International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

A Multi-Scale Dataset With Comprehensive Analysis for Building Detection From Remote Sensing Images for Urban Development Identification

DR. Indhumathi R[1], N. Raja Harsha Vardhan[2], P. Jathin Dev[3]

[1,2,3]Department of Computing Technologies, SRM Institute of Science and Technology, Kattankulathur, India

Abstract
The monitoring of urban development remains vital because it enables authorized city planning and infrastructure control and sustainable expansion. This research delivers a solid building identification mechanism which leverages a multi- scale dataset while employing the YOLOv8 deep learning methodology. High-resolution aerial and satellite imagery makes up the main element in the dataset since it contains precise bounding boxes for buildings, roads, and trees throughout different types of urban areas. The YOLOv8 model efficiently detects objects of all sizes while functioning in real time with high accuracy levels through its optimized tuning method. The detection system evaluation relies on standard performance metrics which include precision, recall, mean Average Precision (mAP) and Intersection over Union (IoU) for a complete assessment of detection accuracy and reliability. Through geospatial analytics integration the system delivers detailed examination of urban growth patterns with actionable land use and infrastructure trends data analysis. These implementation methods deliver an expandable and effective answer to handle the difficulties of contemporary urban development.

Keywords: Traffic Light Detection, Visually Impaired Assistance, Real-Time Classification, Deep Learning.

Introduction

In recent times, Urban development has changed much, artificial intelligence (AI) along with machine learning (ML) acts as a most powerful 'game changing' technique in 'Smarter', 'Efficient' and 'Sustainable Cities'. Increasing urban populations and limited resources necessitate the use of powerful solutions provided by the AI and ML technologies to current problems faced by urban communities. These technologies allow a city to take advantage of massive amount of data from sensors, traffic systems, energy grids and other social media platforms to extract pertinent information. Traffic management, crime prevention, energy distribution, waste management and infrastructure planning are some of examples for which AI and ML are being used for predictive analytics. For example, ML models can anticipate traffic congestion, the ML models can optimise the path for public transportation, and AI's enhanced urban safety systems will detect anomalous situations, predict future threats. Sometimes, AI could analyze satellite images and population data to help with urban planning decisions on the type of zoning, land use and environmental conservation. Additionally, AI driven simulations enable planner to understand the long term impact of the development projects on urban ecosystems. Along with reducing energy waste with the integration of AI in smart buildings and homes, automatic adjustments of lighting, heating and cooling systems, depending on how the buildings and homes are used, help condense energy. Also, machine learning algorithms are key at maintaining environmental sustainability by the monitoring of the water usage, air quality and carbon emissions. As a whole, the application of AI and ML to urban development not only helps improve the operation efficiency and brings resilience, inclusivity and innovation, but also sets the stage for the development of the intelligent urban ecosystem.

LITERATURE REVIEW

In the work of Y. Du (2024), the Long Short Term Memory (LSTM) algorithm was used to perform traffic flow analysis. Short term traffic flow prediction was found to benefit enormously from the application of LSTM, as shown by the study. Model is trained on historical data and forecasted (with high execution efficiency) future traffic trends so traffic authorities can anticipate congestion, become more effective in their decision making and promote urban sustainable development.

According to M.S. Tanaka (2025), a digital twin computing method was developed for predicting general road traffic volumes. The proposed system is a simulation of real- time urban traffic dependent on sensed data and supplemented with machine learning techniques in case of limited inputs. Having minimal sensing infrastructure allows it to be used specifically for cities, and through simulation of urban environments, it also allows for strategic traffic control.

There are various approaches which they developed, the first is a hybrid model A. Sharma, S. Bhatnagar, and S. Dhariwal (2024) developed the model which is a combination between YOLO for the detection of vehicles and ARIMA for time series forecasting. It predicts congestion trends at intersections using historically detected vehicles traffic pattern. The hybrid framework enables a real time, accurate traffic surveillance with significant values to the urban traffic management systems.

Giganti et al. (2024) suggested a Graph Neural Network (GNN) based methodology to forecast NO_2 levels with past and future covariates. The model incorporates the environmental and the calendar related influences on air pollution using a spatiotemporal graph structure. It was shown that future covariates including weather forecasts can improve on the accuracy of prediction by traditional methods relying only on historical traffic data.

In this work of J. Zhang, J. Chen, G. Yang, C. Wu (2024), the Multi-Angle Attention Convolutional Neural Network (MACACNN) was used for traffic flow prediction. CNNs with coordinate attention mechanisms are used to analyze traffic data from front, side and top views in order to capture spatiotemporal dependencies in the model. The model was tested on the TaxiBJ dataset and the results were better than baseline methods on prediction accuracy, and it is able to predict more accurately across the more complex urban networks.

IIn 2024, K. Zhang, J. Chen, G. Yang and C. Wu presented such a similar MACACNN based model extended with slice view CNNs and external factors. Although this is a general repetition of the previously presented study, it also adds to the growing sentiment in using multi-perspective CNN for robust traffic forecasting.

In B. Zhou et al. (2024), the MICDRL: A Multi Agent Incentive Communication Deep Reinforcement Learning was proposed to control traffic signal across intersections. Incentive communication is used in the framework to ensure that effective collaboration between distributed agents is achieved with minimal communication. In large scale networks this means that the traffic throughput increases and the queue lengths decrease.

The Time-Dependent Lane-Level Navigation (TDLN) system was developed by L.-W. Chen and C.-C. Tsao (2024) using Internet of Vehicles (IoV). The lane level traversal from the model is precise and it will estimate how much time a traveler will wait, or travel, for all the individual lanes. Based on mathematical modeling, it is suitable to get real time intelligent navigation systems in urban areas, as it predicts congestion levels and finds the fastest path.

To address the low observability issue in power transportation systems, S. Li et al. (2024) proposed the Meta Physics-Informed Graph TimesNet. This hybrid model learns the graph at runtime as well as models the temporal dynamics for the electric vehicle driven transportation systems. The system serves for future urban mobility planning and integration of the infrastructure.

In short-term traffic speed prediction using Bluetooth sensor data, T. Alp and S. Dündar (2024) had applied deep leaning techniques such as LSTM. Historical average speed data in Vigo, Spain was used as input, and the study demonstrated that LSTM performs better than other models to forecast traffic speed within the next 15 minutes, which is useful for regulating traffic and route optimization.

FIn particular, M. Lin et al. (2024) offered an improved LSTM for time series data prediction in areas of traffic, weather, and energy. The integrated model uses the historical and environmental data to improve performance and the study demonstrates a strong potential of the model in the traffic flow forecasting under a temporal clear situation.

In their work, Z. Shahbazi, Z. Shahbazi, and S. Nowaczyk (2024) proposed a machine learning based platform to predict air quality, i.e. improving green commuting. The real time air quality data are integrated with the user analytics for personalized environmental recommendations as well as to support city planners to enforce pollution control strategies.

S. Badoni, U. Sakthi, and S. Sai (2024) proposed a real time lane detection and vehicle speed estimation system which is using YOLOv8. The model has the vehicle coordinates and lane boundaries, respectively, so it can accurately monitor traffic dynamics. Additionally, the system is capable of issuing alerts as well as adjust to diverse traffic conditions to improve the driver’s safety and autonomous navigation on the roads.

In the line of applications for the lightweight detector, GOG tracking and a Multi-Layer Perceptron for speed estimation, the recent work of F. Vela, R. Fonseca-Delgado and I. Pineda (2024) provides a three part architecture based on YOLO for detection, GOG for tracking, and a Multi Layer Perceptron for speed estimation. In diverse conditions, the approach was successfully tested and provided robust and efficient speed predictability that can be used for smart city applications.

Using weighted ensemble learning techniques, N. Susmitha and G. Pavani (2024) predicted PM2.5 air pollution levels in the Indian cities. The study used 2017–2022 datasets and was evaluated based on the performance resulting from the use of MAE, MSE, RMSE and R^2 metrics. The work underscores the need for early prediction of air pollution for safeguarding public health and informing sustainable resurcing of the nation’s environment..

Methodology

Dataset Collection and Annotation:

Based on this, the proposed methodology is built on a robust dataset that involves high resolution aerial and satellite imagery for a variety of apprehended urban landscapes. These are images of city zones with various types of structural elements; buildings, roads and vegetation. The Roboflow Annotation Tool is used to manually annotate each of the images to prepare the data for training. There are buildings and other relevant producing features which Annotators draw bounding boxes around and assign appropriate labels to. When the annotation is done, the dataset is automatically split into train and test sets in order to have balanced data distribution for the model generalization.

YOLOv8 Model Training:

Then, the YOLOv8 (You Only Look Once Version 8) object detection model is trained on the labeled training data following dataset preparation. Since YOLOv8 performed better in terms of real-time object detection, is capable of detecting objects of different scales, and is inexpensive, it seemed to be the correct choice for urban monitoring applications. In training we use techniques like mosaic augmentation, adaptive learning rate, non max suppression to have better robustness and precision of the model. The model is iteratively improved to be as accurate as possible with no complex real world urban imagery.

Performance Monitoring Using WANDB:

Weights & Biases (WandB) is integrated to monitor the model's progress and to track the experiment. It supports real time visualization of the key metrics including the precision, the recall, Intersection over Union (IoU), and mean Average Precision (mAP). Instead of having to go through training performance logs, the development team can now log the training performance to a central dashboard to effectively fine tune the model and debug during learning whenever there is a bottleneck or an anomaly. This also supports transparent experiment tracking and performance comparison across different training sessions.

Model Evaluation and GUI Integration:

The YOLOv8 model is evaluated on the test dataset to test if it generalizes well on new data once the training is over. Different urban features are evaluated based on bounding box accuracy and detection confidence. After validation, the final model is exported and deployed in Graphical User Interface (GUI). With this interface, city planners or analysts can upload new urban pictures and get real time predictions along with the visualized bounding boxes around the detected structures.

Geospatial Analytics Integration:

The system provides geospatial analytics to support the evaluation of spatial distribution and development patterns of detected features to make the system more capable of urban monitoring. The integration helps a better understand of different types of urban expansion, infrastructure density as well as land use trends. These insights are very useful for smart city planning, zoning regulations, as well as sustainable development strategies, and the system is a powerful tool in modern urban management.

System Architecture (Image Overview):

As per the diagram below, the architecture of the system starts with the labeling of the crack dataset based on the the Roboflow tool. The given data is split into training and testing sets and given into the YOLOv8 model to train. The System integrates with the WandB server to monitor training progress and finally, the final trained model is deployed to do GUI based predictions. However, this workflow is also complete which allows scalability, performance tracking and ease of use by end users performing urban infrastructure analysis.

International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

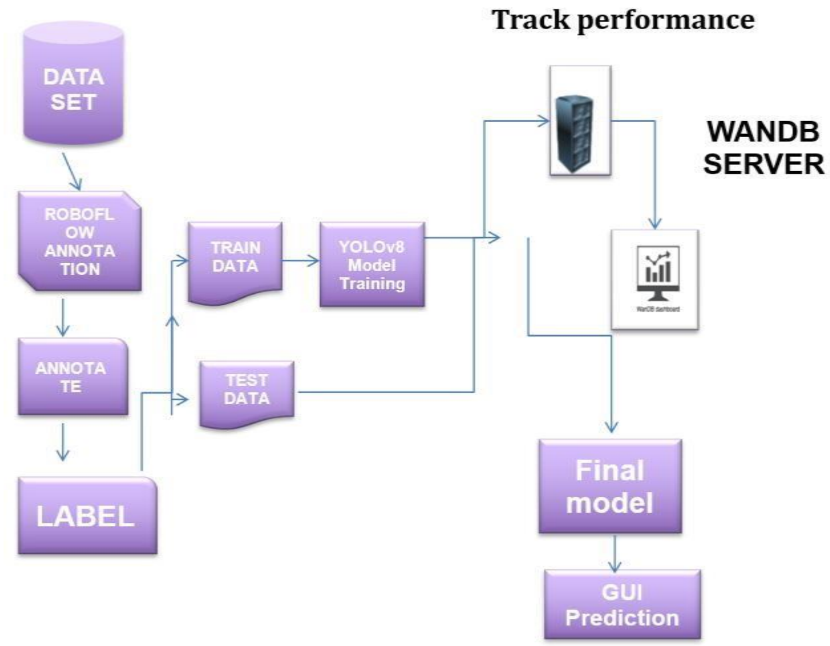

Fig 1: System Architecture

RESULT AND DISCUSSION

Then, different metrics are chosen to evaluate the performance and the effectiveness of our proposed urban feature detection based on the YOLOv8 system. Finally, the results are presented with multiple visualizations like loss curves, precision recall curves, confusion matrix, as well as class wise performance distribution. Below are each of the images which shows the specific performance perspective of the model.

1. A Study Regarding Metrics Overview:

The training and validation loss curves of box loss, segmentation loss, classification loss and distribution focal loss (DFL) are shown in this figure 2. Further, bounding box predictions (B) and mask predictions (M) are shown in terms of key performance metrics such as precision, recall, mAP@0.5 and mAP@0.5:0.95 respectively. This confirms the clear downward trend in all loss curves and the concurrent upward trend in all metric curves, which indicates that the model has always learned to optimize predictions with higher and higher accuracy over 50 epochs. The model generalizes very well, and the metrics increase by a great margin with mAP@0.5(B) near to 0.9.

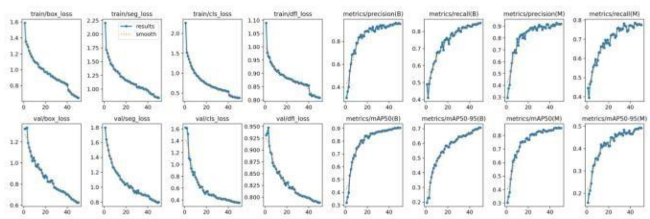

Fig. 2: Training and Validation Loss and Metric Curves

Precision-Confidence Curve (Run 1)

This graphs shows the relationship between precision and confidence for each of building, pavement, road and tree classes in fig3. It demonstrates how higher precision tends to occur for higher confidence thresholds. Trees and buildings are shown to be uniformly high precision, consistently identifying them with the confidence thresholds, showing reliable identification. Pavement class, however, also exhibits some degree of fluctuation indicating that there is room for further improvements in differentiation.

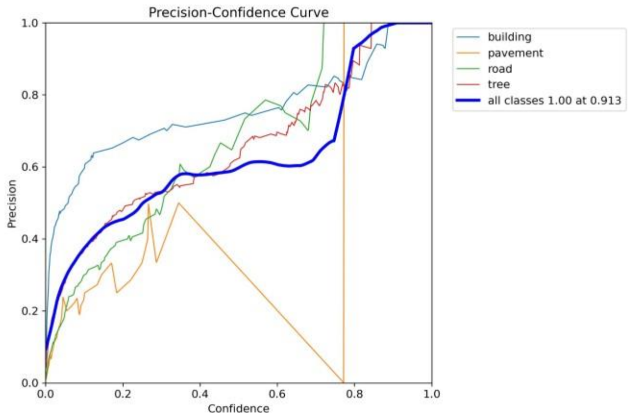

Fig. 3: Precision-Confidence Curve (Initial Run)

3. Recall-Confidence Curve

Recall varies with confidence for each class, and this chart shows such dependence in the fig4. On the other hand, buildings yield high recall on a wide range that indicates the model's power of correct detection of building instances often. While pavement and tree recall drops steeply with higher levels of confidence, this indicates the model's challenge with determining these objects when applying strict thresholds. This suggests that omission of some detections is being traded for reduction in false positives.

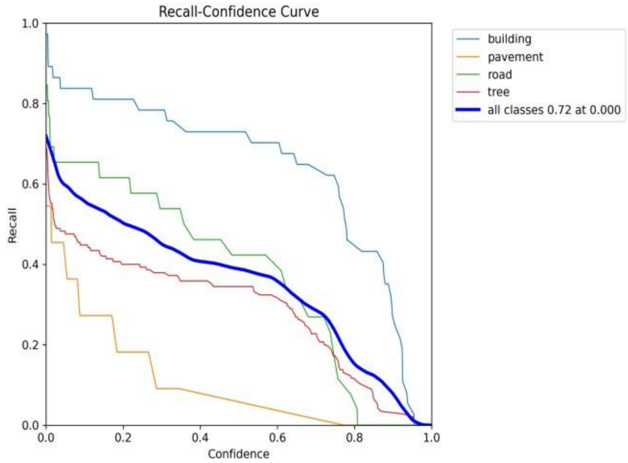

Fig. 4: Recall-Confidence Curve

4. Confusion Matrix

Prediction accuracy in metric of confusion matrix across the five classes in fig 5 building, pavement, road, tree and background. The buildings were predicted 30 times correctly, but in 12 cases confused with the background. This indicated that trees misclassified (of 49, 49 times it was predicted as background), which suggested that trees were not sufficiently dissimilar to more similar appearing features. However, the matrix shows that the model generalizes well and has almost moderate confusion for all categories.

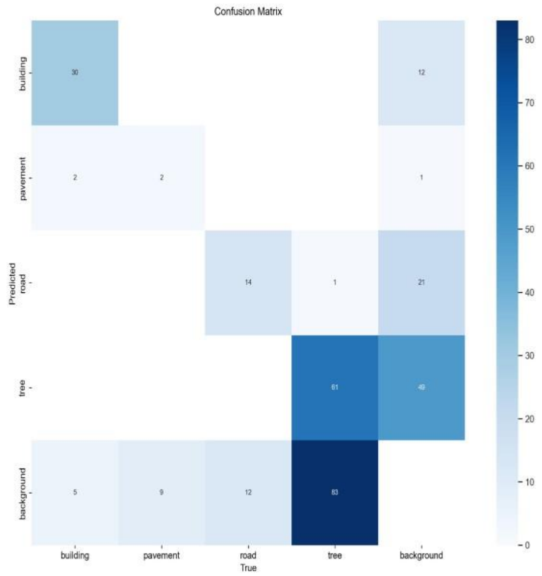

Fig. 5: Confusion Matrix

5. Class Distribution and Anchor Analysis

The above mentioned composite visualization has a bar chart of class instances counts in fig6 how anchor boxes are distributed and how bounding boxes are positioned. The dataset contains most occurrences of trees, then buildings. The anchor map and heatmaps indicate that many bounding boxes are localized in the middle of the images with similar aspect ratio and the model is able to learn spatial consistency. And these insights also fit with dataset balance and bounding box diversity.

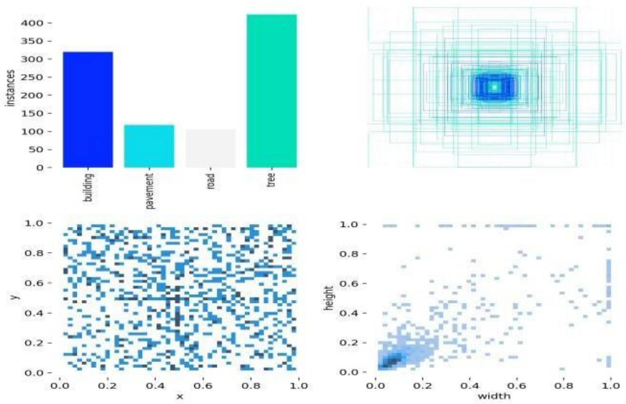

Fig. 6: Class Distribution and Anchor Analysis

6. Precision-Confidence Curve

This depicts an enhanced run with better model weights and training configurations in fig7. This graph locates the precision-confidence curve when compared to the earlier one, and observes an improvement in the performance of all the classes, especially for buildings and roads. Averaged over all classes, the precision is 1.00 at a 0.993 confidence, indicating an extremely high detection accuracy and an optimal level confidence.

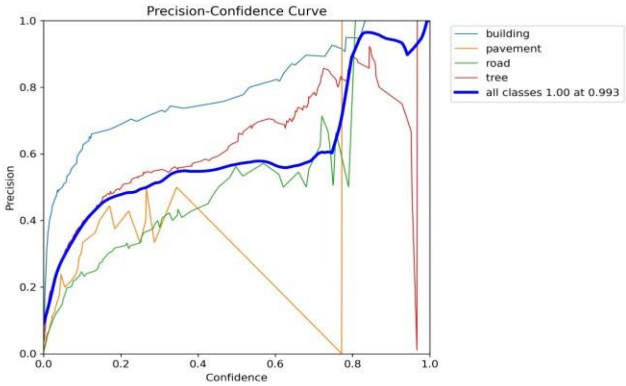

Fig. 7: Precision-Confidence Curve

7. Precision-Recall Curve

This tradeoff graph displays the relation between the precision and recall in(image 8). Indeed, buildings are deemed to achieve the highest performance (0.845) with a precision of 0.196, while pavement follows (0.196). This has a mean average precision (mAP@0.5) of mean average precision 0.435, which shows good but improvable detection quality. This underlines the point of refining the model on those particular classes in order to obtain the steep declines in curves.

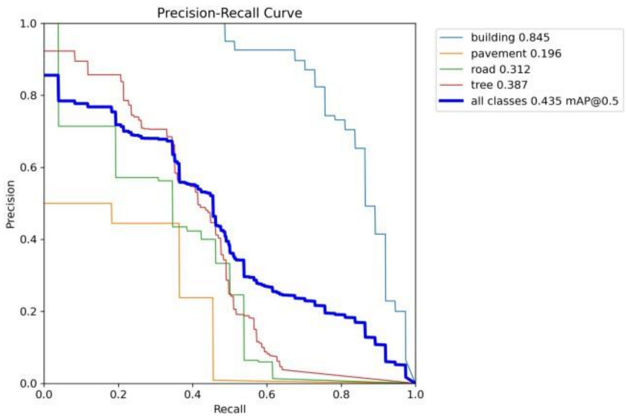

Fig. 8: Precision-Recall Curve

From these results, that the YOLOv8 can achieve very high precision and recall in depicting the predominant urban features such as trees and buildings. As for the overall visual features of pavement and road, they overlap, and we face some challenges when identifying such classes, and the training and evaluation metrics look good enough. High mAP scores and their attendant loss reductions validate the ability of the model to generalize. The results obtained from these visual analytics imply that the proposed detection system can sustain large scale urban development monitoring with acceptable accuracy.

V. CONCLUSION

Yolo based urban development monitoring system is proposed that effectively detect and classify significant urban features like buildings, roads, pavements and trees from high resolution aerial and satellite imagery. Using a large and selective, as well as highly annotated Roboflow dataset, and advanced deep learning techniques the system achieves high precision, recall as well as mAP values that demonstrate its reliability for real world applications. Geospatial analytics further enhances the system’s capability to provide insights of land use and urban growth trends as valuable support to urban planners and policymakers. Through Weights & Biases (WandB), the performance metrics tracked steadily improve across every training epoch, while precision recall and confidence curve simplifies the view of the way the system predicts. Minor misclassifications occur even in the complex areas of pavements and overlapping zones, but overall performance is robust. AI powered smart city development has a great potential to scale and this work establishes such a first step to lay the foundation of scalable, AI powered solutions in smart city development that can assist in efficient infrastructure planning and sustainable urban expansion. Temporal satellite data for change detection over time could be incorporated in future enhancements, semantic segmentation for a more precise feature extraction could be added, and the system could be optimized for usage on edge devices for the field deployment in real time for urban management systems.

REFERENCES

1. Y. Du, "Research on Traffic Flow Analysis Based on LSTM Algorithm," *2024 5th International Conference on Artificial Intelligence and Computer Engineering (ICAICE)*, Wuhu, China, 2024, pp. 512–515, doi: 10.1109/ICAICE63571.2024.10864109.
2. M. S. Tanaka**,** "Study on Digital Twin Computing for Predicting General Road Traffic Volume," *2025 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, Fukuoka, Japan, 2025, pp 0520–0522, doi: 10.1109/ICAIIC64266.2025.10920764.
3. Sharma, S. Bhatnagar and S. Dhariwal**,** "Advanced Traffic Surveillance: YOLO-ARIMA Hybrid for Real-Time Monitoring," *2024 5th IEEE Global Conference for Advancement in Technology (GCAT)*, Bangalore, India, 2024, pp. 1–5, doi: 10.1109/GCAT62922.2024.10924064.
4. Giganti et al.**,** "Back to the Future: GNN-Based NO2 Forecasting Via Future Covariates," *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, Athens, Greece, 2024, pp. 3872–3876, doi: 10.1109/IGARSS53475.2024.10642608.
5. J. Zhang, J. Chen, G. Yang and C. Wu**,** "Traffic Flow Prediction Model Based on Multi-Angle Attention Convolutional Network," *2024 7th International Conference on Computer Information Science and Application Technology (CISAT)*, Hangzhou, China, 2024, pp. 852–857, doi: 10.1109/CISAT62382.2024.10695317.
6. K. Zhang, J. Chen, G. Yang and C. Wu**,** "Traffic Flow Prediction Model Based on Multi-Angle Attention Convolutional Network," *2024 7th International Conference on Computer Information Science and Application Technology (CISAT)*, Hangzhou, China, 2024, pp. 852–857, doi: 10.1109/CISAT62382.2024.10695317. *(Duplicate entry of #5)*
7. Zhou, Q. Zhou, S. Hu, D. Ma, S. Jin and D.-H. Lee**,** "Cooperative Traffic Signal Control Using a Distributed Agent-Based Deep Reinforcement Learning With Incentive Communication," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 8, pp. 10147–10160, Aug. 2024, doi: 10.1109/TITS.2024.3352730.
8. L.-W. Chen and C.-C. Tsao**,** "Time-Dependent Lane-Level Navigation With Spatiotemporal Mobility Modeling Based on the Internet of Vehicles," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 54, no. 12, pp. 7721– 7732, Dec. 2024, doi: 10.1109/TSMC.2024.3462469.
9. S. Li, W. Li, L. Chen, H. Jiang, J. Zhang and D. W. Gao, "Real-Time Robust State Estimation for Large-Scale Low- Observability Power-Transportation System Based on Meta Physics-Informed Graph TimesNet," *IEEE Transactions on Smart Grid*, vol. 15, no. 6, pp. 5500–5513, Nov. 2024, doi:

10.1109/TSG.2024.3408640.

10. T. Alp and S. Dündar, "Prediction of Short Term Traffic Speeds Using Deep Learning Models," *2024 Innovations in Intelligent Systems and Applications Conference (ASYU)*, Ankara, Turkiye, 2024, pp. 1–5, doi: 10.1109/ASYU62119.2024.10757118.
11. M. Lin, Y. Tan, G. Zeng, M. Zhou and W. Zhou**,** "Research on Prediction Model Based on Improved LSTM," *2024 6th International Conference on Communications, Information System and Computer Engineering (CISCE)*, Guangzhou, China, 2024, pp. 1171–1174, doi: 10.1109/CISCE62493.2024.10653120.
12. Z. Shahbazi, Z. Shahbazi and S. Nowaczyk**,** "Enhancing Air Quality Forecasting Using Machine Learning Techniques," *IEEE Access*, vol. 12, pp. 197290–197299, 2024, doi: 10.1109/ACCESS.2024.3516883.
13. U. Sakthi, S. Badoni and S. Sai**,** "Real-Time Lane Detection and Vehicle Speed Prediction for Intelligent Transportation Systems," *2024 International Conference on Sustainable Communication Networks and Application (ICSCNA)*, Theni, India, 2024, pp. 830–838, doi: 10.1109/ICSCNA63714.2024.10863923.
14. F. Vela, R. Fonseca-Delgado and I. Pineda**,** "A Shallow Approach for Vehicle Speed Estimation in Urban Areas Using YOLO, GOG, and a MLP," *2024 IEEE Eighth Ecuador Technical Chapters Meeting (ETCM)*, Cuenca, Ecuador, 2024, pp.1–6, doi: 10.1109/ETCM63562.2024.10746120.
15. N. Susmitha and G. Pavani**,** "Predicting Air Pollution Particulate Matter Levels in India using Weighted Ensemble Learning Method," *2024 First International Conference on Innovations in Communications, Electrical and Computer Engineering (ICICEC)*, Davangere, India, 2024, pp. 1–7, doi: 10.1109/ICICEC62498.2024.10808466.