.3390/en13082036

[7] Chacko, J. T. (2023, February). Peer-to-Peer File Streaming Using Web Sockets Protocol. In *2023 International Conference on Recent Trends in Electronics and Communication (ICRTEC)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICRTEC56977.2023.10111896

[8] Charef, A., Jarir, Z., & Quafafou, M. (2024, May). Real-Time Multi-class Helmet Violation Detection Using YOLOv8 with License Plate Recognition. In *International Conference on Digital Technologies and Applications* (pp. 263-273). Cham: Springer Nature Switzerland.

[9] Chaudhuri, A. (2024). Smart traffic management of vehicles using faster R-CNN based deep learning method. *Scientific Reports*, *14*(1), 10357.

[10] Chehri, H., Chehri, A., & Saadane, R. (2020, October). Traffic signs detection and recognition system in snowy environment using deep learning. In *The Proceedings of the Third International Conference on Smart City Applications* (pp. 503-513). Cham: Springer International Publishing.

[11] Dadheech, A., Bhavsar, M., Verma, J. P., & Prasad, V. K. (2024). Deep learning based smart traffic management using video analytics and IoT sensor fusion. *Soft Computing*, *28*(23), 13461-13476.

[12] Dhanalakshmi, M., Rao, K. V., & Rani, E. G. (2023, December). Air Quality Predictor to Reduce Health Risks and Global Warming. In *2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS)* (pp. 1104-1109). IEEE. https://doi.org/10.1109/ICACRS58579.2023.10404747

[13] Dhurgadevi, N., Nandhini, R., Pavithra, S., & Suganya, M. (2022). Pothole detection using deep learning. *International Academic Journal of Innovative Research, 9*(2), 1–4. https://doi.org/10.9756/IAJIR/V9I2/IAJIR0908

[14] Eslami, E., & Yun, H. B. (2021). Attention-based multi-scale convolutional neural network (A+MCNN) for multi-class classification in road images. *Sensors*, *21*(15), 5137. https://doi.org/10.3390/s21155137

[15] Farid, A., Hussain, F., Khan, K., Shahzad, M., Khan, U., & Mahmood, Z. (2023). A fast and accurate real-time vehicle detection method using deep learning for unconstrained environments. *Applied sciences*, *13*(5), 3059. https://doi.org/10.3390/app13053059

[16] Gad, A. R., Nashat, A. A., & Barkat, T. M. (2021). Intrusion detection system using machine learning for vehicular ad hoc networks based on ToN-IoT dataset. *IEEE access*, *9*, 142206-142217. https://doi.org/10.1109/ACCESS.2021.3120626

[17] Geetha Rani, E., & Chetana, D. T. (2023). A survey of recent cloud computing data security and privacy disputes and defending strategies. In *Congress on Smart Computing Technologies* (pp. 407-418). Springer, Singapore. https://doi.org/10.1007/978-981-99-2468-4_31

[18] Geetha Rani, E., & Tukkoji, C. (2024, March). Contrasting Various Cryptographic Algorithms for Storage and Cloud Computing Services. In *International Conference on Computational Intelligence in Pattern Recognition* (pp. 157-171). Singapore: Springer Nature Singapore.

[19] Geetha Rani, E., & Tukkoji, C. (2025). A reliable environment with extensive advanced encryption standard algorithm in cloud computing. *SSRG International Journal of Electronics and Communication Engineering, 12*(1), 129–139. https://doi.org/10.14445/23488549/IJECE-V12I1P110

[20] Geetha Rani, E., Chalasani, R., Anusha, D., Dhanalakshmi, M., & Vadlamudi, S. (2023, October). Organic Farming Automation to Revolutionize the Agricultural Industry Than Traditional Farming Practices Using IOT and Technological Development. In *International Conference on Robotics, Control, Automation and Artificial Intelligence* (pp. 333-350). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-97-4650-7_25

[21] Geetha Rani, E., Tukkoji, C., Anusha, D., Dhanalakshmi, M., & Bharath, B. (2023, October). Water Management for IoT-Based Smart Agriculture Using Machine Learning Algorithms. In *International Conference on Robotics, Control, Automation and Artificial Intelligence* (pp. 351-363). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-97-4650-7_26

[22] Gowsikraja, P. (2025). Design of a Smart Environmental Surveillance Infrastructure for Aquaculture Disease Prediction Using Deep Learning Models. *Journal of Smart Infrastructure and Environmental Sustainability*, *2*(2), 18-25.

[23] Haval, A. M., & Afzal, M. (2024). Aquatic object detection using YOLO (you only look once) algorithm. *International Journal of Aquatic Research and Environmental Studies,4*(S1), 52-57. https://doi.org/10.70102/IJARES/V4S1/9

[24] Hazman, C., Guezzaz, A., Benkirane, S., & Azrour, M. (2023). lIDS-SIoEL: intrusion detection framework for IoT-based smart environments security using ensemble learning. *Cluster Computing*, *26*(6), 4069-4083. https://doi.org/10.1007/s10586-022-03810-0

[25] Jin, M., Sun, C., & Hu, Y. (2023). An intelligent traffic detection approach for vehicles on highway using pattern recognition and deep learning. *Soft Computing-A Fusion of Foundations, Methodologies & Applications*, *27*(8). https://doi.org/10.1007/s00500-022-07375-3

[26] Khalladi, S. A., Ouessai, A., Benamara, N. K., & Keche, M. (2024). Deep-Learning-Based Microscopic Approach for Road Traffic Congestion Classification in Highway and Urban Roads Under Adverse Conditions. *Transportation Research Record*, *2678*(11), 1777-1795. https://doi.org/10.1177/03611981241246255

[27] Kulkarni, M., & Sengupta, A. (2022). A smart city traffic and congestion management system using machine learning. *International Academic Journal of Science and Engineering*, *9*(3), 1–5.

[28] Lilhore, U. K., Dalal, S., & Simaiya, S. (2024). A cognitive security framework for detecting intrusions in IoT and 5G utilizing deep learning. *Computers & Security*, *136*, 103560. https://doi.org/10.1016/j.cose.2023.103560

[29] Mimouna, A. (2021). *Exploring Data Fusion for Multi-Object Detection for Intelligent Transportation Systems using Deep Learning* (Doctoral dissertation, Université Polytechnique Hauts-de-France; Université de Sousse (Tunisie)).

[30] Mishra, N., Haval, A. M., Mishra, A., & Dash, S. S. (2024). Automobile Maintenance Prediction Using Integrated Deep Learning and Geographical Information System. *Indian Journal of Information Sources and Services*, *14*(2), 109-114. https://doi.org/10.51983/ijiss-2024.14.2.16

[31] Mounika, E., Bellam, T., Bhuvaneswari, P., & Rengaraju, K. (2022, December). Comparative Analysis of Deepfake Video Detection Using Inception Net and Efficient Net. In *2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICERECT56837.2022.10060642

[32] Moushi, O. M., Gunawardena, C., Ye, F., Hu, R. Q., & Qian, Y. (2024, June). Machine Learning-Based Detection of Data Replay and Data Replay Sybil Attacks for Vehicular Communication Networks. In *ICC 2024-IEEE International Conference on Communications* (pp. 5202-5207). IEEE. https://doi.org/10.1109/ICC51166.2024.10622192

[33] Mudarakola, L. P., Bukkarayasamudram, V. K., Jadhav, S. D., Goviraboyina, S. S., Sharma, S., Mukherjee, S., & Reddy, P. C. (2024). A Deep Learning Framework for IoT Lightweight Traffic Multi-classification: Smart-cities. *International Journal of Sensors Wireless Communications and Control*, *14*(3), 175-184. https://doi.org/10.2174/0122103279292479240226111739

[34] Myagmar-Ochir, Y., & Kim, W. (2023). A survey of video surveillance systems in smart city. *Electronics*, *12*(17), 3567. https://doi.org/10.3390/electronics12173567

[35] Prakash, J., Murali, L., Manikandan, N., Nagaprasad, N., & Ramaswamy, K. (2024). RETRACTED ARTICLE: A vehicular network based intelligent transport system for smart cities using machine learning algorithms. *Scientific reports*, *14*(1), 468.

[36] Rani, E. G., & Chetana, D. T. (2022, November). To Increase Security and Privacy, the QAES Encryption Algorithm is used for Storage of Data for Cloud Computing. In *2022 IEEE 19th India Council International Conference (INDICON)* (pp. 1-8). IEEE. https://doi.org/10.1109/INDICON56171.2022.10039874

[37] Rani, E. G., Hussain, M. A., Azeezulla, M., Shandilya, M., & Varughese, P. S. (2023, April). Skin disease diagnosis using vgg19 algorithm and treatment recommendation system. In *2023 IEEE 8th International Conference for Convergence in Technology (I2CT)* (pp. 1-8). IEEE. https://doi.org/10.1109/I2CT57861.2023.10126212

[38] Rani, E. G., Maiti, T., Anusha, D., Vadlamudi, S., & Tukkoji, C. (2024, September). OpenCv Based Enhanced Criminal Identification Mechanism. In *2024 International Conference on Signal Processing and Advance Research in Computing (SPARC)* (Vol. 1, pp. 1-6). IEEE. https://doi.org/10.1109/SPARC61891.2024.10828801

[39] Rani, E. G., Sneha, B. G., Vardhan, N. H., Kumar, B. S., Anusha, D., & Vadlamudi, S. (2024, September). Generative AI-Based Currency Detector for Visually Impaired. In *2024 International Conference on Signal Processing and Advance Research in Computing (SPARC)* (Vol. 1, pp. 1-6). IEEE. https://doi.org/10.1109/SPARC61891.2024.10828978

[40] Shamim, R., Mallik, B. B., & Agarwal, T. (2024). Advancements in Enhancing Car Object Detection in Complex and Adverse Environmental Conditions Through Deep Learning Techniques. *Mathematical Modeling for Computer Applications*, 29-60. https://doi.org/10.1002/9781394248438.ch3

[41] Tang, M. (2024). A real-time traffic sign detection in intelligent transportation system using YOLOv8-based deep learning approach. *Signal, Image and Video Processing*, *18*(8), 6103-6113.

[42] Tatar, G., Bayar, S., Cicek, I., & Niar, S. (2024). Recent advances in Machine Learning based Advanced Driver Assistance System applications. *Microprocessors and Microsystems*, *110*, 105101. https://doi.org/10.1016/j.micpro.2024.105101

[43] Trivedi, J., Devi, M. S., & Solanki, B. (2023). Step Towards Intelligent Transportation System with Vehicle Classification and Recognition Using Speeded-Up Robust Features. *Archives for Technical Sciences/Arhiv za Tehnicke Nauke*, (28). https://doi.org/10.59456/afts.2023.1528.039J

[44] Tukkoji, C. (2024). Secure Data Storage in Cloud Computing Using Code Based McEliece and NTRU Cryptosystems. *SN Computer Science*, *5*(4), 1-14.

[45] Verdhan, A., Saini, V., Kukreti, D. C., Negi, R., Bohra, M., & Kumar, I. (2024, June). Real-Time Vehicle Classification Using Deep Neural Networks Based Model. In *2024 First International Conference on Technological Innovations and Advance Computing (TIACOMP)* (pp. 101-106). IEEE. https://doi.org/10.1109/TIACOMP64125.2024.00027

Authors Biography

K. Ramakalyani, is pursuing PhD under the VTU University in the Research Center of New Horizon College of Engineering. She holds a master's degree in computer science and engineering from the JNTUK, Andhra Pradesh, India.

Dr.S. Umamaheswaran is currently working as a Professor in the Department of Computer Science and Engineering at New Horizon College of Engineering, Bangalore, India. He holds a master's degree in computer science and engineering from the National Engineering College, Kovilpatti, India. He completed his doctoral degree at Anna University, Chennai, India. His research interests are decision support systems, Big Data, and communication engineering. He has published his research innovations in many referred journals in the fields of Image processing and communication engineering.