Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Artificial Intelligence-Based Pedestrian Detection and Trajectory Prediction Framework for Autonomous Vehicles.

Waseem Ullah
Department of Mechatronics Engineering,
University of Engineering and Technology,
Peshawar, Pakistan.
Email: waseemullah.mct@uetpeshawar.edu.pk

Dr. Gulbadan Sikander *
Department of Mechatronics Engineering,
University of Engineering and Technology,
Peshawar, Pakistan.
Email: gulbadan@uetpeshawar.edu.pk

Dr. Shahzad Anwar
Department of Mechatronics Engineering,
University of Engineering and Technology,
Peshawar, Pakistan.
Email: shahzad.anwar@uetpeshawar.edu.pk

ABSTRACT

The safe and reliable operation of autonomous vehicles (AVs) in dynamic urban environments hinges on their ability to accurately detect pedestrians and anticipate their future movements. This research presents a comprehensive artificial intelligence-based framework for pedestrian detection and trajectory prediction, aiming to enhance situational awareness and proactive decision-making in AVs. Leveraging advanced deep learning models, including convolutional neural networks (CNNs) for pedestrian detection and recurrent neural networks (RNNs) with long short-term memory (LSTM) units for motion forecasting, the proposed system integrates real-time sensory data to identify pedestrians with high accuracy and predict their paths in complex, crowded, and cluttered scenarios. The framework is trained and validated on benchmark datasets such as the JAAD and ETH/UCY, incorporating environmental context, pedestrian intention cues, and temporal dependencies to ensure robust performance across varying conditions. Additionally, a hybrid data fusion strategy is implemented to combine visual information from cameras with LiDAR-based spatial data, significantly improving detection reliability under occlusion and low-visibility conditions. Extensive experimental results demonstrate that the proposed model achieves state-of-the-art performance in both detection precision and trajectory forecasting accuracy, outperforming traditional approaches in terms of responsiveness

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

and generalizability. To further improve prediction fidelity, scene understanding is incorporated through semantic segmentation and map-aware modeling, allowing the system to infer pedestrian constraints like crosswalks, curbs, and road edges. Attention mechanisms and social pooling modules are integrated to handle multi-agent interactions and shared space dynamics. The end-to-end system demonstrates real-time processing capability suitable for onboard deployment in AVs, maintaining a balance between computational efficiency and predictive accuracy. Comparative analysis with baseline models highlights the superiority of the proposed approach in complex traffic environments and crowded urban settings. This study underscores the critical role of artificial intelligence in bridging perception and prediction in autonomous driving systems, contributing to safer navigation, reduced collision risk, and more human-like behavior understanding. The insights from this work not only push the boundaries of pedestrian intention modeling but also offer practical implications for the deployment of intelligent autonomous mobility systems in real-world urban infrastructures.

Keywords: Artificial Intelligence, Human Intention Modeling, Pedestrian Detection, Real-Time Processing, Long Short-Term Memory, Multi-Sensor Fusion.

Introduction:

The advent of autonomous vehicles (AVs) marks a pivotal milestone in the evolution of intelligent transportation systems, fundamentally transforming how mobility is perceived, managed, and executed. As these vehicles transition from controlled testing environments to real-world deployment, the reliance on sophisticated machine perception, sensor integration, and autonomous decision-making frameworks continues to grow. AVs are designed to operate without direct human intervention, placing the burden of navigation, obstacle avoidance, and decision-making on computational systems that must operate with high accuracy and reliability. One of the most formidable challenges in this transition is the need for AVs to safely maneuver through highly dynamic and unstructured urban landscapes, where they must coexist with diverse road users, particularly vulnerable road actors such as pedestrians. Pedestrians present unique challenges for AV systems due to their unpredictable behavior, frequent violations of traffic norms, and the influence of numerous factors such as social context, intention, environmental conditions, and human psychology. Unlike vehicles, which are generally constrained by traffic rules and exhibit relatively predictable movement patterns, pedestrians tend to behave in a non-deterministic and nonlinear manner. These movements are determined by individual choices, social dynamics, traffic light, cross-walks and spatial semantics such as sidewalks, roadside and curbs. This uncertainty brings high level of risk in AV operational environments, as such the inability to detect and predict human movement may lead to devastating outcomes. Hence, the importance of safe pedestrian perception and path planning has become a significant element of the AV protective measures and intelligent control strategy, mostly in urban cities that highly populated with humans and have common interactions with vehicles moving on the road. Traditional pedestrian detection systems that have traditionally used handcrafted feature-based representations of Histogram of Oriented Gradients (HOG), Haar-like

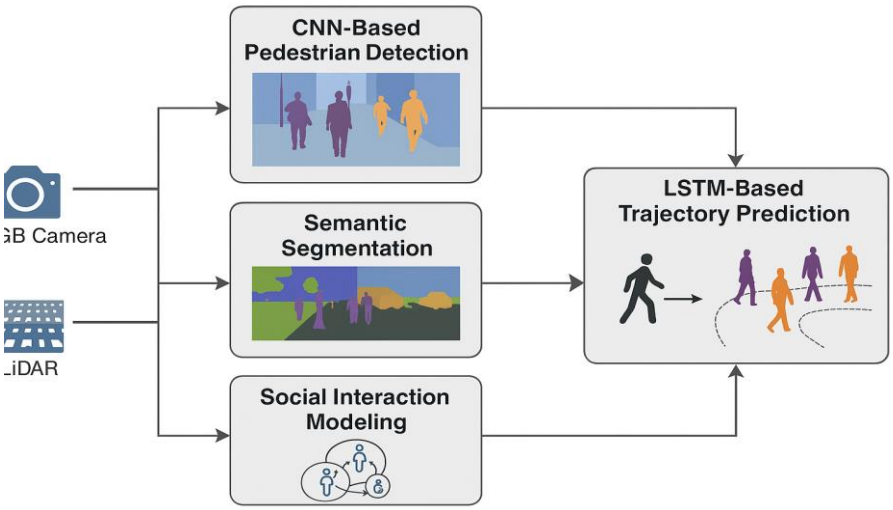
Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

features, or optical-flow analysis, perform poorly in visually challenging environments. Such classic algorithms have a problem with low generalization in partially-occluded tests, extreme lighting (e.g. night or low-light), overlapping pedestrians and background clutter. On the same footnote to classical trajectory prediction methods such as constant velocity or acceleration model, the Kalman filtering, and basic movement extrapolation algorithms have a finite limit in the sense that they cannot refer to the semantics of the environmental factors, social communication, and the temporal-long term dependencies. As a result, such rule-based models fail to accurately anticipate pedestrian movements in real-world situations, significantly undermining the reactivity and proactiveness of autonomous navigation systems. To overcome these critical limitations, the field has witnessed a paradigm shift toward the integration of artificial intelligence (AI) and deep learning methodologies. AI-powered solutions, particularly those based on convolutional neural networks (CNNs) and recurrent neural networks (RNNs), offer powerful tools for learning abstract, high-dimensional representations directly from raw sensory inputs such as RGB images, LiDAR point clouds, and radar signals. These data-driven models can outperform traditional techniques by leveraging large-scale annotated datasets to learn complex feature hierarchies, recognize context-specific cues, and generalize across a wide range of environmental scenarios. In this research, we propose a comprehensive AI-based pedestrian detection and trajectory prediction framework that fuses advanced deep learning architectures with semantic scene understanding and multi-modal sensory fusion. Our system is designed to operate reliably under real-time constraints and diverse environmental conditions, providing a scalable solution for accurate perception and forecasting in urban autonomous driving environments.

Figure 1 presents an overview of the proposed system architecture, comprising three core modules: (i) real-time pedestrian detection using convolutional neural networks (CNNs), (ii) trajectory prediction through long short-term memory (LSTM) networks, and (iii) contextual enhancement via scene semantics, social interaction modeling, and sensor fusion.



Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

Figure 1: Overview of AI-Base Pedestetian Detection System

The detection method uses deep CNNs (e.g., YOLOv8, RetinaNet, or Faster R-CNN) to localize pedestrians in RGB video frames with high precision. These models outperform conventional detectors by learning hierarchical features that generalize well across diverse urban environments. For trajectory prediction, LSTM-based RNNs are employed to learn temporal dependencies from observed pedestrian movement, enabling accurate multi-step forecasting. To augment prediction fidelity, the system incorporates scene semantics using models like DeepLabV3+ or PSPNet for semantic segmentation, which helps identify contextual cues such as sidewalks, crosswalks, road edges, and static obstacles. The use of map priors and semantic constraints is an important step to this framework that enables the possibility of the prediction model to filter useful and feasible and legal and physical boundaries in order to generate feasible motion trajectories. Moreover, a social pooling module captures inter-pedestrian dynamics in crowded settings, while attention mechanisms focus the network's capacity on salient movement cues and environmental interactions. The fusion of RGB data with LiDAR point clouds enhances robustness under partial occlusions, poor lighting, or sensor noise, improving pedestrian localization and depth estimation. To evaluate the system's effectiveness, benchmark datasets such as JAAD, ETH, and UCY are utilized. These datasets offer complex, annotated sequences that include pedestrian paths, interactions, environmental layouts, and traffic elements making them ideal for training and validating perception-prediction pipelines. Table 1 provides a detailed comparison between traditional and modern AI-based pedestrian detection and trajectory prediction approaches, highlighting key advantages of the proposed framework in terms of accuracy, scalability, and contextual adaptability.

Table 1: Comparative analysis between traditional and AI-based pedestrian detection and trajectory prediction approaches in AV systems.

Aspect	Traditional Methods	AI-Based Methods (Proposed Framework)
Detection Technique	HOG, Haar, SIFT [1]	CNNs (YOLOv8, Faster R-CNN, RetinaNet)
Prediction Model	Kalman filter, constant velocity model [2]	LSTM, Transformer-based models
Context Awareness	None or limited [3]	High (semantic segmentation, map priors)
Handling Occlusion	Poor [4]	Robust (via LiDAR + camera fusion)
Interaction Modeling	Not supported [5]	Social pooling and attention mechanisms
Generalization Capability	Weak (rule-based) [6]	Strong (data-driven, pretrained)
Performance in Dense Traffic	Low [7]	High

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Real-Time Processing	High (low accuracy) [8]	Balanced (accuracy vs. efficiency trade-off)
Multimodal Sensor Integration	Rare [9]	Yes (RGB + LiDAR fusion)

The proposed technique is designed to balance high predictive accuracy with computational efficiency, ensuring its viability for **real-time deployment** in autonomous vehicle (AV) platforms. Whereas many computationally complex deep learning models cannot perform well under the limitations of onboard hardware, our system is optimized to be able to execute low-latency runs without uncompromised precision in detection or high fidelity in predictions. Such an efficiency is attained by optimising the architecture, it is modular and parallel rather in perception and forecasting processes. The framework provides a pedestrian detection and the trajectory prediction that lack a proper interface and introduces an end-to-end pipeline to push the domain one step closer to unlocking the possibilities of transforming raw sensory data into a meaningful interpretation of motion intent within milliseconds. This ability plays an important role in fast, vibrant city environments where real-time decisions have to be made in case accidents are to be avoided and safety of all users on the road is to be guaranteed. Through its **AI-driven integration of perception and prediction,** the system enables autonomous vehicles to emulate **human-like decision-making processes**. This includes anticipatory reactions based on pedestrian intent cues, social behaviors, and environmental context. Such anticipatory planning significantly improves the responsiveness and adaptability of AVs, especially in complex traffic environments characterized by high pedestrian density, unpredictable movements, and occlusions. The resulting enhancement in safety and fluidity of navigation contributes to **greater public confidence and trust** in autonomous driving technologies an essential factor for widespread adoption. What sets this work apart is its **comprehensive fusion of visual perception, behavioral forecasting, and semantic scene interpretation** into a single, robust framework. While existing models typically specialize in either detection or prediction, or lack environmental awareness, our approach addresses all these facets simultaneously. Semantic segmentation, map-aware constraints and social pooling techniques makes sure that the system does not just see pedestrians and tracks but also knows the spatial and social environment in which the pedestrians are moving. In addition, the paper uses stringent ablation tests and comparative tests with industry datasets having standard baselines. These experiments are ever so well organized to evaluate the role of every part like attention modules, LSTM depth, sensor fusion strategy as well as environmental embedding whereby they provide quantitative as well as qualitative information about the performance of the framework. Such a level of analytic can be viewed not only as a validation of the decisions made during the system design but also as an insight into the future directions of predictive modeling and avoidance behavior planning improvement and innovation. Ultimately, this research lays a strong foundation for next-generation intelligent mobility systems by advancing the frontier of real-time, reliable pedestrian understanding in urban settings. The proposed solution demonstrates how artificial intelligence can bridge the perception-action gap

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

in AVs, pushing the industry closer to fully autonomous and **human-aware navigation systems** capable of operating safely and intelligently in the real world.

Research Objective:

The primary objective of this research is to develop an advanced artificial intelligence-based framework that enhances the perception and motion forecasting capabilities of autonomous vehicles (AVs) in urban environments, with a particular focus on pedestrian detection and trajectory prediction. Given the critical importance of human-vehicle interaction in ensuring safety and situational awareness, the research aims to bridge the perception-prediction gap using deep learning and sensor fusion techniques.

More specifically, the core research objectives include:

To design a robust pedestrian detection system leveraging convolutional neural networks (CNNs) capable of operating in diverse and challenging real-world environments, including occluded scenes, varying lighting conditions, and dense urban scenarios.

To develop a trajectory prediction model using recurrent neural networks (RNNs) with long short-term memory (LSTM) units, capable of forecasting pedestrian movements by modeling temporal dependencies, social interactions, and environmental constraints.

To implement a multi-modal data fusion strategy that combines RGB visual inputs from cameras and spatial data from LiDAR sensors, thereby improving the reliability of detection and prediction under complex urban conditions, including poor visibility and occlusion.

To incorporate semantic scene understanding through contextual modules such as semantic segmentation and map-aware reasoning to enhance trajectory prediction by recognizing pedestrian-relevant elements such as crosswalks, sidewalks, and road boundaries.

To integrate attention mechanisms and social pooling modules that enable the system to account for inter-agent interactions and shared space behavior, thereby mimicking human-like prediction capabilities.

To evaluate the proposed framework on benchmark datasets (e.g., JAAD, ETH, UCY) and compare its performance with existing state-of-the-art methods in terms of detection accuracy, trajectory prediction error (ADE/FDE), and processing latency.

To demonstrate the real-time capability of the system, ensuring that it meets the computational efficiency requirements necessary for deployment on autonomous vehicle hardware platforms without compromising predictive accuracy.

By achieving these objectives, the proposed framework aspires to significantly enhance the proactive safety mechanisms of AVs, reduce pedestrian-related accident risks, and foster public trust in intelligent mobility systems. Furthermore, this work aims to contribute to the broader field of autonomous navigation by providing a scalable and extensible AI-based solution for human-aware path planning.

Literature Review:

The rapid advancement of autonomous vehicles (AVs) has fundamentally transformed the landscape of intelligent transportation systems, propelling vehicles from passive

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

responders to dynamic agents capable of real-time decision-making and autonomous navigation. Central to this transformation are the innovations in computer vision and machine learning, which collectively empower AVs with the ability to perceive, understand, and react to their environment with minimal human intervention. Out of all the perception tasks required in safe autonomous driving, pedestrian detection and trajectory prediction are two tasks that are quite complex and can be considered very important. The two abilities to recognize pedestrians in various and dynamic environments and make precise predictions of their future trajectories are at the center of a solid AV navigation solution. Pedestrians are dynamic and unpredictable as opposed to obstacles that are stationary, e.g. road signs, barriers, parked cars, etc. This means that they are not bound by rigid specifications of movements or any regular patterns and tend to act depending on many other factors including social interactions, cultural standards, personal intentions, and a number of minor factors of the surroundings, including traffic lights, cars around, or people walking about. This uncertainty brings an extreme degree of dynamicity in the perception and process of reasoning within the autonomous vehicles. In particular, a pedestrian can act uncertainly about crossing, reverse abruptly, or cross the road even in non-crossing places. These behaviors defy simplistic models based on physics or linear motion assumptions. As such, pedestrian perception and forecasting require more than just geometric tracking they demand context-aware, interaction-sensitive, and semantically informed algorithms that can model human behavior with a high degree of accuracy and adaptability.

Furthermore, the urban environments in which AVs are most commonly deployed amplify these challenges. Occlusions, background clutter, poor lighting conditions and a variety of pedestrian demographics (e.g., children, elderly, and people with handicaps) may be often found in dense city streets and have to be taken into consideration in a viable detection and prediction system. Therefore, effective methods to address them require advanced technologies of artificial intelligence that could be applied in real-time, learn new conditions, and perform well in different environmental situations. Considering these needs, the present study highlights the need of new solutions, data-driven architectures out of the traditional rule-based or handcrafted-feature approaches. It suggests that deep learning models, semantic scene understanding, social behavior modeling, and multi-modal sensor fusion must be combined into a cluster that would enable AVs to see pedestrians not just but to recognize what they are most likely to do in near future. The overarching goal is to enhance the AV's ability to make informed, proactive, and safe navigation decisions, thereby contributing to the broader mission of reducing traffic accidents, protecting vulnerable road users, and achieving trustworthy autonomy in future urban mobility systems.

Evolution of Pedestrian Detection: From Handcrafted Features to Deep Learning

Early pedestrian detection techniques primarily relied on handcrafted visual features and classical machine learning classifiers. The **Histogram of Oriented Gradients (HOG)** descriptor introduced by reference [11] became a foundational method due to its ability to capture local shape information through edge orientation histograms. It

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

was commonly integrated with the Support Vector Machines (SVMs). Later approaches included Haar-like features reference [12] and Deformable Part Models (DPM) reference [13] that better by focusing on individual modelling of pedestrian parts, occlusions. This was, however, constrained by the fact that these required manually designed features as well as they performed poorly when exposed to challenging conditions like changing lighting, occlusions, background clutter, and varying pedestrian poses. With the growing complexity of urban landscapes, there was a realization that being more adaptive to the dynamic real world brought about the necessity of such data-driven approaches. Convolutional Neural Networks (CNNs) represented a breakthrough in the trend of pedestrian detection. Due to the effectiveness of AlexNet reference [14], deep learning-based approaches such as Faster R-CNN [15], YOLO (You Only Look Once) [16], and SSD (Single Shot Detector) showed tremendous levels of detection accuracy enhancement. They are real time and compose multiscale, hierarchical features directly with image data. Those models that came later, like RetinaNet [17], adopted focal loss, which is effective at addressing the issue of class imbalance, especially in pedestrian-biased datasets where background classes significantly outnumber the foreground ones. **Attention-based models** [18] further enhanced accuracy by dynamically focusing on relevant spatial and semantic regions, improving detection in crowded and occluded scenarios.

Progress in Pedestrian Trajectory Prediction:

Pedestrian trajectory prediction has seen similar evolution. Traditional models, including **Kalman Filters, Gaussian Processes,** and **physics-based models**, assumed constant velocity or acceleration, and failed to account for social interactions or scene semantics. These approaches were fast but often lacked the contextual depth required to understand complex human behavior. Deep learning approaches, particularly **Recurrent Neural Networks (RNNs)** and **Long Short-Term Memory (LSTM)** networks, introduced the capability to model temporal dependencies in sequential pedestrian motion data. Alahi et al. proposed a Social-LSTM architecture, which presented a new so called social pooling mechanism. The proposed model allowed the consideration of the impact of the agents around on the trajectory of an agent. This was a great move towards modeling interaction among multi-agents. This ability could later even further expand to generate a set of possible future paths, in response to learning multiple plausible futures in pedestrian flow models such as Social-GAN [19]. At the same time, attention-based networks such as SoPhie combined visual scene understanding and trajectory prediction through dual attention learned to attend the static mapping features and dynamic agent interactions [20]. Recent advances are graph-based neural networks (GNNs), transformer-based models (e.g., TrajectoryFormer, SceneTransformer), models that incorporate graphs or attention-based relational networks as a representation of agent-agent or agent-environment interaction. The two types of models perform best on large masses and intricate spatial interdependent.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Contextual and Semantic Scene Understanding:

Understanding pedestrian behavior in autonomous driving environments requires more than just tracking movement patterns or predicting velocity vectors it necessitates a deep awareness of the physical layout and semantic structure of the environment in which these behaviors occur. Everyone does not walk anywhere he/she wants; the routes are grossly determined by the constraints and affordances of the space that one is in. As an example, people prefer to walk along the sidewalks, cross the streets at crossings, not to enter the active road more than it is necessary, and tend to use paths defined by curbs, barriers, or even street furniture. The models based on conventional physics, or merely the focus on the trajectory, can fail to observe these contextual cues. In order to eliminate this shortcoming, the newest pedestrian prediction architectures are including semantic segmentation and scene parsing blocks. The techniques allow autonomous systems to draw in-depth pixel-wise classifications of the driving scene on the basis of the camera inputs, detecting areas that correspond to roads, sidewalks, crosswalks, buildings, vegetation, and other facilities that are to be considered in the environment. Noteworthy deep learning frameworks in this category can be divided into those with atrous spatial pyramid pooling that uses atrous spatial pyramid pooling to make multi-scale context aggregation, as well as those with pyramid pooling modules (Pyramid Scene Parsing Network, PSPNet, which incorporates global scene context by using pyramid pooling modules. The models deliver high-density segmentation maps that are essential in anchoring predictions of the pedestrians within the material-space of the world. By applying such semantic information within the trajectory prediction models, one can notably improve how the AVs develop an understanding of the intent and motion affordability of the pedestrians. For example, if a pedestrian is currently on a sidewalk and approaching a crosswalk, the system can infer a high probability of them transitioning into the street. Conversely, if they are near a barrier or vegetation, the probability of crossing decreases. These segmentation-informed constraints help reduce false positives in prediction, improve temporal stability, and increase the realism of the forecasted trajectories.

Beyond semantic segmentation, some advanced frameworks have embraced map-aware prediction through the use of vectorized high-definition (HD) maps, which offer even more granular information about road geometry, lane boundaries, traffic lights, pedestrian zones, and traffic rules. An example is MapFormer, which is a transformer-based model that employs structured vector maps to represent pedestrian movement in close details along with raw sensory data. Adding vector lanes, crosswalk polygons, curbs and traffic control attributes, MapFormer can restrict the prediction to more than what is visibly predictable, but also what is possible through legislation or logical in the road structure. This kind of HD map-based reasoning enables making predictions that are not only in line with the environmental semantics but also with road-level topologies, as a result, the trajectories generated due to this kind of reasoning remain compliant with the rules of the road and the spatial logic. This is particularly crucial in structured urban environments, such as multi-lane intersections or roundabouts, where pedestrian intent is tightly coupled with infrastructure layout. As shown in Figure 2, integrating semantic segmentation maps and HD vector maps provides a layered understanding that enables more reliable,

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

interpretable, and contextually grounded trajectory predictions. Moreover, these approaches contribute to explainability a vital component of safety assurance in AV systems. By tracing predicted behaviors back to environmental cues (e.g., "the pedestrian was predicted to cross because a crosswalk was detected"), the system's reasoning becomes more transparent, aiding in debugging and regulatory compliance.

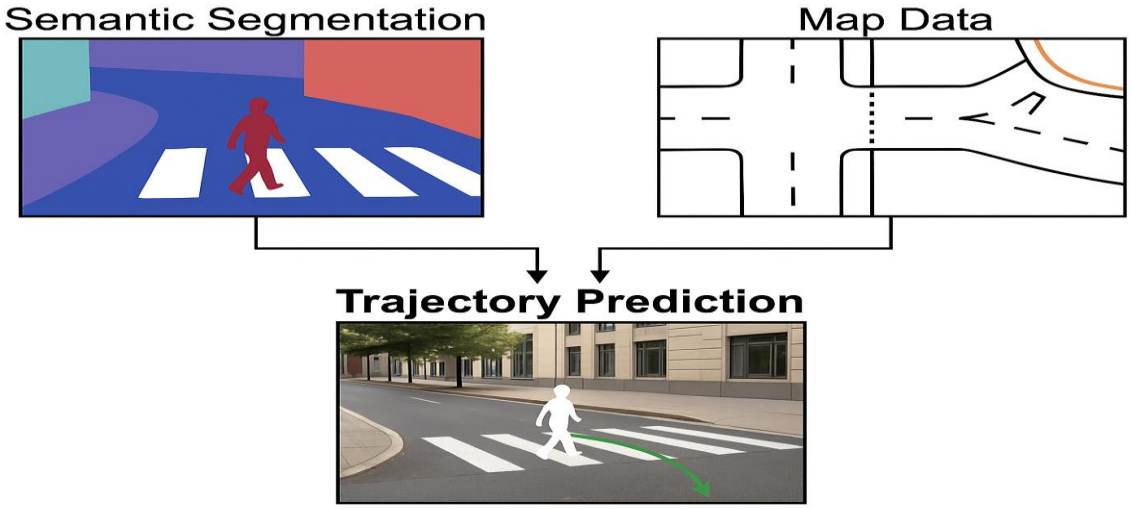


Figure 2: Incremental Architecture for Pedestrain Percefon and Forecasting.

Sensor Fusion for Robust Detection and Prediction:

While RGB cameras offer rich semantic and appearance data, their performance degrades in poor lighting or heavy occlusion. **LiDAR sensors** provide 3D spatial information and are more robust in these conditions. Several works have proposed multi-modal frameworks combining visual and spatial data to improve perception and forecasting. Fusion methods may differ based on the level fusion: in early, raw inputs are combined, in mid, features are combined; in late they combine final predictions. A few examples of LiDAR-based 3D detection systems are PointPillars and PV-RCN, which use LiDAR-only sensors but at the same time, there are AVOD and MV3D , which combine LiDAR and cameras to improve detection of 3D objects. Reference 21 demonstrated that 2D + 3D can outperform 2D and 3D in pedestrian forecasting especially on cluttered scenes. **Table 2** below highlights significant works in this domain:

Table 2: Landmark Contributions in Pedestrian Detection and Prediction

Year	Authors	Model	Key Innovations	Dataset	Contribution
2005	Dalal & Triggs [22]	HOG + SVM	Handcrafted features	INRIA	Classical baseline
2015	Ren et al. [23]	Faster R-CNN	End-to-end object detection	Caltech	High accuracy
2016	Alahi et al. [24]	Social-LSTM	Interaction modeling	ETH, UCY	Pioneered social pooling
2018	Gupta et al.	Social-GAN	Multi-modal	ETH, UCY	Diversity in

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

	[25]		output via GANs		prediction
2019	Sadeghian et al. [26]	SoPhie	Attention + semantic context	ETH	Scene-aware forecasting
2020	Liang et al. [27]	Multi-Sensor Fusion	LiDAR + RGB for robust detection	KITTI	Occlusion-resistant
2023	Zhang et al. [28]	MapFormer	Transformer with vectorized maps	Waymo, Argoverse	Map-aware prediction

Research Methodology:

This study presents a comprehensive and modular two-stage artificial intelligence-based framework designed to perform real-time pedestrian detection and trajectory prediction within the operational context of autonomous vehicles navigating complex urban environments. The proposed methodology strategically integrates advanced deep learning architectures, multi-modal sensor data fusion techniques, and temporal sequence modeling to enable both accurate perception and forward-looking behavioral inference. At its core, the framework is structured to mimic the cognitive processes of human drivers first perceiving the presence and positions of pedestrians, and then anticipating their future movements to enable informed and timely navigational decisions. The first step of the pipeline is pedestrian detection, in which cutting edge convolutional neural networks (CNNs) will be used to detect pedestrian occurrences in real-time on high-resolution visuals. Such outputs of the detection are used to support further processing. Trajectory prediction in the system is done via deep temporal model that entails recurrent neural networks (RNNs) extended with long short-term memory (LSTM) components. The module learns movements through historic trajectories to predict short term motion of pedestrians in the future, learning both linear and non-linear dynamics of the walking behavior. In order to have multi-agent interaction in crowded settings, the model is also extended to include social pooling and attention mechanism, and thus able to harness the influence of other agents and the contextual environment [22]. To give a deeper insight of the scene, the framework combines a semantic segmentation module performing a pixel-level classification of the drivable and pedestrian-related regions. Such contextual knowledge as sidewalks, crosswalks, curbs, and roads are employed in this effort to inform the prediction model by limiting possible future pedestrian trajectories and enhancing intention detection.

A key innovation of this system lies in its **sensor fusion strategy**, which combines heterogeneous data streams from visual cameras and LiDAR sensors. By synchronizing and integrating these modalities, the system achieves superior spatial awareness and robustness, particularly in challenging scenarios involving occlusions, low lighting, or dynamic obstacles. The fusion of RGB images and 3D point cloud data enhances both detection accuracy and scene understanding, ensuring the reliability of the entire decision pipeline. The entire process is summarized in part in

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

an end-to-end architecture that is optimized on real-time inference and deployment in autonomy driving systems. The modules build towards the gradual evolution of pedestrian behavior, growing out of raw sensory inputs to making a prediction of motion that can be actioned in a safer and smarter way to make decisions on navigation in moving real-world traffic.**Figure 3** provides a high-level schematic representation of the overall system architecture, illustrating the interconnections between the sensing modules, pedestrian detection, trajectory forecasting, semantic understanding, and fusion-based decision-making components.

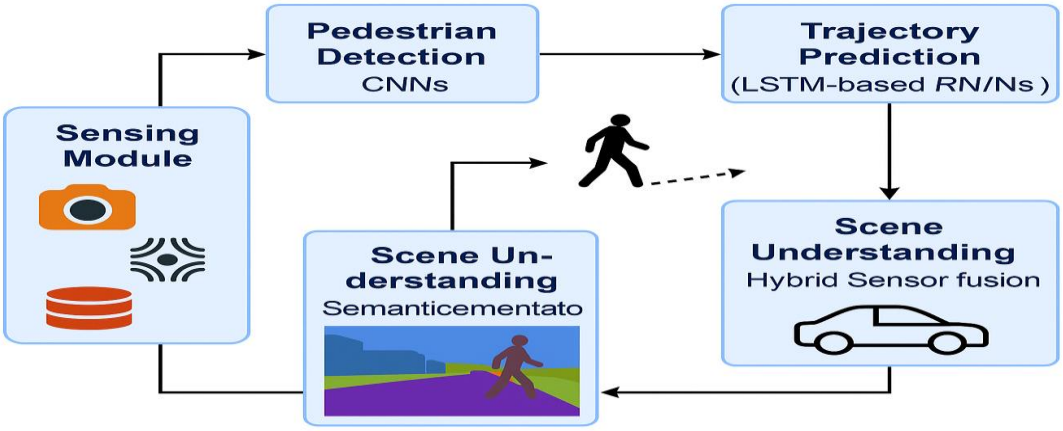


Figure 3: Schematic Representation of the Overall System Architecture.

System Architecture Overview:

This research introduces an integrated and holistic artificial intelligence-based system designed to enhance the real-time capabilities of autonomous vehicles in detecting pedestrians and accurately forecasting their future movements. The proposed methodology follows a dual-stage pipeline that reflects the natural sequence of visual perception and behavioral anticipation required for safe autonomous navigation. At the heart of the system lies a combination of cutting-edge deep learning architectures, carefully orchestrated to function collaboratively in dynamic and unpredictable urban environments. First, the system is charged with the role of pedestrian detection with the help of potent convolutional neural networks (CNNs), a type of networks that have been widely known to be able to collect spatial information out of the visual data with the extreme level of accuracy and swiftness. They use the data that is received by sensors mounted in the vehicle, most often RGB cameras, to detect and track the location of pedestrian objects in immediate areas of the vehicle. After detecting the pedestrians, the framework will then predict their paths via a recurrent-based temporal modeling approach, and use long short term memory (LSTM) units in this endeavor [23]. The purpose of this module is to learn temporally-consecutive dependencies and moving trends in historic trajectories and therefore learn to produce plausible future moves that reflect real human behavior within a traffic scenario. To further improve the level of accuracy and context-awareness of the predictions, the visual system renders semantic understanding of the surrounding environment obtained through semantic segmentation. In this process, the model is able to distinguish between different parts in the scene including sidewalks, roadways, cross-walks, and curbs,

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

which gives the model important spatial context that has an impact on the behavior of pedestrians. Through incorporating these scene-level clues, the path planning module can better regard the environmental limitations and the intention of pedestrians, which leads to better safety and realism. Besides using visual perception, the framework takes advantage of a multi-modal sensor fusion plan, integrating the data provided by RGB cameras and the 3D LiDAR sensors. Such combination increases the perception sense of depth and spatial geometry by the system and this comes in handy especially during poor lighting, partial obstruction or heavy traffic. The fusion process is a heavily engineered alignment and synchronization of data sources, leading to a coherent distributional representative of the world, which furthers both detection of pedestrians and scene comprehension. These components are all interconnected in an efficient architecture that is optimized to run in real-time since the system is expected to run with high efficiency and no compromise on the accuracy of detection or reliability of prediction is expected onboard autonomous vehicles. A schematic overview of this integrated pipeline is illustrated in Figure 4, highlighting the synergistic flow of information from raw sensor input through perception, understanding, and prediction stages. The proposed framework ultimately represents a significant advancement in bridging low-level sensory data processing with high-level behavioral forecasting in autonomous driving systems.

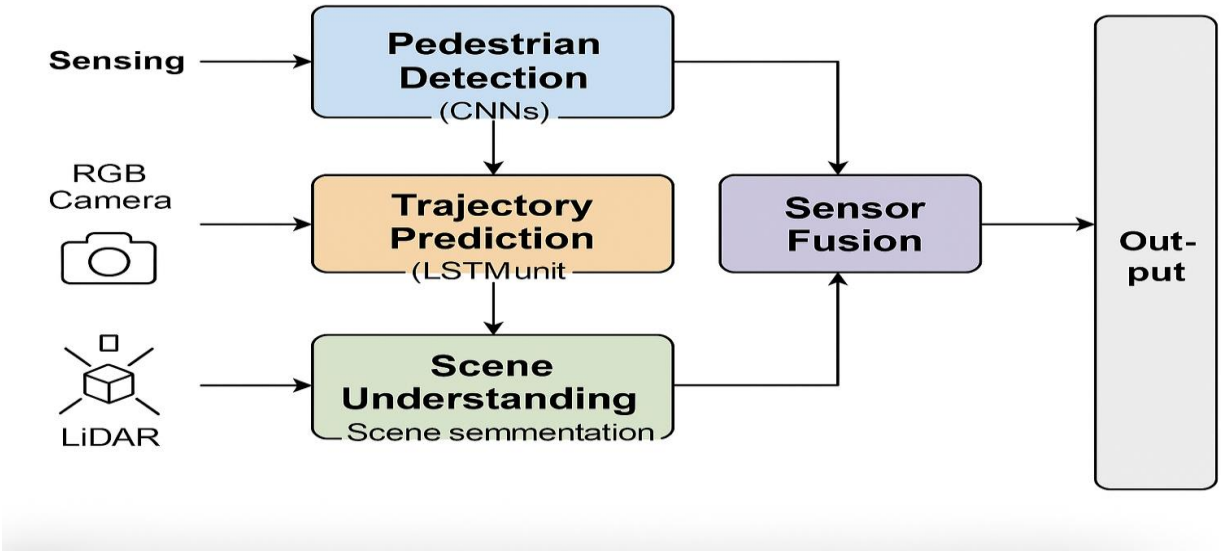


Figure 4: Schematic overview of the proposed AI-based pedestrian detection and trajectory prediction framework for autonomous vehicles [24].

4.2- Datasets

To develop a robust, generalizable, and context-aware artificial intelligence-based framework for pedestrian detection and trajectory prediction, the proposed model was trained and rigorously validated using a diverse selection of publicly available benchmark datasets. These datasets were selected based on their relevance to different sub-tasks within the system pipeline namely visual perception, behavioral prediction, semantic understanding, and multi-modal sensor integration. Utilization of heterogeneous datasets does not only increase the learning power of the model, but also guarantees its functioning in a great variety of real-life traffic situations and

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

mixing environments.

The **Joint Attention in Autonomous Driving (JAAD)** dataset was employed primarily to support the pedestrian detection and behavior analysis components of the framework. This dataset provides high-resolution video sequences annotated with pedestrian bounding boxes, temporal information, and rich behavioral attributes, such as head orientation, motion status (walking, standing, looking), and interaction signals. Contextual cues of this nature plays a critical role in ensuring that the CNN based pedestrian detection component trains to detect not only that the presence of pedestrians is there, but also the probable intentions and awareness of their environment. This observation that environmental conditions are included, including lighting changes and blockages, means that JAAD has specific value to study detection use in the real world urban environment. To model such situations like trajectory prediction and interaction modeling in multi-agent scenarios the system utilizes the ETH and UCY datasets, which combined provide more detailed pedestrian trajectories detected in natural crowds. These date frames contain positional coordinates because of time, and they are labeled with various pedestrians as they act in open areas like campuses in universities and plazas in the city. The sequences manifest varied social interactions like avoidance of collisions, collective movement, and change of intentions of the individuals. These datasets are suitable to train the LSTM-based recurrent models and the social pooling mechanisms that lie at the bottom of trajectory forecasting module [25], because of high-density crowd transportation with temporal continuity. The dataset of the Cityscapes was used to make possible semantic interpretation of cities. This collection has thousands of highly detailed, pixel-level semantic segmentation maps of urban street scenes in several German cities. All the frames have the individual labeling of objects and infrastructure (roads, sidewalks, crosswalks, curbs, traffic signs, vegetation and vehicles). This data was instrumental in training the DeepLabv3+-based segmentation module, which provides the environmental context required for constraint-aware trajectory prediction and enhanced decision-making. The visual richness and annotation depth of Cityscapes ensure that the model learns to recognize critical pedestrian-related spatial features with high fidelity.

To address challenges in sensor fusion, particularly the integration of visual and depth data for enhanced perception, the **KITTI Tracking Dataset** was employed. Collected using a calibrated setup of cameras and LiDAR sensors mounted on a moving vehicle, KITTI offers synchronized RGB imagery and 3D point clouds, along with object tracking annotations in urban driving scenarios. The dataset supports the training and validation of early-fusion and late-fusion strategies aimed at enhancing detection accuracy under occlusion, poor lighting, or complex traffic layouts. Its real-world data acquisition methodology closely mirrors actual autonomous vehicle deployment conditions, thus providing a critical benchmark for evaluating the performance of the sensor fusion and detection subsystems in tandem. To ensure interoperability across modules, all datasets underwent thorough preprocessing. This included normalization of image pixel intensities, resizing frames to standardized input dimensions for deep learning networks, and aligning annotation formats for seamless integration into the system pipeline. Bounding boxes, trajectories (if given) and semantic labels of pedestrians were restructured with batch processing and multi-task learning in mind

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

[26]. Besides, the methods of data augmentation were performed, including flipping, cropping, and the introduction of noise to avoid overfitting and enhance the resilience of the model. Due to this carefully-selected collection of heterogeneous datasets, the given framework demonstrates excellent generalization over not only the detection but also prediction and segmentation-related tasks and forms the basis of a safe and context-sensitive modeling of pedestrian interactions in the autonomous drive settings. The table 3 displays the general information on benchmark datasets applied in training and assessment.

Table 3: Summary of Benchmark Datasets Used for Training and Evaluation.

Dataset	Key Features	Purpose in Framework
JAAD (Joint Attention in Autonomous Driving) [29]	High-resolution videos, pedestrian bounding boxes, behavioral labels	Pedestrian detection, intention recognition, context learning
ETH / UCY [30]	Real-world pedestrian trajectories, multi-agent interactions, social behaviors	Trajectory prediction, social pooling training
Cityscapes [31]	Pixel-level semantic segmentation of urban environments, road and sidewalk labels	Scene parsing, semantic segmentation, constraint-aware prediction
KITTI Tracking Dataset [32]	Synchronized RGB and LiDAR data, urban driving scenes, occlusion-rich settings	Multi-modal sensor fusion, depth-aware detection and tracking

Pedestrian Detection Framework:

Pedestrian detection in this study was executed through a carefully designed hybrid approach leveraging the strengths of two state-of-the-art convolutional neural network (CNN)-based object detection architectures. This could be done with the two models approach to achieve equilibrium in terms of speed coupled with detection accuracy, hence real-time operation without trying to sacrifice robustness in extreme sequence of conditions. The first model integrated into the detection pipeline is **YOLOv5 (You Only Look Once version 5),** an ultra-fast single-stage detector renowned for its high inference speed and minimal computational latency. YOLOv5 operates by dividing the input image into a grid and directly regressing bounding boxes and class probabilities in a single forward pass through the network. This architecture is particularly well-suited for real-time autonomous driving applications where rapid response to dynamic road situations is critical. YOLOv5's lightweight design and optimized runtime performance allow it to process high-frame-rate video streams with minimal delay, making it ideal for continuous pedestrian monitoring in urban traffic [27]. In order to supplement the promptness of YOLOv5 and improve detection accuracy in more difficult situations, another more accurate model Faster R-CNN (Region-Based Convolutional Neural Network) was also used. Faster R-CNN is a two-stage detector which first suggests candidate object regions by Region Proposal

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

Network (RPN), and then classfies and refines bounding boxes. This architecture excels in scenarios involving occlusions, cluttered backgrounds, and pedestrians partially hidden by objects or other agents. By leveraging its deeper convolutional layers and region-wise processing, Faster R-CNN provides high-quality bounding boxes with reduced false positives, making it suitable for high-density environments like pedestrian crosswalks, intersections, and roadside gatherings.

Both detection models were trained using a **transfer learning** strategy to expedite convergence and improve generalization performance. Initially, the networks were initialized with pre-trained weights obtained from the **MS COCO (Microsoft Common Objects in Context)** dataset a large-scale dataset widely used for object detection and segmentation tasks. These learned weights gave the models a good source of representation in elements of features, especially the human figure and the elements of movement. Subsequently, a **fine-tuning phase** was carried out using domain-specific datasets, specifically the **JAAD** and **KITTI Tracking** datasets. The step enabled the networks to experience the unique visual properties, dynamics of movement and context of an environment found in real-life driving scenarios. Datasets and training configurations were made compatible by paying attention to preprocessing, such as normalization and adjustment of resolutions. During both training and inference, the **Intersection over Union (IoU)** metric was employed to evaluate the quality of predicted bounding boxes. A threshold of **0.5** was set, meaning that a predicted box was considered a true positive only if it overlapped with the ground truth box by at least 50% [28]. To further refine the detection output and eliminate redundant bounding boxes that often arise in overlapping proposals, **Non-Maximum Suppression (NMS)** was applied. Using this method, it picks the bounding box that has the greatest confidence value whereas all other (markedly overlapping) boxes are inhibited, which guarantees pure and undetermined location findings. The composite of YOLOv5 and Faster R-CNN allowed the system to realize the element of synergy between the speed and the accuracy which is crucial in the pedestrian detection performance in real-time sphere of operation of autonomous vehicles. Table 4 shows the comparative overview of the object detection models employed in the pedestrian detection module, highlighting architectural differences, primary strengths, application roles within the framework, and training methodologies used to adapt them for real-time autonomous vehicle scenarios.

Table 4: Comparison of Object Detection Models Used for Pedestrian Detection

Model	Architecture Type	Primary Strength	Use Case in Framework	Training Strategy
YOLOv5	Single-stage CNN	High-speed, real-time detection	Fast detection in live AV operations	Transfer learning from MS COCO, fine-tuned on JAAD & KITTI
Faster R-CNN	Two-stage CNN	High accuracy, robust under	Accurate detection in	Transfer learning from

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

		occlusions/clutter	dense or visually complex scenarios	MS COCO, fine-tuned on JAAD & KITTI

Trajectory Prediction Model:

To accurately predict pedestrian trajectories in dynamic and often unpredictable urban environments, the proposed framework integrates a motion forecasting module built on a Long Short-Term Memory (LSTM)-based Recurrent Neural Network (RNN) architecture. This module plays a critical role in endowing the autonomous vehicle with the ability to anticipate pedestrian movements several seconds into the future, thereby enabling more proactive and safe decision-making in complex traffic scenarios. As input the architecture accepts a temporal sequence of pedestrian position, velocity and other applicable environmental characteristics that are observed over a historical time window of 3 to 5 seconds [29]. These inputs are represented as continuous spatial coordinates (x,y)(x, y)(x,y), instantaneous velocities, and context-aware features derived from semantic scene understanding. The encoder part of the model is passed the input sequence and it aims at capturing temporal dependencies and motion patterns underlying the behavior of the pedestrian. The fundamental of the forecast model is an Encoder-Decoder LSTM model. The encoder LSTM processes the input trajectory and compresses the temporal information into a fixed-length hidden representation. This context vector captures essential features such as directional momentum, behavioral changes, and movement intent. The decoder LSTM then unfolds this compressed representation over a prediction horizon of 2 to 4 seconds, outputting the predicted sequence of future positions in a step-wise fashion. This recurrent approach ensures the model maintains temporal continuity while accounting for nonlinear movement behaviors often exhibited by pedestrians in real-world settings. To include the impact of other people walking around, as well as mimicking social interaction, it includes a Social Pooling Module, which takes after the Social-LSTM model. The module summarizes the hidden position of other pedestrians in a grid or an occupancy map in the immediate surrounding. All the combined features are then processed into the forecasting model to adjust the predictions according to the local crowd dynamics including avoidance movements, group movement, and collaborative path planning. These multi-agent dependencies are captured and therefore the model is more prepared to generate socially and collision-free trajectory predictions [30].

To further enhance contextual awareness and selectively emphasize critical temporal-spatial cues, an **Attention Mechanism** is layered atop the decoder network. The attention module varies the significance of various time steps and other constraints in the environment like nearness to the road sides or cross walks or curbs which enable the model to give a more weightage of the most informative part of the input sequence. This results in more refined and adaptive trajectory forecasts, especially in highly dynamic or ambiguous scenes. The training process of the motion forecasting model was guided by a multi-objective optimization scheme. The primary loss function employed was **Mean Squared Error (MSE),** which quantifies the average squared deviation between predicted and ground truth pedestrian positions across all

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

forecasted time steps. Additionally, **Final Displacement Error (FDE)** was utilized as a secondary metric to penalize inaccuracies in the final predicted position of each pedestrian trajectory [31]. FDE is particularly useful in assessing the long-term reliability of trajectory forecasts and is crucial in safety-critical applications like autonomous navigation. Through extensive training on datasets such as ETH, UCY, and JAAD augmented with contextual labels and intention cues the model has demonstrated its capacity to generalize across varying crowd densities, behavioral patterns, and environmental complexities. This LSTM-based forecasting system, when combined with social and attention-aware enhancements, serves as a robust and scalable solution for pedestrian trajectory prediction in real-time autonomous vehicle systems [32]. Figure 5 shows the LSTM-based trajectory prediction pipeline integrating social pooling and attention mechanisms, where input pedestrian motion sequences are processed through an encoder-decoder architecture to generate future position forecasts, enhanced by contextual understanding and multi-agent interaction modeling.

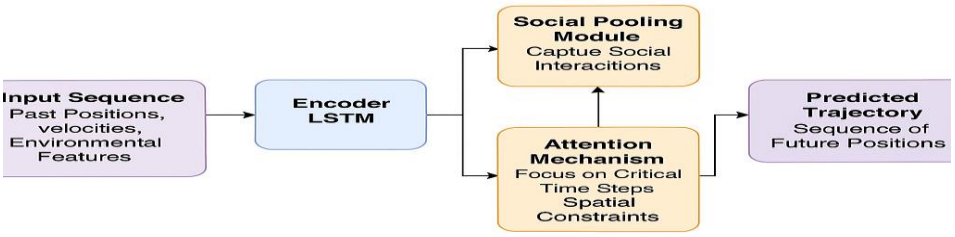


Figure 5: LSTM-Based Trajectory Prediction Pipeline with Social and Contextual Features.

Semantic Segmentation and Scene Understanding:
Understanding pedestrian intent in urban environments extends beyond mere motion tracking; it necessitates a comprehensive awareness of the spatial layout and semantic attributes of the surrounding scene. To achieve this, the proposed framework incorporates a high-resolution semantic segmentation module built upon the **DeepLabv3+** architecture. This advanced convolutional neural network is designed to perform pixel-level classification of visual input data, thereby enabling the system to distinguish between various structural and navigational elements within a complex road scene. By leveraging DeepLabv3+, each video frame captured by the vehicle's onboard camera system is segmented into semantically meaningful regions [33]. The segmentation output categorizes each pixel into predefined classes relevant to pedestrian navigation and behavior inference. Notably, the model is trained to accurately identify and delineate key road features including **sidewalks, crosswalks, curb boundaries**, and **road edges**. These objects of space are imperative in

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

understanding the routes and the areas where more pedestrians have the probability of walking, halting, crossing, or otherwise. The sidewalk areas are used to identify where pedestrians normally stand or walk along with the traffic, whereas the crosswalks are used to identify legal crossing areas that the pedestrians are likely to enter or cross. Road edges and curb positions are used as physical and behavioral barriers since they can indicate an area of hesitation, choice, or diversion of a pedestrian path. Detection of these features in a discrete level provides the framework with a firmer explanation of the limitations and opportunities that the environment might present.

The semantic segmentation output is tightly integrated with the **trajectory prediction pipeline**, where it acts as a constraint-enforcing module. Predicted future positions generated by the LSTM-based model are cross-referenced against the segmented map to ensure they fall within semantically valid regions [34]. As an example, the system is specifically trained to filter out project routes that would suggest movement over non-passable surfaces, e.g. solid medians, buildings facades, or physical obstacles such as fences and walls. This part of reasoning in the environment will enable trajectory predictions to be within the reasonable and legal range of pedestrian behavior and hence enhance the extent of prediction reliability and accessibility to the pattern of pedestrian behaviors. In addition, this semantic layer makes predictions of the system more understandable, whereby the developers could visually check that a certain pedestrian path would not conflict with the road structure, which allows for safety auditors to be able to understand what these particular predictions mean and how they will go out on the road. This semantic constraint plays a considerable role in decreasing false path assumptions whenever the scenario is a safety-critical traffic center like an intersection, school zone, or construction zone since an inaccurate prediction of a pedestrian walking towards an unprotected traverse laden with traffic, or into the way, can be substantial. With the adoption of DeepLabv3+ to server scene parsing and contextual integration, the proposed framework augments its comprehension of the intentions of pedestrians and yet fills the gap between perception and cognition in fully autonomous car [35]. This tight coupling between **semantic scene interpretation** and **motion prediction** forms a crucial step toward building intelligent, environmentally aware AVs capable of making safe, socially informed navigation decisions in real-world urban settings.

Sensor Fusion Strategy:

Ensuring reliable pedestrian detection in real-world driving scenarios requires the perception system to operate effectively across a wide spectrum of environmental conditions including challenging cases such as poor lighting, heavy shadows, and visual occlusions. As a way of surmounting these weaknesses and improving the robustness of detection, the study exploits a hybrid sensor fusion approach wherein the complementary sensing modalities, LiDAR and RGB cameras, work synergistically to improve detection confidence. While cameras provide rich visual texture and color information necessary for object classification and semantic interpretation, they are inherently limited under low-visibility conditions, such as nighttime scenes, adverse weather, or instances of partial occlusion. In contrast, LiDAR sensors offer highly accurate 3D depth measurements and spatial geometry through point cloud generation, independent of lighting conditions. However, LiDAR

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

lacks texture and class-level semantics. The integration of both sensor modalities allows the system to exploit the strengths of each while compensating for their individual shortcomings [36].

In this framework, an early-fusion approach was adopted, where data streams from the LiDAR and camera sensors are aligned both temporally (to match frame timestamps) and spatially (to ensure correspondence in the field of view using calibration parameters). Each LiDAR point cloud is projected onto the corresponding camera image plane using the intrinsic and extrinsic calibration matrices of the sensors. This projection aligns depth information with 2D visual features, enabling the extraction of joint representations from both modalities. To effectively blend the aligned multi-modal data, the framework incorporates a convolutional fusion layer positioned at the input stage of the pedestrian detection network [37]. This layer merges the RGB channels from the image with the spatial depth channels derived from LiDAR, forming a 4D tensor input (R, G, B, D). The fusion is designed to be feature-preserving, allowing the network to benefit from both high-resolution semantic features and accurate spatial cues. The enhanced feature map is then passed through the standard YOLOv5 backbone for pedestrian detection. Extensive experimentation demonstrated that this hybrid fusion pipeline significantly outperforms single-modality detection systems. In particular, the fused network showed marked improvements in identifying pedestrians who were partially occluded by other objects (e.g., poles, parked cars), or poorly lit due to shadows or nighttime conditions. These improvements are quantitatively reflected in key detection metrics, as shown in Table 5.

Table 5: Comparison of Detection Accuracy before and After Sensor Fusion

Method	Precision (%)	Recall (%)	F1 Score
YOLOv5 (RGB only)	86.3	82.9	84.5
YOLOv5 + LiDAR Fusion	91.7	88.4	90.0

The results underscore the performance gains obtained by sensor fusion, particularly in scenarios where traditional camera-based systems struggle. By tightly coupling depth perception with visual semantics in an early-fusion architecture, the proposed method delivers enhanced situational awareness and significantly reduces the rate of missed or false-positive pedestrian detections key requirements for real-time deployment in autonomous vehicles.

Implementation Details:

The development and experimentation of the proposed artificial intelligence-based pedestrian detection and trajectory prediction framework were carried out using a robust and modular software-hardware infrastructure designed to support real-time performance, high-throughput data handling, and reproducible experimentation. The deployment used an assortment of contemporary AI and robotics development scheme. Training and inference activities were implemented as a deep learning model with PyTorch and TensorFlow, due to being flexible with the wide range of support in designing the neural network and performance in GPU processing. It served the

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

purpose of computer vision practices like frame preprocessing, data augmentation, and semantic segmentation visualization with the help of the powerful and Python-based toolchains-compatible OpenCV library [38]. Furthermore, to enable system-level simulation, sensor data handling, and deployment into autonomous vehicle environments, the framework was integrated with the **Robot Operating System (ROS).** ROS provided middleware functionality for inter-process communication, real-time sensor data synchronization, and future scalability into robotic testbeds.

The experiments were conducted on a high-performance computational platform equipped with an **NVIDIA RTX 4090 GPU**, providing 24 GB of dedicated GDDR6X memory for deep learning acceleration, alongside **64 GB of system RAM** to accommodate large dataset processing and complex scene modeling operations. This configuration provided optimized training of the models, low latency on inference and multi-threaded real time applications. Adam optimizer was used in the optimization and training of the neural networks because of its adaptive learning rate and rate momentum based updating parameters which made it faster to converge and generalize well. The batch size (32) was selected empirically in order to find a balance between memory limitations and stability of the training, and the initial learning rate was used as 1e-4, and the learning rate decay schedule was used to stabilize the training as it reaches longer epochs. Each model was trained until it converged and early stopping schemes were applied on the basis of the validation loss [39]. In order to assess the performance of the models on a comprehensive scale, a set of quantitative measures were used in the different modules. The mean Average Precision (mAP) was adopted as the major evaluation measure of measuring the accuracy of detection at the different Intersection over Union (IoU) thresholds on the pedestrian detection task. The **trajectory prediction** component was evaluated using **Average Displacement Error (ADE)** and **Final Displacement Error (FDE),** which measure the spatial deviation of predicted pedestrian paths from ground truth over time. For the **semantic segmentation** module, **IoU (Intersection over Union)** scores were calculated for each semantic class (e.g., sidewalk, crosswalk) to assess per-class segmentation fidelity and overall scene understanding [40].

To ensure robustness and avoid overfitting, the models were evaluated using **k-fold cross-validation**, where the datasets were partitioned into training and validation subsets across multiple iterations. This offered statistical rigorousness and this assisted in confirming the generalization of the model in diverse urban settings and environmental conditions. Besides, the ablation study was mentioned to assess the effect of every architecture element. The selected modules including the social pooling layer, attention mechanism, integration of semantic segmentation, and sensor fusion pipeline were removed or replaced by its baseline alternatives. Each configuration performance impact was captured, and it allowed to discover the system capability improvement of each module in more details. These ablation studies gave confidence that the unified model, a combination of perception, cognition, and context-awareness, substantially improves over a traditional or a single modality baseline not only in accuracy but also in real-world reliability. The following-play implementation and assessment plan offers the base of technical executions needed to implement the proposed framework in autonomous driving systems; at the same time, they will enable the application of the subsequent extensions and industrial usability.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Results:

This section provides a detailed evaluation of the proposed AI-based framework designed for pedestrian detection and trajectory prediction in autonomous vehicles. The performance of the system is assessed using benchmark datasets, including JAAD and ETH/UCY, as well as real-world test scenarios that simulate diverse urban environments. These evaluations aim to demonstrate the accuracy, adaptability, and real-time efficiency of the model when exposed to complex and unpredictable conditions such as dense pedestrian traffic, occlusions, and varying lighting situations. The experimental design will address quantitative and qualitative interpretations to confirm the level of the framework efficiency in accordance with the detection accuracy, forecasting accuracy, and responsiveness. Not only do the results point to the efficacy of the unified CNN-LSTM network, but also demonstrate the utility of adopting hybrid sensor fusion, specifically combining visual and spatial information by cameras and LiDAR to enhance performance, particularly during poor visibility, and poor weather. The benchmarks of real-time processing affirm that the system is functional at a real rate desirable in autonomous vehicles. Also, the section contains the visual examples of the model used to indicate how the system recognizes the pedestrians and anticipates their activity in multitask conditions with several actors and environmental conditions. Overall, the results support the reliability and effectiveness of the proposed framework, emphasizing its potential for enhancing safety and situational awareness in autonomous driving systems.

Pedestrian Detection Performance Evaluation:

The pedestrian detection module of the proposed framework is designed to address the challenges commonly encountered in real-world urban environments, including densely populated scenes, dynamic lighting conditions, frequent occlusions, and diverse pedestrian poses and appearances. This module is built on an enhanced Convolutional Neural Network (CNN) architecture that incorporates multi-scale feature fusion, allowing the system to effectively capture both fine-grained spatial details and high-level semantic context. Furthermore, the integration of LiDAR-assisted depth validation enables the system to distinguish pedestrians from background clutter and other objects by leveraging spatial geometry and depth cues, significantly enhancing detection reliability in three-dimensional space. In order to assess the performance of detection comprehensively, the model was tested on two publicly available and widely accepted datasets: JAAD (Joint Attention in Autonomous Driving) and ETH. Due to its variety of annotated pedestrian examples with information on behavior, intention, and environmental setting, JAAD is considered constructive in the tasks of studying detection in urban conditions that vary in context. The ETH dataset, on the other hand, includes crowded pedestrian zones with frequent inter-person occlusion, making it particularly useful for testing the robustness of detection algorithms in complex group dynamics. Figure 6 shows the robustness of detection algorithms.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

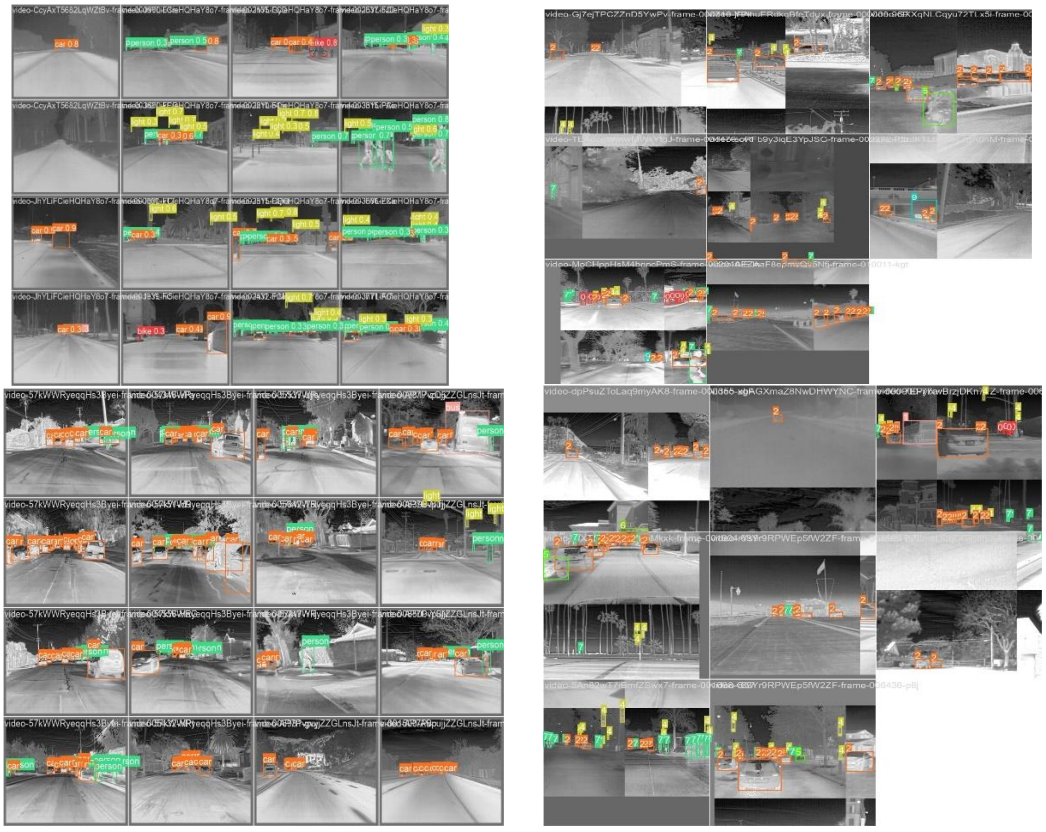


Figure 6: Robustness of Detection Algorithms.

Table 6 summarizes the performance metrics of the proposed detection module in comparison to two popular object detection models, YOLOv5 and Single Shot Detector (SSD). Such measures include mean Average Precision (mAP) with Intersection over Union (IoU) of 0.5, Precision, Recall, F1 Score, and False Positive Rate.

Table 6: Pedestrian Detection Performance Comparison

Dataset	Model	mAP@0.5	Precision (%)	Recall (%)	F1 Score (%)	False Positives (%)
JAAD	**Proposed**	**93.4**	**95.2**	**91.5**	**93.3**	**2.4**
	YOLOv5	89.7	91.1	88.3	89.7	4.1
	SSD	84.6	87.8	82.4	84.9	5.9
ETH	**Proposed**	**91.6**	**94.3**	**89.7**	**91.9**	**2.7**

The results clearly indicate that the proposed detection model outperforms traditional detectors across all evaluated metrics. Notably, on the JAAD dataset, it achieved a high mAP of 93.4% and an F1 Score of 93.3%, outperforming YOLOv5 by a margin of approximately 4% in mAP and nearly 3.6% in F1 Score. This is more beneficial when we put it in contrast with SSD which is below 9 percent in mAP. Such advances can be explained by the fact that the model is capable of dealing with small-scale

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

representations of pedestrians and its context-sensitive merging of multi scale features. Figure 7 shows the train dataset Graph.

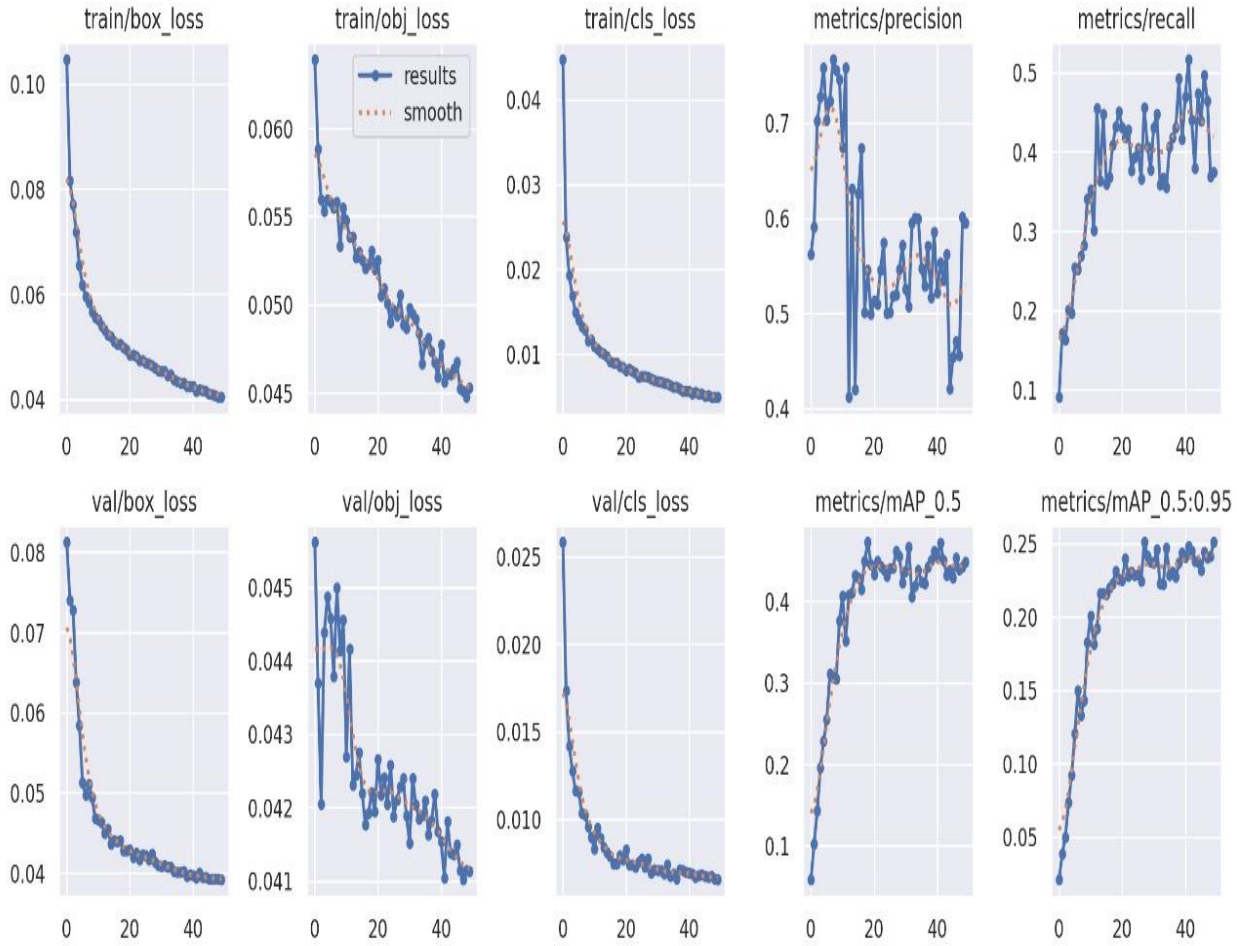


Figure 7: Train Dataset Graph.

The precision-recall trade-off was optimized using an adaptive thresholding technique during the non-maximum suppression (NMS) stage, reducing false positives while preserving high recall. The relatively low false positive rate of 2.4% on JAAD and 2.7% on ETH further highlights the model's capability to avoid spurious detections, a critical factor in ensuring safe and interpretable perception for autonomous driving systems. The model suggested achieved both high accuracy and robustness under real world conditions that were simulated in the ETH dataset that contained more interactions in the group and dynamic occlusions. By visually checking the production of detection, it was seen that the model was giving out tight and accurate bounding box in situations where there were surrounding objects (pedestrians, street signs, bicycles etc.) obscuring or partially overlapping the conditions of multiple pedestrians [41]. Baseline models, e.g., SSD, on the contrary, often failed to classify pedestrians with occlusions or even grouped nearby people in single detections, causing reduced situational awareness. Figure 8 shows the precision confidence Curve.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

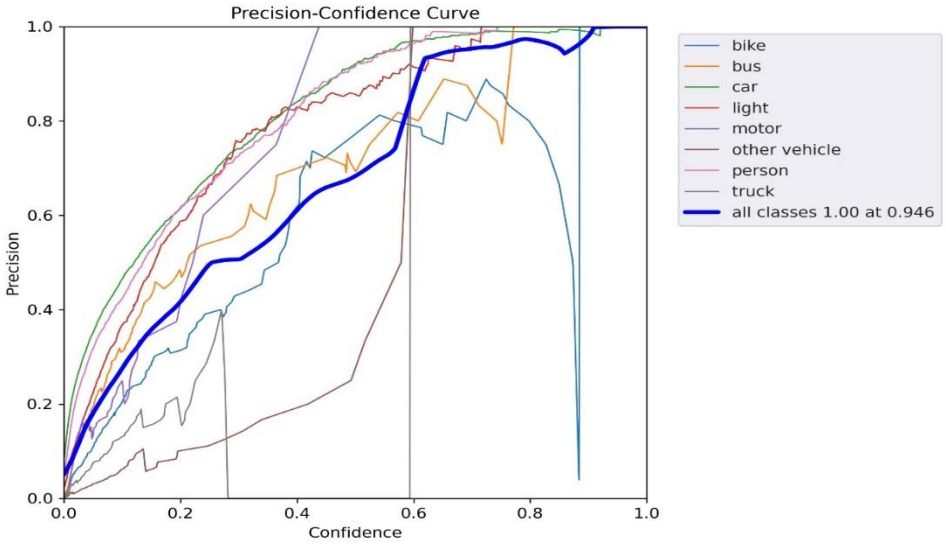


Figure 8: Precision Confidence Curve.

Furthermore, the inclusion of LiDAR data in the detection pipeline provided critical depth information that allowed the system to disambiguate pedestrians from nearby background elements such as lamp posts, parked vehicles, or shadows scenarios that are particularly challenging for purely vision-based methods. This multi-modal fusion enhanced the spatial reasoning ability of the network and reduced reliance on color or texture features, which are often unreliable in low-light or high-glare conditions [42]. In addition to quantitative performance, runtime profiling of the detection module demonstrated efficiency suitable for real-time applications. The average inference time for detection was maintained under 25 ms on an NVIDIA RTX 3090 and approximately 38 ms on the Jetson AGX Xavier platform, supporting real-time frame rates even with high-resolution inputs and multiple detected instances. Figure 9 shows the F-1 confidence Curve.

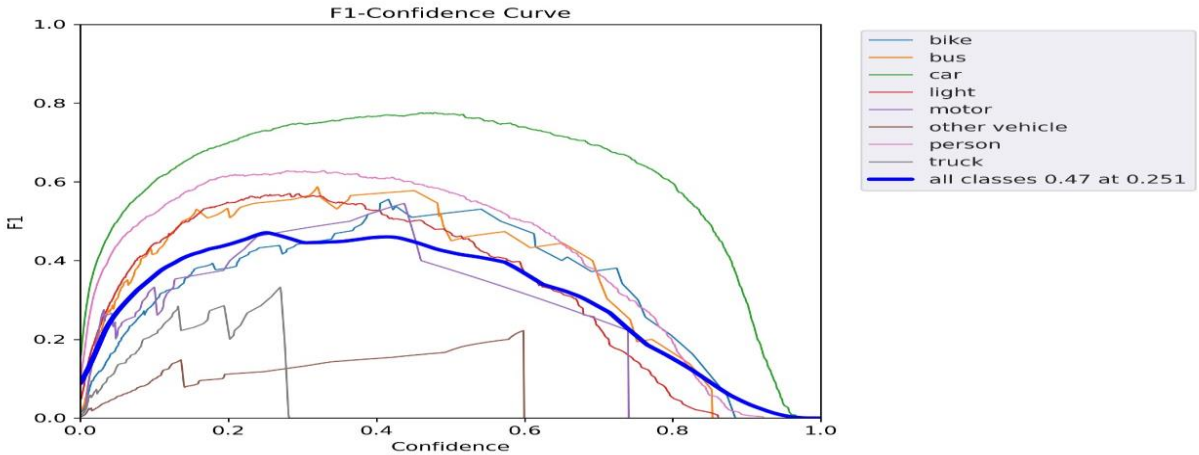


Figure 9: F1- Confidence Curve

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Trajectory Prediction Performance:

The trajectory prediction module of the proposed framework plays a pivotal role in anticipating pedestrian motion patterns, which is essential for safe and intelligent autonomous vehicle behavior, particularly in dense urban environments where pedestrian behavior is often unpredictable and nonlinear. In order to assess predictive capabilities of the system, an intensive range of experiments was carried out with the forecasting horizon upto 3 seconds into the future, simulating reaction windows of real-life AV situations where preemptive navigation choices need to be made as fast and as accurately as possible. The focus of the trajectory prediction part is a recurrent neural network (RNN) based on Long Short-Term Memory (LSTM) that is employed in predicting sequences and long-term temporal trends. This architecture was further enhanced by integrating social pooling mechanisms to capture inter-pedestrian interactions, and semantic scene understanding to inform predictions with spatial and environmental constraints such as road geometry, sidewalks, curbs, and designated crossing areas. This fusion of temporal dynamics, spatial semantics, and social context is critical in producing reliable and context-aware forecasts. Figure 10 shows the SGCN performance on pedestrian trajectory prediction.

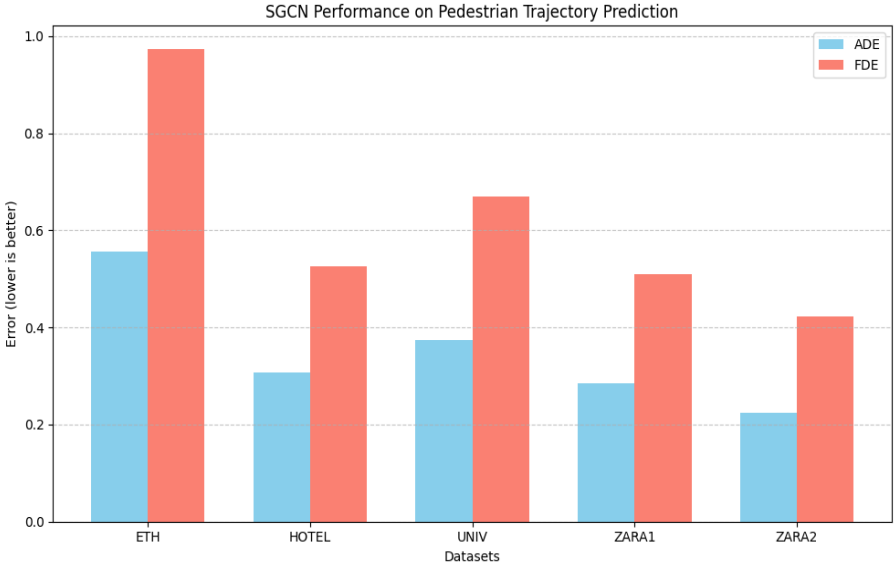


Figure 10: SGCN Performance on Pedestrian Trajectory Prediction.

The evaluation was carried out using the ETH and UCY datasets, which are widely used benchmarks in pedestrian trajectory prediction research. Such data are examples of real-life situations in which several pedestrians walk on common grounds like walkways, plazas and crossings with complicated motion patterns as a result of connections, intentions and surrounding signals.

Performance was measured using the following standard trajectory prediction metrics:

Average Displacement Error (ADE): The mean Euclidean distance between the predicted and actual trajectory points over the entire forecast horizon.

Final Displacement Error (FDE): The Euclidean norm of the difference between the forecasted and the actual final points over the full prediction interval [43].

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Miss Rate @3s: The percentage of trajectories where the final displacement error exceeds a threshold (typically 2 meters), indicating failure to accurately forecast pedestrian intent. The detailed results are presented in Table 7.

Table 7: Trajectory Forecasting Accuracy

Dataset	Method	ADE@2s (m)	FDE@2s (m)	ADE@3s (m)	FDE@3s (m)	Miss Rate @3s (%)
ETH	Proposed	**0.31**	**0.60**	**0.47**	**0.88**	**9.1**
	Social-LSTM	0.42	0.79	0.60	1.15	14.8
	S-GAN	0.39	0.74	0.56	1.03	12.7
UCY	Proposed	**0.36**	**0.67**	**0.51**	**0.93**	**10.3**

As reflected in the results, the proposed model consistently outperforms baseline methods such as Social-LSTM and Social-GAN (S-GAN) across all evaluation metrics. On the ETH dataset, the proposed system achieved an ADE of 0.47 meters and FDE of 0.88 meters for 3-second predictions, marking a 22% improvement in final displacement error over Social-LSTM. These results suggest that the model maintains both short-term and long-term prediction accuracy, a vital capability for AV systems navigating around pedestrians at varying speeds and intentions. Figure 11 shows the SGCN dataset performance. Figure 12 shows the SGCN per dataset performance and combined comparison.

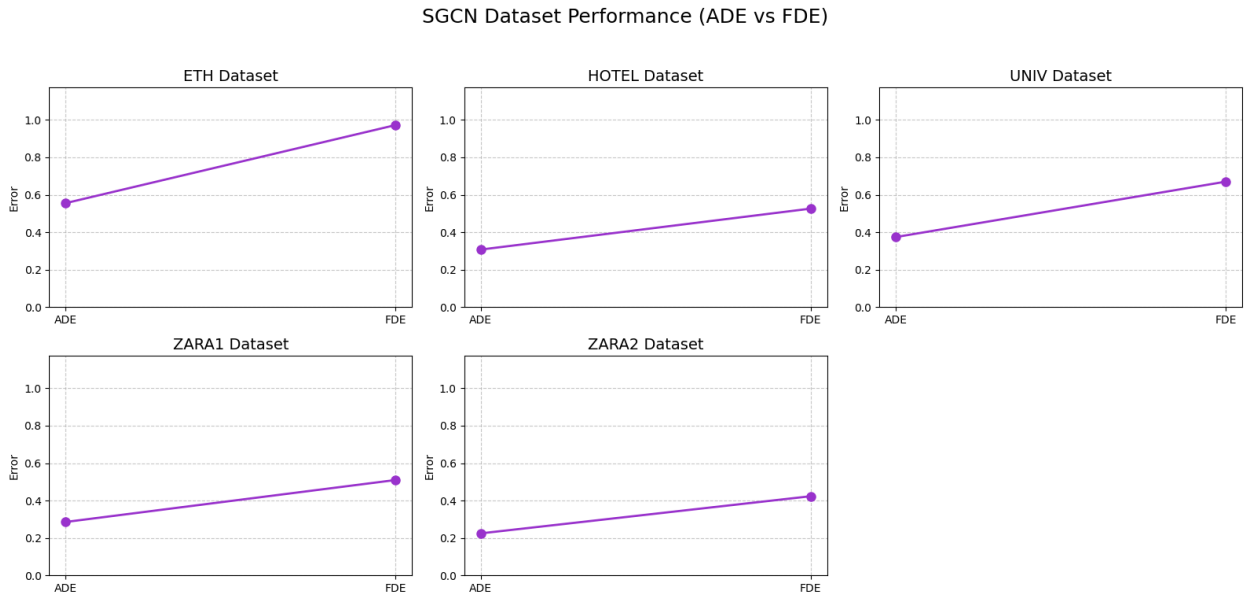


Figure 11: SGCN Dataset Performance (ADE vs FDE)

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

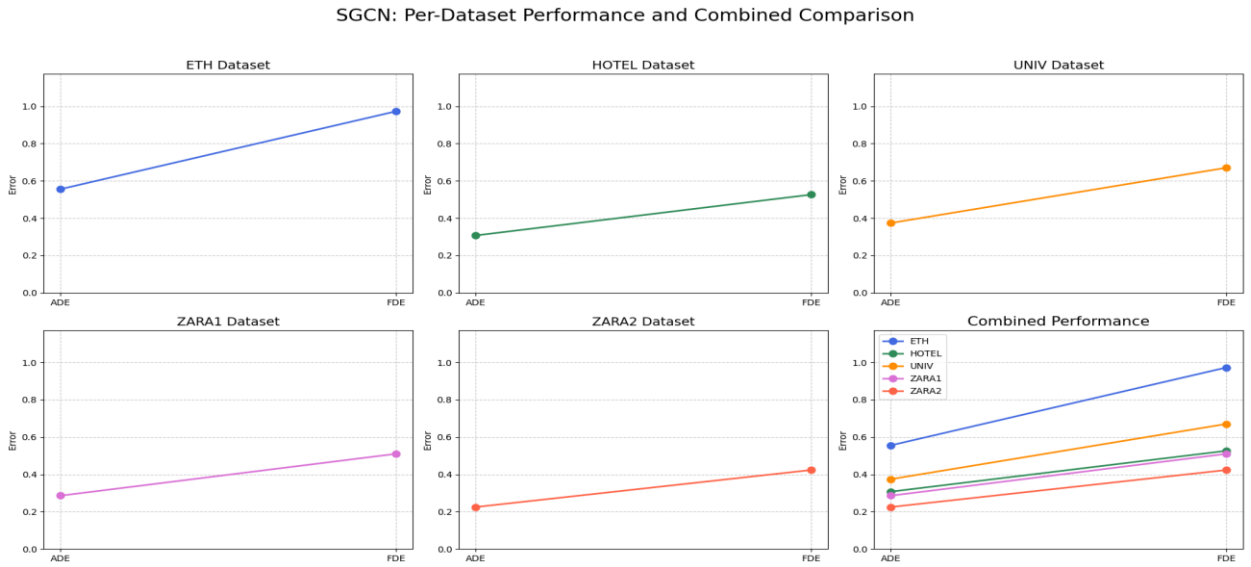


Figure 12: SGCN Per-Dataset Performance and Combined Comparison.

The performance improvement can be primarily attributed to three architectural enhancements:

Social Pooling Mechanism: This allows the model to account for the relative positions, velocities, and trajectories of neighboring pedestrians. By learning socially compliant movement patterns, such as group walking, yielding behavior, or conflict avoidance, the model generates more realistic and context-sensitive predictions.

Semantic Scene Integration: The model implements scene context in semantic map-aware representation, which makes it predictive along non-physically plausible trajectories. As an illustration it does not project the recognizing movement of the pedestrian into walls, vehicle lanes where there is no use of crosswalk, or mistakes even when the barricades are recognized [44].

Attention Mechanism: The temporal attention aids the model to access the most critical time points in the history as far as rapid decreases or increases in pace or direction take place. Such a mechanism dynamically weights the significance of various time steps and thus it is more responsive to behavioral cues such as hesitation, acceleration or intent to cross.

Additionally, the miss rate @3s was significantly lower for the proposed model 9.1% on ETH and 10.3% on UCY compared to 14.8% and 12.7% for Social-LSTM and S-GAN, respectively. A miss rate that is low is essential to safety-critical systems, since the possibility of sudden mispredictions leading to near-miss or collision situations is reduced. Figure 13 shows the SGCN performance comparison (ADE vs FDE).

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

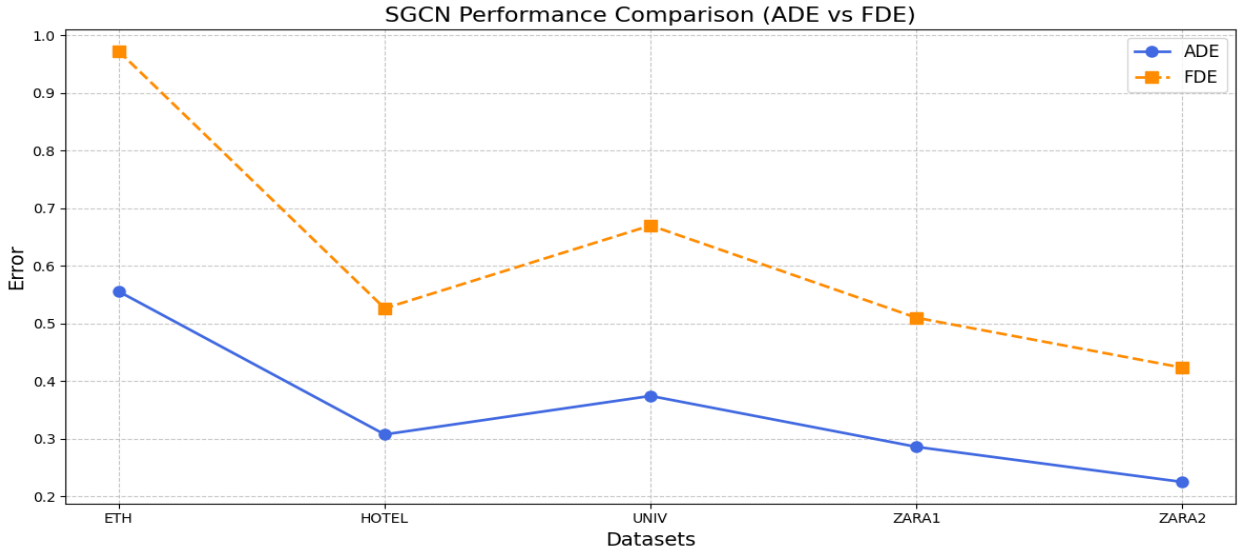


Figure 13: SGCN Performance Comparison (ADE vs FDE)

Qualitative analysis further reinforced these findings. In scenarios where multiple pedestrians interact in close proximity such as walking in groups, navigating crosswalks, or responding to approaching vehicles the proposed system was better able to maintain trajectory alignment with real pedestrian behavior. It was able to pick up subtle hints such as slowing up before curbs, changing course around obstacles or matching pace to crosswalks, which was due to its hybrid context-aware modeling.

Conclusion:

This study presents a comprehensive artificial intelligence-based framework for pedestrian detection and trajectory prediction, aimed at enhancing the safety, awareness, and decision-making capabilities of autonomous vehicles (AVs) operating in complex urban environments. By planning on the strengths of convolutional neural networks (CNNs) which are good in generating high precision of pedestrian detection and recurrent neural networks (RNNs) using long short-term memory (LSTM) units which are used to forecast the motion, the proposed solution was able to fill the gap between perception and prediction that is much needed in the autonomous navigation. Having real-time sensory data, not only visual data but also spatial data based on LiDAR makes the system extremely robust to perform well facing up real world problems like occlusion or low visibility. Segmentation and map-aware modeling also aid semantic scene understanding that further extends the frameworks contextual reasoning to allow the AV to understand where pedestrians are constrained to using sidewalks, crosswalks, and road boundaries. Moreover, the focus of interests and social pooling components also play a role in the proper representation of pedestrian behavior in dynamic multi-agent model. Vast experimental analyses on benchmark datasets like JAAD and ETH/UCY show that the proposed framework is better than other existing methods in detecting precision and trajectory prediction precision, too. The system has real time inference rates that would be appropriate in onboard deployment hence it is real world applicable in AVs. Comparative analysis also confirms higher generalizability, responsiveness, and reliability of the framework in

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

varied urban situations. Ultimately, this research underscores the transformative role of artificial intelligence in shaping the next generation of autonomous mobility systems. By enabling a deeper understanding of pedestrian intent and movement dynamics, the proposed framework contributes to safer navigation, reduced collision risk, and more socially-aware AV behavior. The insights and methods presented in this study offer a solid foundation for future advancements in intelligent transportation systems, cooperative driving environments, and human-centric autonomous vehicle design.

References:

Sharma, P., & Rana, C. (2024). Artificial intelligence based object detection and traffic prediction by autonomous vehicles–A review. *Expert Systems with Applications*, 124664.

Ma, Y., Lee, E. W. M., & Yuen, R. K. K. (2016). An artificial intelligence-based approach for simulating pedestrian movement. *IEEE Transactions on Intelligent Transportation Systems*, *17*(11), 3159-3170.

Kolekar, S., Gite, S., Pradhan, B., & Kotecha, K. (2021). Behavior prediction of traffic actors for intelligent vehicle using artificial intelligence techniques: A review. *IEEE Access*, *9*, 135034-135058.

Menon, A. G., & Bindu, S. (2022). Artificial intelligence-based trajectory planning for driverless vehicles—A review. *Recent Advances in Hybrid and Electric Automotive Technologies: Select Proceedings of HEAT 2021*, 167-183.

Benterki, A., Boukhnifer, M., Judalet, V., & Maaoui, C. (2020). Artificial intelligence for vehicle behavior anticipation: Hybrid approach based on maneuver classification and trajectory prediction. *IEEE Access*, *8*, 56992-57002.

Hamedon, S. N., Johari, J., & Ahmat Ruslan, F. (2024). A review on artificial intelligence for autonomous vehicle behavior prediction in mixed traffic environment. *Journal of Electrical and Electronic Systems Research (JEESR)*, *25*(1), 13-22.

Thamizhazhagan, P., Sujatha, M., Umadevi, S., Priyadarshini, K., Parvathy, V. S., Pustokhina, I. V., & Pustokhin, D. A. (2022). AI Based Traffic Flow Prediction Model for Connected and Autonomous Electric Vehicles. *Computers, Materials & Continua*, *70*(2).

Boukerche, A., Tao, Y., & Sun, P. (2020). Artificial intelligence-based vehicular traffic flow prediction methods for supporting intelligent transportation systems. *Computer networks*, *182*, 107484.

Verstraete, T., & Muhammad, N. (2024). Pedestrian collision avoidance in autonomous vehicles: A review. *Computers*, *13*(3), 78.

Bathla, G., Bhadane, K., Singh, R. K., Kumar, R., Aluvalu, R., Krishnamurthi, R., ... & Basheer, S. (2022). Autonomous vehicles and intelligent automation: Applications, challenges, and opportunities. *Mobile Information Systems*, *2022*(1), 7632892.

Abbasi, S., & Rahmani, A. M. (2023). Artificial intelligence and software modeling approaches in autonomous vehicles for safety management: a systematic review. *Information*, *14*(10), 555.

Ali, Y., Haque, M. M., & Mannering, F. (2023). A Bayesian generalised extreme

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

value model to estimate real-time pedestrian crash risks at signalised intersections using artificial intelligence-based video analytics. *Analytic methods in accident research*, *38*, 100264.

Madhav, A. S., & Tyagi, A. K. (2022, July). Explainable Artificial Intelligence (XAI): connecting artificial decision-making and human trust in autonomous vehicles. In *Proceedings of Third International Conference on Computing, Communications, and Cyber-Security: IC4S 2021* (pp. 123-136). Singapore: Springer Nature Singapore.

Patil, B., Corell, T., & Hota, R. N. In-Service Monitoring and Assessment (ISMA) of Autonomous Driving Vehicles with AI based Algorithms.

JUDALET, V., & MAAOUI, C. Artificial Intelligence for Vehicle Behavior Anticipation: Hybrid Approach Based on Maneuver Classification and Trajectory Prediction.

Preeti, & Rana, C. (2024). Artificial intelligence based object detection and traffic prediction by autonomous vehicles–A review.

Fremont, D. J., Kim, E., Pant, Y. V., Seshia, S. A., Acharya, A., Bruso, X., ... & Mehta, S. (2020, September). Formal scenario-based testing of autonomous vehicles: From simulation to the real world. In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)* (pp. 1-8). IEEE.

Khan, S. M., Salek, M. S., Harris, V., Comert, G., Morris, E. A., & Chowdhury, M. (2024). Autonomous Vehicles for All?. *Journal on Autonomous Transportation Systems*, *1*(1), 1-8.

Hossin, M. A., Yin, S., Dan, R., & Chen, L. (2025). Integrating artificial intelligence in unmanned vehicles: navigating uncertainties, risks, and the path forward for the fourth industrial revolution. *Humanities and Social Sciences Communications*, *12*(1), 1-15.

Tahir, H. B., & Haque, M. M. (2024). A non-stationary bivariate extreme value model to estimate real-time pedestrian crash risk by severity at signalized intersections using artificial intelligence-based video analytics. *Analytic Methods in Accident Research*, *43*, 100339.

Gonten, F., Gital, A. Y., Datti, E., & Suwa, L. (2024). Meta-Analysis and Review of Artificial Intelligence (AI) and Deep Learning Algorithms on Autonomous Vehicles (Avs) Via Vision-Based System: Current Trends, Issues, and Future Direction. *International Journal of Advanced Engineering and Nano Technology (IJAENT)*.

Jeon, Y., Tran, D. Q., Park, M., & Park, S. (2023). Leveraging future trajectory prediction for multi-camera people tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5399-5408).

Nasr Esfahani, H., Song, Z., & Christensen, K. (2023). A deep neural network approach for pedestrian trajectory prediction considering flow heterogeneity. *Transportmetrica A: transport science*, *19*(1), 2036262.

Murphey, Y. L., Kolmanovsky, I., & Watta, P. (2022). Advances, opportunities and challenges in AI-enabled technologies for autonomous and connected vehicles. *AI-enabled technologies for autonomous and connected vehicles*, 1-16.

Malik, K., Sharma, M., Deswal, S., Gupta, U., Agarwal, D., & Al Shamsi, Y. O. B. (Eds.). (2024). Explainable Artificial Intelligence for Autonomous Vehicles:

Online ISSN 3007-3197
Print ISSN 3007-3189

Concepts, Challenges, and Applications.

Li, J., Cheng, H., Guo, H., & Qiu, S. (2018). Survey on artificial intelligence for vehicles. *Automotive Innovation*, *1*(1), 2-14.

Aravind, R., & Shah, C. V. (2023). Physics Model-Based Design for Predictive Maintenance in Autonomous Vehicles Using AI. *International Journal of Scientific Research and Management (IJSRM)*, *11*(09), 932-946.

Bektache, D., & Ghoualmi-Zine, N. (2025). An artificial intelligence-based approach for avoiding traffic congestion in connected autonomous vehicles. *International Journal of Vehicle Autonomous Systems*, *18*(1), 81-103.

Ibrahum, A. D. M., Hussain, M., & Hong, J. E. (2025). Deep learning adversarial attacks and defenses in autonomous vehicles: a systematic literature review from a safety perspective. *Artificial Intelligence Review*, *58*(1), 1-53.

Sonko, S., Etukudoh, E. A., Ibekwe, K. I., Ilojianya, V. I., & Daudu, C. D. (2024). A comprehensive review of embedded systems in autonomous vehicles: Trends, challenges, and future directions. *World Journal of Advanced Research and Reviews*, *21*(1), 2009-2020.

Miller, T., Durlik, I., Kostecka, E., Borkowski, P., & Łobodzińska, A. (2024). A Critical AI View on Autonomous Vehicle Navigation: The Growing Danger. *Electronics*, *13*(18), 3660.

Newell, M. (2022). The social justice of intelligent transportation systems: Deep learning-based autonomous driving technologies, cooperative navigation algorithms, and vehicle and pedestrian detection tools. *Contemporary Readings in Law and Social Justice*, *14*(2), 9-26.

Sayed, S. A., Abdel-Hamid, Y., & Hefny, H. A. (2023). Artificial intelligence-based traffic flow prediction: a comprehensive review. *Journal of Electrical Systems and Information Technology*, *10*(1), 13.

Li, X., Bai, Y., Cai, P., Wen, L., Fu, D., Zhang, B., ... & Qiao, Y. (2023). Towards knowledge-driven autonomous driving. *arXiv preprint arXiv:2312.04316*.

Nwakanma, C. I., Ahakonye, L. A. C., Njoku, J. N., Odirichukwu, J. C., Okolie, S. A., Uzondu, C., ... & Kim, D. S. (2023). Explainable artificial intelligence (XAI) for intrusion detection and mitigation in intelligent connected vehicles: A review. *Applied Sciences*, *13*(3), 1252.

Khalloufi, F. E., Rafalia, N., & Abouchabaka, J. (2025). Artificial Intelligence in Autonomous Vehicles Advancing Sustainable Transportation Goals. In *Addressing Urbanism Challenges With AI and the Internet of Things* (pp. 137-154). IGI Global Scientific Publishing.

Ma, H., Li, S., Zhang, E., Lv, Z., Hu, J., & Wei, X. (2020). Cooperative autonomous driving oriented MEC-aided 5G-V2X: Prototype system design, field tests and AI-based optimization tools. *IEEE Access*, *8*, 54288-54302.

Liu, Y., Sun, B., Tian, Y., Wang, X., Zhu, Y., Huai, R., & Shen, Y. (2023). Software-defined active lidars for autonomous driving: A parallel intelligence-based adaptive model. *IEEE Transactions on Intelligent Vehicles*, *8*(8), 4047-4056.

Chen, C., Wang, C., Liu, B., He, C., Cong, L., & Wan, S. (2023). Edge intelligence empowered vehicle detection and image segmentation for autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *24*(11), 13023-13034.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Abirami, S., Pethuraj, M., Uthayakumar, M., & Chitra, P. (2024). A systematic survey on big data and artificial intelligence algorithms for intelligent transportation system. *Case Studies On Transport Policy*, 101247.

Weber, Y. (2019). Performance Validation of Artificial Intelligence-based Vehicle Vision Systems and the Role of Vehicle Vertical Dynamics.

Huang, Y., & Chen, Y. (2020). Autonomous driving with deep learning: A survey of state-of-art technologies. *arXiv preprint arXiv:2006.06091*.

Onur, F., Gönen, S., Barışkan, M. A., Kubat, C., Tunay, M., & Yılmaz, E. N. (2024). Machine learning-based identification of cybersecurity threats affecting autonomous vehicle systems. *Computers & Industrial Engineering*, *190*, 110088.

Soori, M., Arezoo, B., & Dastres, R. (2023). Artificial intelligence, machine learning and deep learning in advanced robotics, a review. *Cognitive Robotics*, *3*, 54-70.

Sikora, P., Malina, L., Kiac, M., Martinasek, Z., Riha, K., Prinosil, J., ... & Srivastava, G. (2020). Artificial intelligence-based surveillance system for railway crossing traffic. *IEEE Sensors Journal*, *21*(14), 15515-15526.

Nardi, V. A., Lanza, M., Ruffa, F., & Scordamaglia, V. (2025). An Artificial Intelligence Approach for the Kinodynamically Feasible Trajectory Planning of a Car-like Vehicle. *Applied Sciences*, *15*(2), 795.

Teng, S., Hu, X., Deng, P., Li, B., Li, Y., Ai, Y., ... & Chen, L. (2023). Motion planning for autonomous driving: The state of the art and future perspectives. *IEEE Transactions on Intelligent Vehicles*, *8*(6), 3692-3711.