ISSN: 3079-6342

Assessment of Perception Accuracy in Autonomous Driving under Complex Urban Conditions with Sensor Integration

James Anderson[1], Ru Chen[1], Emily Zhao[2], David K. Lee[1,2*], Yanqi Sun[3]

[1]Department of Computer Engineering, California State University, Fullerton, CA, USA

[2]School of Automation and Information Engineering, Sichuan University of Science & Engineering, Zigong, China

[3]Department of Information Technology, University of North Texas, Denton, TX, USA

Corresponding Author

[*] Corresponding Author: David K. Lee (ddklee@fullerton.edu)

Abstract

Dynamic and complex traffic scenarios in urban environments impose stringent requirements on the perception capability of autonomous driving systems. In this study, we develop a perception model that integrates data from LiDAR, cameras and millimeter-wave radar through multimodal sensor fusion and employs a large-scale Transformer-based architecture. By adopting the Bird's Eye View (BEV) representation and a multi-scale feature enhancement mechanism, the proposed model significantly improves the accuracy of 3D object detection and semantic interpretation. At the architectural level, we introduce a cross-modal attention mechanism and a sparse attention module, which enhance the model's perception performance in challenging situations such as occlusion, drastic lighting changes, and densely clustered targets. Experiments on the nuScenes and KITTI datasets show that the proposed model outperforms existing approaches such as BEVFormer and VoxelNet in terms of mean Average Precision (mAP), Intersection over Union (IoU), and stability. The model consistently achieves high recognition accuracy and robust adaptability across various urban driving scenarios.

Keywords

Multimodal fusion; BEV representation; Transformer-based model; 3D object detection; autonomous driving perception.

1. Introduction

Urbanization is progressing at an unprecedented rate across the globe. According to data from the United Nations, by 2023, more than 55% of the world's population resided in urban areas and this figure is expected to rise to nearly 70% by 2050 [1]. As large populations continue to concentrate in cities, urban traffic volumes have increased rapidly, resulting in worsening congestion problems [2]. For example, data released by the Beijing Transportation Development Research Institute in 2024 show that during weekday evening peak hours, the average congestion index in central Beijing reached 2.0. On certain road segments, the average vehicle speed dropped to just 15 kilometers per hour and the annual economic loss caused by traffic congestion was estimated at approximately 30 billion [3]. A similar situation has been observed in Shanghai. According to transportation data from 2024, during weekday morning and evening peak periods, 25% of roads in the city center experienced congestion, leading to annual economic losses exceeding 25 billion RMB. At the same time, traffic safety remains a pressing concern. Frequent urban traffic accidents pose serious threats to public safety and

ISSN: 3079-6342

property [4]. Based on the 2024 Traffic Accident Statistical Yearbook issued by the National Bureau of Statistics, deaths from urban road traffic accidents accounted for 62.3% of all traffic fatalities nationwide and the overall accident rate continues to increase annually. In this context, autonomous driving technology is regarded as a promising solution that could significantly improve current traffic conditions [5].

Over the past decade, autonomous driving technology has advanced considerably [6]. It has evolved from basic driver-assistance functions, such as adaptive cruise control and automatic emergency braking, to more advanced stages approaching high and full automation [7]. Some companies have already implemented relatively mature autonomous systems in constrained environments such as mining sites and ports [8]. For example, an autonomous transport fleet deployed by a mining company has safely delivered over five million tons of ore since its launch in 2022 [9]. Compared to traditional manual operations, transport efficiency has increased by 30%. However, in dynamic and unstructured urban environments, autonomous driving systems still face numerous challenges, and the road to full-scale commercial deployment remains long [10]. The complexity of urban environments can be observed across multiple dimensions. First, urban roads are characterized by a high density of traffic participants, including pedestrians, bicycles and motor vehicles, all exhibiting diverse and unpredictable behaviors. According to traffic monitoring data, pedestrian volumes during peak hours in major commercial areas can reach several thousand people per hour [11]. Mixed traffic involving bicycles and vehicles is also common [12]. Sudden pedestrian crossings, random lane changes by cyclists and frequent merging by vehicles introduce significant uncertainty, making perception tasks more difficult for autonomous systems [13,14]. Second, urban road networks contain numerous complex structures such as intersections, roundabouts and sharp curves [15]. These segments often follow different traffic rules and require context-specific driving behaviors, thereby demanding strong scene understanding and decision-making capabilities. Third, lighting conditions in cities change rapidly. Low visibility caused by sunrise and sunset glare, building shadows, or poor night lighting can significantly impair sensor performance [16]. Studies have indicated that under low-light conditions, image contrast from cameras can drop by 30%–40%, and object detection accuracy can fall by approximately 25% [17,18]. Additionally, harsh weather conditions—such as heavy rain, fog and dust—occur frequently in urban areas, further undermining sensor reliability [19]. For instance, in foggy weather, the effective detection range of LiDAR may be reduced to 50%–60% of its normal performance and the accuracy of millimeter-wave radar is also considerably affected [20].

Most conventional autonomous driving perception systems rely on a single type of sensor. Taking cameras as an example, although they can capture detailed visual information, their performance deteriorates significantly under extreme conditions such as low illumination, strong backlight or severe weather [21]. Due to limitations in imaging principles, image quality declines sharply in these scenarios, resulting in a notable decrease in object detection and recognition accuracy. For instance, a mainstream camera model achieves only a 30% detection accuracy for targets located 100 meters away at night without auxiliary lighting, compared to 85% under sufficient daylight conditions [22]. LiDAR provides accurate 3D spatial information, but it shows limited capability in detecting small or distant targets [23]. Moreover, its high cost and performance instability in complex environments remain challenges. For example, a high-resolution LiDAR unit from a well-known brand is priced in the hundreds of thousands of RMB and under rainy conditions, it can miss up to 15% of targets smaller than 0.5 meters in diameter at a distance of 50 meters [24]. Millimeter-wave radar also has clear shortcomings in capturing the precise position and shape of objects, making it difficult to meet the stringent precision requirements of high-level autonomous driving [25]. Its typical ranging error ranges from 0.2 to 0.5 meters. To address the limitations of individual sensors, multimodal sensor fusion has become a growing area of interest. This approach integrates data from LiDAR, cameras and

ISSN: 3079-6342

millimeter-wave radar, taking full advantage of the strengths of each sensor [26]. Specifically, cameras provide rich texture and semantic content, which supports accurate object classification and recognition; LiDAR delivers high-precision distance and spatial positioning, enabling the construction of accurate 3D maps; and millimeter-wave radar is effective at measuring velocity and motion and can still function reliably in adverse weather conditions [27]. The integrated use of these heterogeneous sensing modalities substantially enhances the system's perception capability in complex environments. In parallel, large-scale Transformer architectures have achieved major breakthroughs in both natural language processing and computer vision [28]. Due to their self-attention mechanism, Transformer models can efficiently capture long-range dependencies in sequence data and exhibit superior performance in complex scene processing tasks. In natural language processing, Transformer-based models such as the GPT series have achieved remarkable results in text generation and question answering. In the field of computer vision, Transformer models have shown outstanding performance in image classification, object detection and semantic segmentation [29]. Applying Transformer models to multimodal sensor fusion and perception tasks in autonomous driving holds great potential [30]. Their powerful feature extraction and relational reasoning capabilities are expected to further improve the accuracy and robustness of perception in complex urban environments.

This study proposes an innovative perception model that combines multimodal sensor fusion with a large-scale Transformer architecture. The goal is to address the core challenges of perception in complex urban environments and enhance the overall performance of autonomous driving systems. By deeply integrating data from multiple sensors and leveraging the strengths of the Transformer framework, the proposed model aims to achieve accurate detection and recognition of various objects in traffic scenes, thereby providing reliable and precise input for decision-making modules in autonomous vehicles. This work lays a technical foundation for the safe and efficient deployment of autonomous driving in urban areas.

2. Methods

2.1. Multimodal Sensor Data Acquisition and Fusion Basis

This study employs the AT128 LiDAR sensor developed by Hesai Technology to obtain high-resolution 3D point cloud data. The device provides a vertical resolution of 0.1°, a horizontal resolution of 0.2°, and a maximum detection range of 200 meters, enabling accurate and efficient acquisition of spatial information about the surrounding environment. In parallel, the Sony IMX586 camera is used to capture texture and color information, offering a resolution of 4000 × 3000 pixels and a frame rate of 60 frames per second, which ensures detailed visual data collection. Additionally, the Bosch ARS408 millimeter-wave radar is used to measure object velocity and distance variation, with a velocity accuracy of ±0.1 m/s and a distance accuracy of ±0.2 m. Before experiments, all sensors were carefully time-synchronized and spatially calibrated to ensure consistency in both temporal and spatial dimensions. The time synchronization error was controlled within ±10 microseconds, and the spatial calibration error was kept within ±5 centimeters.

2.2. Multi-Scale Feature Fusion Based on BEV Representation

In this work, we adopt the Bird's Eye View (BEV) representation framework, projecting data from different sensors into a unified BEV coordinate system. Specifically, LiDAR point clouds are projected onto the BEV plane to generate feature maps such as density and height maps [31]. Camera images are processed using depth estimation algorithms to calculate pixel-wise depth values, which are then transformed into the BEV coordinate system and fused with the LiDAR-based features [32]. For the millimeter-wave radar, velocity and distance data are used

ISSN: 3079-6342

to construct a velocity map within the BEV domain. On top of this projection framework, a multi-scale feature extraction network is constructed. Convolution and pooling layers are used to extract features at various scales from the raw data. These features are then fused for object detection tasks. During the extraction process, the resolution of low-scale feature maps is set to one-quarter of the original input, while high-scale feature maps are set to one-sixteenth. Feature maps of different scales are combined using a weighted fusion strategy, where the weights are optimized through experiments to achieve the best fusion performance.

2.3. Cross-Modal Attention Mechanism

To improve the effectiveness of multimodal data fusion, this study introduces a cross-modal attention mechanism based on the self-attention structure of the Transformer. Specifically, feature representations from LiDAR, camera and millimeter-wave radar are input into the cross-modal attention module. By computing attention scores across different modalities, the model determines the relative contribution of each modality during the fusion process [33]. In implementation, features from each modality are first projected into query (Q), key (K), and value (V) spaces. The attention score is calculated by performing dot products between queries and keys, followed by normalization. These scores are then used to weight the corresponding values. The resulting weighted sum yields the fused feature representation. In the attention computation, a scaling factor of 8 is applied to balance computational efficiency and attention expressiveness, enabling effective and efficient multimodal fusion.

2.4. Transformer-Based Large-Scale Model Architecture

A Transformer model composed of encoder and decoder components is constructed to process the fused multimodal features. The encoder includes multiple Transformer blocks, each consisting of a multi-head self-attention layer and a feedforward neural network layer, which together enhance the model's ability to capture complex dependencies among features. The decoder receives as input both the encoder's output and the features refined by the cross-modal attention mechanism. The final output is processed by fully connected layers to perform classification and regression tasks, resulting in object detection outputs. The model is trained in an end-to-end manner, where parameters are optimized by minimizing a combination of cross-entropy loss and mean squared error loss. To prevent overfitting, dropout regularization is applied and a cosine annealing learning rate schedule is adopted to improve training stability [34]. During training, the initial learning rate is set to 0.001, the dropout rate is set to 0.2, and the number of training epochs is set to 100, ensuring that the model fully learns the features and patterns within the data.

3. Results and Discussion

3.1. Experimental Setup

To thoroughly verify the performance of the proposed model, experiments were conducted on the nuScenes and KITTI datasets. The nuScenes dataset is a large-scale benchmark for autonomous driving, encompassing a diverse range of urban traffic scenarios, including variations in weather, lighting conditions, and traffic density. It provides LiDAR point clouds, camera images, and precise annotation information such as object categories, positions, and dimensions. In total, the dataset includes 1,000 scenes, each lasting 20 seconds, covering 10 different object types. The KITTI dataset is a classical and widely adopted dataset in the field of autonomous driving, primarily focused on object detection and scene understanding in urban road environments. It contains extensive real-world driving data, including 7,481 images in the training set and 7,518 images in the test set. During the experiments, each dataset was divided into training, validation, and test sets in a ratio of 7:2:1. The Adam optimizer was used for model training, with an initial learning rate of 0.001, dynamically adjusted using a cosine annealing

ISSN: 3079-6342

schedule. The training process was carried out for 100 epochs. In each epoch, the training set was shuffled to improve the generalization ability of the model. For performance evaluation, mean Average Precision (mAP) and Intersection over Union (IoU) were adopted as the primary metrics [35,36]. The mAP metric reflects the average precision under different recall thresholds, while IoU measures the overlap ratio between predicted and ground-truth bounding boxes. These metrics provide a comprehensive evaluation of the model's detection capability.

3.2. Model Performance Evaluation

On the nuScenes dataset, the proposed model achieved an mAP of 0.78, which represents a significant improvement compared to mainstream methods such as BEVFormer (0.72) and VoxelNet (0.68). The detailed comparison results are summarized in Table 2. In terms of the IoU metric for vehicle detection, the proposed model achieved an average IoU of 0.75, again outperforming BEVFormer (0.70) and VoxelNet (0.65). On the KITTI dataset, the proposed model also demonstrated strong performance in 3D object detection tasks, achieving an mAP of 0.82. This result surpasses the performance of BEVFormer (0.78) and VoxelNet (0.75), indicating the effectiveness of the model in diverse urban driving environments.

Table 1. mAP Comparison of Different Models on the nuScenes and KITTI Datasets

Model	mAP on nuScenes	mAP on KITTI
Proposed Model	0.78	0.82
BEVFormer	0.72	0.78
VoxelNet	0.68	0.75

A thorough analysis of the experimental results indicates that the proposed model demonstrates clear advantages in handling complex scenarios. For example, in occlusion settings, the cross-modal attention mechanism effectively captures the complementary features among different sensor modalities, allowing the model to accurately detect partially obscured objects [37]. In scenes with significant illumination changes, the model maintains a high detection accuracy by leveraging the multi-scale feature enhancement mechanism and image enhancement preprocessing [38]. Furthermore, in high-density traffic scenarios, the strong relational modeling capability of the Transformer structure enables the model to clearly distinguish between multiple objects, thereby reducing target confusion and significantly improving detection accuracy and stability.

3.3. Comparative Analysis with Other Methods

Compared with BEVFormer, the proposed model achieves an innovative improvement in the method of multimodal sensor data fusion. BEVFormer mainly focuses on BEV feature integration based on attention mechanisms. In contrast, this study not only adopts the BEV representation paradigm but also introduces a cross-modal attention mechanism, which allows for a deeper exploration of the internal correlations between different sensor modalities. This design enables the proposed model to achieve better object detection performance under complex conditions. Compared with VoxelNet, which primarily performs object detection based on voxelized point cloud data and relies heavily on the spatial structure of the input, the proposed model integrates multimodal data and a Transformer-based architecture. This combination allows the model to extract features from more diverse sources of information while reducing its dependence on rigid spatial structures. As a result, the proposed model demonstrates stronger adaptability in various application scenarios. However, the proposed model also presents certain aspects that require improvement. For instance, the computational complexity of the model is relatively high, which may impose more demanding hardware

ISSN: 3079-6342

requirements in real-world deployment. According to our measurements, when processing a single frame containing 10,000 LiDAR points, an image with a resolution of 4000×3000 pixels, and millimeter-wave radar data, the inference time of the proposed model is approximately 0.15 seconds. In comparison, the inference time for BEVFormer is 0.10 seconds, and for VoxelNet, it is 0.08 seconds. Detailed results are shown in Table 3. This higher inference time is primarily attributed to the substantial computational overhead of the Transformer architecture, especially when processing large-scale data inputs. Future research may explore model compression and acceleration techniques—such as pruning and quantization—to reduce the model's computational complexity and improve its inference speed. These optimizations would facilitate the practical deployment of the model in real-world autonomous driving scenarios.

Table 2. Inference Time Comparison of Different Models

Model	Inference Time (per frame)
Proposed Model	0.15 s
BEVFormer	0.10 s
VoxelNet	0.08 s

4. Conclusion

This work proposes a multimodal perception framework that combines LiDAR, camera, and millimeter-wave radar with a Transformer-based architecture to improve object detection in urban autonomous driving environments. The integration of cross-modal attention mechanisms and Bird's Eye View (BEV) representation enables the model to effectively fuse heterogeneous sensor data and capture complex spatial dependencies. Experimental validation on nuScenes and KITTI datasets indicates consistent improvements over established baselines, including BEVFormer and VoxelNet, in both mAP and IoU metrics. The model demonstrates robustness in challenging scenarios such as occlusion, low illumination, and high-density traffic, suggesting its suitability for complex urban contexts. Nonetheless, the relatively high inference time—mainly attributed to the computational load of the Transformer layers—remains a constraint for deployment in latency-sensitive applications. Future research should consider model compression techniques to mitigate computational costs while maintaining detection performance. Overall, the proposed framework contributes a technically sound and empirically validated approach to advancing perception capabilities in autonomous driving systems.

References

[1] Zhao, R., Hao, Y., & Li, X. (2024). Business Analysis: User Attitude Evaluation and Prediction Based on Hotel User Reviews and Text Mining. arXiv preprint arXiv:2412.16744.

[2] Lv, G., Li, X., Jensen, E., Soman, B., Tsao, Y. H., Evans, C. M., & Cahill, D. G. (2023). Dynamic covalent bonds in vitrimers enable 1.0 W/(m K) intrinsic thermal conductivity. Macromolecules, 56(4), 1554-1561.

[3] Xiao, Y., Tan, L., & Liu, J. (2025). Application of Machine Learning Model in Fraud Identification: A Comparative Study of CatBoost, XGBoost and LightGBM.

[4] Gong, C., Zhang, X., Lin, Y., Lu, H., Su, P. C., & Zhang, J. (2025). Federated Learning for Heterogeneous Data Integration and Privacy Protection.

[5] Shih, K., Han, Y., & Tan, L. (2025). Recommendation System in Advertising and Streaming Media: Unsupervised Data Enhancement Sequence Suggestions.

ISSN: 3079-6342

[6] Jiang, G., Yang, J., Zhao, S., Chen, H., Zhong, Y., & Gong, C. (2025). Investment Advisory Robotics 2.0: Leveraging Deep Neural Networks for Personalized Financial Guidance.

[7] Lv, G., Li, X., Jensen, E., Soman, B., Tsao, Y. H., Evans, C. M., & Cahill, D. G. (2023). Dynamic covalent bonds in vitrimers enable 1.0 W/(m K) intrinsic thermal conductivity. Macromolecules, 56(4), 1554-1561.

[8] Wang, Y., Shao, W., Lin, J., & Zheng, S. (2025). Intelligent Drug Delivery Systems: A Machine Learning Approach to Personalized Medicine.

[9] Zhang, B., Han, X., & Han, Y. (2025). Research on Multimodal Retrieval System of e-Commerce Platform Based on Pre-Training Model.

[10] Wang, Y., Jia, P., Shu, Z., Liu, K., & Shariff, A. R. M. (2025). Multidimensional precipitation index prediction based on CNN-LSTM hybrid framework. arXiv preprint arXiv:2504.20442.

[11] Ge, G., Zelig, R., Brown, T., & Radler, D. R. (2025). A review of the effect of the ketogenic diet on glycemic control in adults with type 2 diabetes. Precision Nutrition, 4(1), e00100.

[12] Zhang, L., & Liang, R. (2025). Avocado Price Prediction Using a Hybrid Deep Learning Model: TCN-MLP-Attention Architecture. arXiv preprint arXiv:2505.09907.

[13] Yang, M., Wang, Y., Shi, J., & Tong, L. (2025). Reinforcement Learning Based Multi-Stage Ad Sorting and Personalized Recommendation System Design.

[14] Peng, H., Ge, L., Zheng, X., & Wang, Y. (2025). Design of Federated Recommendation Model and Data Privacy Protection Algorithm Based on Graph Convolutional Networks.

[15] Chen, F., Liang, H., Yue, L., Xu, P., & Li, S. (2025). Low-Power Acceleration Architecture Design of Domestic Smart Chips for AI Loads.

[16] Liang, R., Ye, Z., Liang, Y., & Li, S. (2025). Deep Learning-Based Player Behavior Modeling and Game Interaction System Optimization Research.

[17] Zheng, Z., Wu, S., & Ding, W. (2025). CTLformer: A Hybrid Denoising Model Combining Convolutional Layers and Self-Attention for Enhanced CT Image Reconstruction. arXiv preprint arXiv:2505.12203.

[18] Gui, H., Fu, Y., Wang, B., & Lu, Y. (2025). Optimized Design of Medical Welded Structures for Life Enhancement.

[19] Gui, H., Fu, Y., Wang, Z., & Zong, W. (2025). Research on Dynamic Balance Control of Ct Gantry Based on Multi-Body Dynamics Algorithm.

[20] Freedman, H., Young, N., Schaefer, D., Song, Q., van der Hoek, A., & Tomlinson, B. (2024). Construction and Analysis of Collaborative Educational Networks based on Student Concept Maps. Proceedings of the ACM on Human-Computer Interaction, 8(CSCW1), 1-22.

[21] Hu, J., Zeng, H., & Tian, Z. (2025). Applications and Effect Evaluation of Generative Adversarial Networks in Semi-Supervised Learning. arXiv preprint arXiv:2505.19522.

[22] Song, Z., Liu, Z., & Li, H. (2025). Research on feature fusion and multimodal patent text based on graph attention network. arXiv preprint arXiv:2505.20188.

[23] Zhang, G., Hu, X., You, Z., Zhang, J., & Xiao, Y. (2025). Intelligent Decision Optimization System for Enterprise Electronic Product Manufacturing Based on Cloud Computing.

[24] Fan, P., Liu, K., & Qi, Z. (2025). Material Flow Prediction Task Based On TCN-GRU Deep Fusion Model.

[25] Xu, J., Wang, H., & Trimbach, H. (2016, June). An OWL ontology representation for machine-learned functions using linked data. In 2016 IEEE International Congress on Big Data (BigData Congress) (pp. 319-322). IEEE.

[26] Fu, Y., Gui, H., Li, W., & Wang, Z. (2020, August). Virtual Material Modeling and Vibration Reduction Design of Electron Beam Imaging System. In 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA) (pp. 1063-1070). IEEE.

[27] Gui, H., Zong, W., Fu, Y., & Wang, Z. (2025). Residual Unbalance Moment Suppression and Vibration Performance Improvement of Rotating Structures Based on Medical Devices.

[28] Chen, F., Liang, H., Li, S., Yue, L., & Xu, P. (2025). Design of Domestic Chip Scheduling Architecture for Smart Grid Based on Edge Collaboration.

ISSN: 3079-6342

[29] Liang, R., Feifan, F. N. U., Liang, Y., & Ye, Z. (2025). Emotion-Aware Interface Adaptation in Mobile Applications Based on Color Psychology and Multimodal User State Recognition. Frontiers in Artificial Intelligence Research, 2(1), 51-57.

[30] Qin, F., Cheng, H. Y., Sneeringer, R., Vlachostergiou, M., Acharya, S., Liu, H., ... & Yao, L. (2021, May). ExoForm: Shape memory and self-fusing semi-rigid wearables. In Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-8).

[31] Xu, K., Xu, X., Wu, H., & Sun, R. (2024). Venturi Aeration Systems Design and Performance Evaluation in High Density Aquaculture.

[32] Wang, G., Qin, F., Liu, H., Tao, Y., Zhang, Y., Zhang, Y. J., & Yao, L. (2020). MorphingCircuit: An integrated design, simulation, and fabrication workflow for self-morphing electronics. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4(4), 1-26.

[33] Xu, K., Xu, X., Wu, H., Sun, R., & Hong, Y. (2023). Ozonation and Filtration System for Sustainable Treatment of Aquaculture Wastewater in Taizhou City. Innovations in Applied Engineering and Technology, 1-7.

[34] Wang, Y., Wen, Y., Wu, X., Wang, L., & Cai, H. (2025). Assessing the Role of Adaptive Digital Platforms in Personalized Nutrition and Chronic Disease Management.

[35] Yang, M., Wu, J., Tong, L., & Shi, J. (2025). Design of Advertisement Creative Optimization and Performance Enhancement System Based on Multimodal Deep Learning.

[36] Peng, H., Tian, D., Wang, T., & Han, L. (2025). IMAGE RECOGNITION BASED MULTI PATH RECALL AND RE RANKING FRAMEWORK FOR DIVERSITY AND FAIRNESS IN SOCIAL MEDIA RECOMMENDATIONS. Scientific Insights and Perspectives, 2(1), 11-20.

[37] Peng, H., Dong, N., Liao, Y., Tang, Y., & Hu, X. (2024). Real-Time Turbidity Monitoring Using Machine Learning and Environmental Parameter Integration for Scalable Water Quality Management. Journal of Theory and Practice in Engineering and Technology, 1(4), 29-36.

[38] Zheng, J., & Makar, M. (2022). Causally motivated multi-shortcut identification and removal. Advances in Neural Information Processing Systems, 35, 12800-12812.