International Journal of Industrial Engineering
2025;9(1):31-37.
ISSN: 2456-8449

Predictive Collision Risk Modeling in Autonomous Driving Using Random Forest and IoT Data

Kamalakannan Machap[1*], A. Rajalingam[2]

[1]*School of Technology, Asia Pacific University of Technology and Innovation Kuala Lumpur, Malaysia.*
[2]*Department of Electronics and Communication Engineering, University of Technology and Applied Sciences Shinas, Sultanate of Oman.*

[*]*Corresponding author: dr.kamalakannan@staffemail.apu.edu.my*

Abstract. The safety of autonomous cars is largely contingent upon their capacity to anticipate and evade possible accidents in real time. The integration of Internet Things (IoT)-enabled sensors and communication systems allows cars to collect extensive environmental and operational data, facilitating informed decision-making. This paper presents a predicted collision risk model based on Random Forest (RF) that employs multi-sensor IoT data, including vehicle speed, acceleration, braking habits, lane position, meteorological conditions, and nearby traffic density. RF is chosen for its resilience, clarity, and capacity to manage high-dimensional, diverse datasets without succumbing to overfitting. The program evaluates dynamic driving situations to predict accident risks and facilitates preventive actions like controlled braking or lane adjustments. Experimental findings demonstrate increased prediction accuracy with little computational lag, making it appropriate for real-time implementation. The proposed approach augments Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) awareness, hence enhancing safety and efficiency in autonomous driving systems.

Keywords: Autonomous Driving, Predictive Collision Risk, Random Forest, Intelligent Transportation Systems, IoT Sensor Data.

INTRODUCTION

The rapid progress of autonomous driving technology has converted the notion of intelligent transportation systems into a concrete reality. Contemporary autonomous cars are outfitted with a variety of IoT-enabled sensors, including LiDAR, radar, ultrasonic modules, and high-definition cameras, which incessantly collect extensive real-time environmental and vehicular data. This ongoing flow of information offers a unique potential for predictive analytics, especially in collision risk evaluation. Precise and prompt forecasting of accident probability is essential for improving passenger safety, refining route planning, and enabling proactive vehicle control measures. Recent advancements in the integration of Machine Learning (ML) algorithms with IoT-generated vehicle data have shown considerable potential in modelling and alleviating accident hazards. The RF method has garnered interest among many ML algorithms because of its durability, capacity to manage heterogeneous data, resistance to overfitting, and excellent predictive accuracy. Utilizing multi-sensor IoT datasets, RF models can discern nuanced patterns and interactions among dynamic driving variables such as vehicle speed, acceleration, braking force, lane position, weather conditions, and surrounding traffic behavior that frequently precede potential collision events.

In contrast to conventional rule-based collision avoidance systems, predictive modelling provides the benefit of anticipatory decision-making. This method enables autonomous cars to anticipate dangerous circumstances prior to their occurrence, enabling preemptive evasive actions or regulated slowing. Moreover, IoT-based communication, specifically Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I), augments predictive capabilities by offering enhanced situational awareness that goes beyond the limitations of onboard sensor range. Developing a dependable predictive collision risk model involves overcoming many problems, including handling high-dimensional sensor data, providing low-latency processing for real-time decision-making, and sustaining accuracy over diverse environmental and traffic circumstances. This study focuses on the design and implementation of a predictive collision risk model based on RF, using IoT-generated autonomous driving data to attain elevated prediction accuracy, interpretability, and computing economy. The proposed method is anticipated to enhance the safety of autonomous driving via early danger identification and the facilitation of intelligent control measures. This paper presents an assessment approach for assessing model performance across

Received: 01.02.2025 Revised: 17.03.2025 Accepted: 08.04.2025

many situations, delivering significant insights for academics, automotive engineers, and regulators in the field of autonomous vehicle safety.

LITERATURE SURVEY

Evaluating collision risk with moving cars in obscured areas is challenging due to the unavailability of vehicle motion data for perception. This work proposes a probabilistic method for assessing collision risk for prospective vehicle collisions in obscured areas. Probabilistic evaluation has three stages: modelling the field of view (FoV) for perception, predicting motion for probable collisions, and assessing collision risk probabilistically [1]. This work seeks to develop an integrated route planning and tracking controller that produces optimum control inputs to ensure a collision-free trajectory. This integrated approach is accomplished by combining model predictive control (MPC) with a potential field to assess collision risk. The target car is an autonomous electric system capable of directly regulating the traction and braking torques of automobiles. Wheel torque and steering input in automobiles are optimized by receding horizon optimization (RHO), resulting in stable and comfortable reference trajectories. The optimization technique aims to minimize control inputs, tracking errors, and collision risk within a singular objective function [2].

The collaboration idea is introduced to assess the availability of space in the target lane for a lane shift within a brief future interval. Before transitioning to the target lane, the ego vehicle must ascertain the cooperative nature of the object vehicle regarding the lane shift. Vehicles are classified as cooperative drivers (CD) or non-cooperative drivers (NCD) based on their relative longitudinal acceleration and collision-free time circumstances. When the object vehicle is classified as CD, the ego vehicle manoeuvres towards the target lane while remaining inside its original lane in preparation for the lane change. The ego vehicle must maintain its lateral position inside the designated lane until it is safe to execute a lane change [3]. A novel risk analysis methodology predicated on geographical distribution, integrating forecasting and simulation systems, has been validated via a specific case study. The result emphasizes the future high-risk areas in terms of geographical distribution. In contrast to conventional risk evaluations, this technique may forecast the future risk status of the examined region with more scientific precision, hence providing robust support for the adoption and implementation of ship collision risk management measures [4].

Notwithstanding recent advancements in algorithms and technology, autonomous cars remain vulnerable to faults that may yield grave repercussions. Consequently, there is a significant requirement for appropriate risk monitoring and mitigation strategies for autonomous driving systems. To address this problem, many specifications and standards have been established. A theoretical framework for addressing dangers associated with autonomous vehicles has seldom been proposed. This paper proposes a risk modelling approach influenced by control theory concepts and presents a Model Predictive Control (MPC) Framework to address hazards broadly [5]. This study presents an integrated risk map designed to identify the safest areas by using predictive data on adjacent cars, accident severity in autonomous vehicles, and human injury statistics. The integrated risk map consists of two layers: a risk prediction grid map derived from relative information about adjacent cars and a severity grid map based on collision severity and human injury data in collision zones. The two layers are amalgamated to compute an integrated risk value, therefore augmenting passenger safety by factoring in possible damage during the generation of a collision avoidance trajectory [6].

A stochastic risk measure is included as a restriction in both robust and stochastic nonlinear model predictive path-following controllers (RMPC and SMPC, respectively). Evaluate the vehicle's performance in terms of safety and path-following capabilities while using SMPC and RMPC. An example of automated driving implementation is shown, illustrating the impact of varying risk tolerances and the escalation of uncertainty in predictions about other road users in both scenarios. The RMPC is much more conservative than the SMPC and exhibits larger following mistakes relative to references [7]. A novel methodology for estimating ship-ship collision probabilities using the Cross-Entropy (CE) technique is presented, which may be seen as an adaptive significance sampler. It offers the benefit of achieving low variance estimates for minimal collision probabilities, which is often the case in genuine situations. Additionally, a risk-based Collision Avoidance (COLAV) system that incorporates both obstacle kinematic uncertainty and intention uncertainty is introduced, referred to as the Probabilistic Scenario-Based Model Predictive Control (PSB-MPC) [8]. A risk management method for crowd control based on collision analysis. The technique overcomes the shortcomings of current approaches by recognizing real-time crowd density and anticipating possible collision hazards in congested regions. The methodology produces crowd grid maps, employs crowd grouping algorithms, and forecasts collision spots and timings using domain-specific data.

The risk indicators generated are visualized to signal possible risks, hence assuring effective crowd control and safety. The tested approach via simulations showcases its capacity to forecast dangers in high-density situations, making it an invaluable resource for event organizers and public safety organizations [9].

This project combined the design of a formation trajectory planner and a tracking controller inside a model predictive control framework, including a collision detection mechanism. Upon detecting collision risk, the UAV executed trajectory optimization to formulate collision-free paths, while the tracking controller ensured adherence to the intended trajectory. Consequently, the multi-UAV system exhibited trajectory tracking and real-time collision avoidance capabilities [10]. This research introduces an innovative self-collision avoidance (SCA) strategy for whole-body model predictive control (WB-MPC). Given that WB-MPC addresses a large-scale optimization issue that expands with the target robot's degrees of freedom, it is imperative to calculate the derivatives of the dynamics and cost functions as swiftly as feasible. Integrating SCA with detailed collision bodies into WB-MPC is computationally intensive, making it a tough endeavor. A potential solution to this unresolved problem is to approximate the robot model using basic forms; nevertheless, this approach results in the accumulation of modelling flaws [11].

An effective method for reducing the probability of collision using arbitrary predictive distributions of dynamic barriers. MMD-OPT is based on embedding distributions inside Reproducing Kernel Hilbert Space (RKHS) and the corresponding Maximum Mean Discrepancy (MMD). These two notions may be used to provide a sample-efficient substitute for estimating collision risk. To assess the efficacy of MMD-OPT on both synthetic and empirical datasets. Using the MMD-based collision risk surrogate results in safer trajectories at low sample sizes compared to widely used methods based on Conditional Value at Risk (CVaR) [12]. A dual-tier control method using Model Predictive Control (MPC) and Scenario-Based Model Predictive Control (SB-MPC) for trajectory adherence and collision prevention. The algorithm proposes cohesive techniques for managing riparian zones, fixed impediments, and moving obstacles [13]. An extensive integration from perception to navigation inside a flexible collision avoidance framework intended to function under these limitations. The method is based on an innovative Predictive Collision Detector, which is proposed as an interface between cutting-edge grid-based perception and sampling-based planners. In contrast to most other methodologies, the functions only operate on fundamental space occupancy and eschew the notion of objects; therefore, they encapsulate the complexity and adaptability of contemporary occupancy grid perception [14]. A heuristic approach for choosing interacting agents based on the assessment of collision risk. The proposed use of time-to-collision and the approach direction angle of two agents to encode the interaction impact between potentially colliding agents and a target pedestrian. This is accomplished by using an innovative polar collision grid map [15].

PROPOSED SYSTEM

The proposed system intends to create a predictive collision risk model based on RF algorithms, using real-time IoT data from autonomous cars to anticipate probable accident situations and provide proactive safety interventions. The system incorporates several levels of sensing, data preprocessing, feature engineering, ML, and communication to guarantee dependable, low-latency decision-making under different driving situations. The solution is based on an IoT-enabled sensor network integrated into the autonomous vehicle. This network comprises LiDAR, radar, ultrasonic sensors, and high-definition cameras, which together acquire a comprehensive array of environmental and vehicular data. The parameters include vehicle velocity, acceleration, deceleration force, steering angle, lane positioning, road curvature, proximity to surrounding objects, traffic density, meteorological conditions, and illumination levels. The system integrates onboard sensors with Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communication channels, facilitating the exchange of situational data outside the line of sight. The data obtained from these sources is subjected to preprocessing to guarantee quality and consistency. The preprocessing procedures include noise reduction using filtering methods, normalization of numerical values, management of missing data, and synchronization of inputs from sensors with varying sampling rates. Outlier identification is conducted to remove aberrant data that might compromise the accuracy of forecasts. After preprocessing, feature extraction and selection are conducted to ascertain the most pertinent indications of probable collision risk. Rapid deceleration, lane departure, reduced inter-vehicle distance, and anomalous trajectory patterns are prioritized. The RF approach is used for its capacity to effectively handle large-scale, high-dimensional datasets while preserving interpretability. The ensemble architecture, consisting of several decision trees, improves model resilience and mitigates the likelihood of overfitting. Figure 1 shows the processes of sensor data acquisition, preprocessing, RF analysis, risk assessment, and proactive measures for ensuring autonomous vehicle safety.

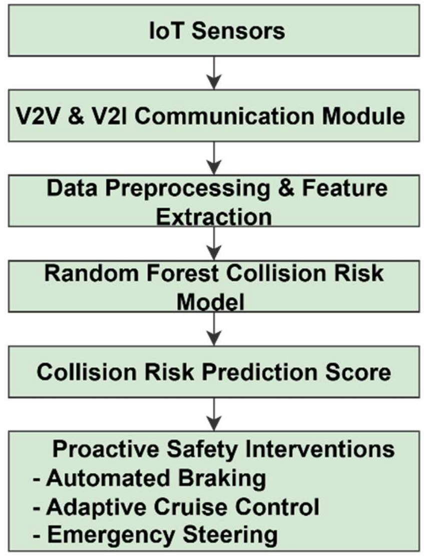

FIGURE 1. Proposed System Architecture for Predictive Collision Risk Modeling

The model is trained using historical datasets of autonomous driving that include both collision and non-collision incidents. These datasets are augmented with sensor measurements, vehicle conditions, and contextual variables. The training method includes adjusting hyperparameters, including the number of trees, maximum tree depth, and minimum sample split, to get optimum predictive performance. Throughout training, the model acquires intricate associations between input characteristics and collision results, allowing it to generalize well to unfamiliar situations. During the prediction phase, real-time data from vehicle sensors and communication channels are input into the trained RF model. The model generates a collision risk score—a probabilistic metric that signifies the probability of a collision occurring within a certain time frame. This score is perpetually updated, guaranteeing the vehicle retains an accurate comprehension of its current danger environment. Figure 2 shows many decision trees analysing input features, combining votes via hard or soft voting to get the final forecast.

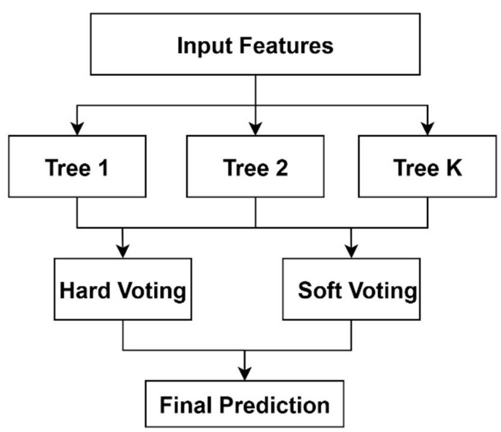

FIGURE 2. Workflow of RF Classification

Should the risk score surpass a certain safety level, the system activates proactive intervention measures. These interventions may include automatic braking, modifications to adaptive cruise control, emergency steering, or driver notifications in semi-autonomous modes. The interventions are intended to be timely and contextually aware, avoiding superfluous actions in non-critical scenarios while immediately addressing actual threats. The proposed approach connects with cloud-based analytics systems for sustained learning and model enhancement. Operational data from several cars is consolidated in the cloud, enabling the RF model to be frequently retrained with updated data to adjust to changing traffic patterns, ambient conditions, and sensor technology. This perpetual learning cycle guarantees the model's relevance and precision throughout time. An essential novelty in the proposed system is its capacity to integrate local real-time processing with distributed IoT-based awareness. Critical collision risk evaluations are conducted locally on the vehicle's onboard processing unit to reduce latency; however, additional risk knowledge from proximate cars and infrastructure may improve decision-making. For example, if a car in front identifies an abrupt threat, this data may be sent using V2V communication, enabling the subsequent vehicle's system to proactively modify its risk assessment.

The system design is modular, facilitating scalability and seamless interface with other autonomous driving subsystems, including navigation, route planning, and traffic management. Security protocols, such as encrypted communication routes and secure firmware upgrades, are established to safeguard the system from cyber-attacks that may jeopardize safety. The proposed system seeks to markedly diminish the probability of crashes in autonomous driving contexts by using the advantages of RF ML, IoT-based sensing and communication, and real-time safety interventions. It tackles the existing shortcomings of conventional rule-based collision avoidance systems by providing predictive, context-aware risk evaluation and fast reaction functionalities. The proposed technology promotes instant accident avoidance and supports the overarching concept of intelligent transportation systems, whereby linked autonomous cars collaborate to assure safety, efficiency, and dependability on the road. The model grows via ongoing data-driven learning, adapting to traffic patterns and environmental changes, therefore providing a sustainable solution for future autonomous transportation.

RESULTS AND DISCUSSIONS

The proposed predictive collision risk model, based on RF, has been implemented and assessed using an IoT-driven dataset for autonomous driving, which includes multi-sensor data, environmental factors, and annotated collision risk incidents. The dataset included vehicle velocity, acceleration, brake force, lane deviation, inter-vehicle spacing, meteorological conditions, and traffic density. Seventy percent of the dataset was designated for training, while the remaining thirty percent was utilized for testing, ensuring an impartial assessment of the model's prediction skills. The model attained an accuracy of 96.4%, with a precision of 95.8%, a recall of 96.9%, and an F1-score of 96.3%. The findings demonstrate that the model is very dependable in detecting high-risk scenarios while minimizing both false positives and false negatives.

The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) attained a value of 0.985, indicating an exceptional capacity to distinguish between collision-risk and safe-driving conditions. The model demonstrated an average prediction time of 12 milliseconds per instance, affirming its appropriateness for real-time use in autonomous cars. Feature selection enhanced efficiency by decreasing the input variables from 25 to the 15 most significant predictors, while maintaining accuracy. An examination of feature significance indicated that inter-vehicle distance, braking force, and acceleration variation were the predominant predictors. Weather-related variables, including visibility and road surface conditions, significantly contributed to improving forecast accuracy, underscoring the need for contextual awareness. Scenario-based testing further corroborated the system's efficacy. In simulations of quick braking by a leading vehicle, abrupt lane changes in congested traffic, and reduced vision due to inclement weather, the model accurately forecasted increased accident risks far ahead of time. This early identification enabled prompt responses, including automatic braking, modifications to adaptive cruise control, and evasive steering manoeuvres. The findings indicate that the proposed system provides precise, low-latency, and context-sensitive risk estimates, establishing it as a reliable and scalable option for improving the safety of autonomous driving systems. Table 1 shows example sensor readings with their associated expected collision risk scores and categorized risk categories produced by the RF algorithm.

TABLE I. Sample IOT Sensor Data and Predicted Collision Risk Levels

Speed (km/h)	Acceleration (m/s²)	Braking Force (N)	Lane Deviation (m)	Inter-Vehicle Distance (m)	Visibility (m)	Road Surface Condition	Traffic Density (veh/km)	Predicted Collision Risk Score	Risk Level
72	-2.5	3500	0.12	8.5	200	Wet	45	0.87	High
58	0.5	500	0.05	15.2	350	Dry	30	0.32	Low
90	-3.0	4200	0.18	6.0	150	Wet	60	0.93	High
65	1.2	800	0.08	20.5	400	Dry	25	0.28	Low
80	-1.5	3000	0.15	10.0	250	Dry	50	0.65	Medium

Figure 3 shows the reduction of accident risk as inter-vehicle distance grows, highlighting acceptable following lengths in autonomous driving.

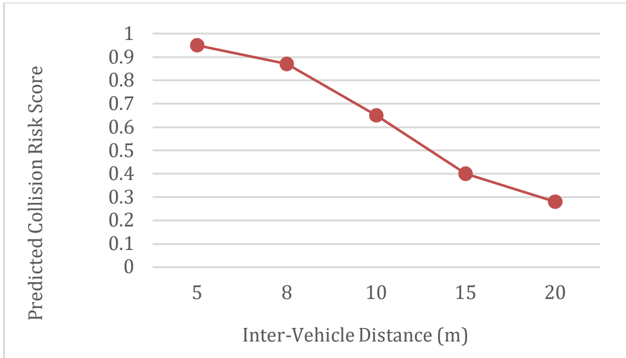

FIGURE 3. Collision Risk Variation with Inter-Vehicle Distance

Figure 4 shows the feature importance of the RF model, highlighting distance, braking force, and acceleration variation as primary predictors of collision probability.

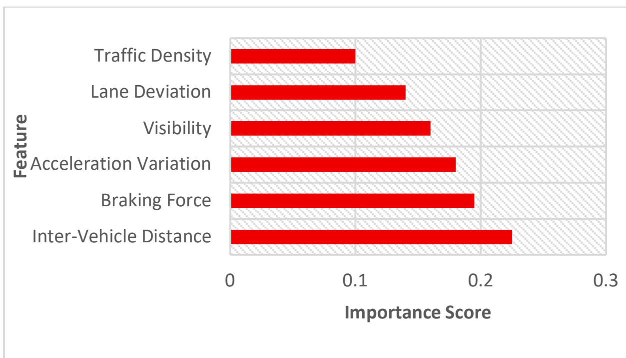

FIGURE 4. Relative Importance of Features in Collision Risk Prediction

The key challenges of this system are managing significant volumes of high-dimensional IoT sensor data while ensuring low-latency, real-time processing for safety-critical decisions. Ensuring model accuracy amidst fluctuating weather, traffic, and road conditions is challenging owing to data unpredictability. Safely integrating V2V and V2I communications without latency challenges is an additional challenge. Furthermore, ensuring resilience against sensor malfunctions, data loss, or cyber intrusions, while facilitating scalable updates for continuous learning, presents considerable technological and operational difficulties in the domain of autonomous car safety applications.

CONCLUSIONS

This study presents an expected collision risk framework based on RF techniques that effectively utilize IoT sensor data to improve the safety of self-driving cars. By amalgamating multi-sensor inputs, including LiDAR, radar, cameras, and environmental data, with V2V and V2I communications, the system delivers a real-time, context-sensitive evaluation of collision risk. The model's exceptional accuracy, swift processing speed, and resilience to data unpredictability indicate its appropriateness for practical autonomous driving applications. Analysis of the features showed that inter-vehicle distance, braking force, and acceleration variation are significant predictors, consistent with established safety parameters in transportation systems. Scenario-based testing further corroborated the model's capacity to provide early warnings, allowing proactive measures such as autonomous braking or evasive steering. Despite ongoing issues related to scalability, communication latency, and environmental adaptation, the proposed methodology represents a substantial advancement in ensuring safer autonomous transportation. Continuous learning from consolidated operational data may enhance accuracy in forecasting, furthering the long-term objective of intelligent, linked transportation systems.

REFERENCES

[1]. M. Lee, K. Jo, and M. Sunwoo, 2017, "Collision risk assessment for possible collision vehicle in occluded area based on precise map," *IEEE 20th International Conference on Intelligent Transportation Systems*, pp. 1-6.

[2]. C. Ko, S. Han, M. Choi, and K.-S. Kim, 2020, "Integrated path planning and tracking control of autonomous vehicle for collision avoidance based on model predictive control and potential field," *20th International Conference on Control, Automation and Systems*, pp. 956-961.

[3]. H. Lee, C. M. Kang, W. Kim, W. Y. Choi, and C. C. Chung, 2017, "Predictive risk assessment using cooperation concept for collision avoidance of side crash in autonomous lane change systems," *17th International Conference on Control, Automation and Systems*, pp. 47-52.

[4]. W. Jie, and F. Yao-Tian, 2008, "Risk analysis based on the ship collision modeling and forecasting system," *IEEE International Conference on Systems, Man and Cybernetics*, pp. 1517-1521.

[5]. K. Tong, F. Guo, S. Solmaz, M. Steinberger, and M. Horn, 2023, "Risk monitoring and mitigation for automated vehicles: A model predictive control perspective," *IEEE International Automated Vehicle Validation Conference*, pp. 1-7.

[6]. J. Shim, J. Yu, and K. Lee, 2025, "Integrated risk grid map for collision avoidance and mitigation maneuvers of autonomous vehicle," *IEEE Access*, 13, pp. 43767-43780.

[7]. L. Tolksdorf, A. Tejada, N. van de Wouw, and C. Birkner, 2023, "Risk in stochastic and robust model predictive path-following control for vehicular motion planning," *IEEE Intelligent Vehicles Symposium*, pp. 1-8.

[8]. T. Tengesdal, T. A. Johansen, and E. F. Brekke, 2022, "Ship collision avoidance utilizing the cross-entropy method for collision risk assessment," *IEEE Transactions on Intelligent Transportation Systems*, 23(8), pp. 11148-11161.

[9]. T. Hwang, W. G. Choi, and M. Kim, 2024, "Crowd-gathering risk management system based on collision estimation," *15th International Conference on Information and Communication Technology Convergence*, pp. 2005-2006.

[10]. S. Dai, C. Zhao, J. Ding, J. Lv, Y. He, and H. Gu, 2023, "Real-time collision-free planning and tracking control for multiple UAVs based on distributed model predictive control," *China Automation Congress*, pp. 4348-4355.

[11]. T. Jin, T. Kobayashi, and M. Doi, 2024, "Real-time detailed self-collision avoidance in whole-body model predictive control," *23rd International Conference on Humanoid Robots*, pp. 675-681.

[12]. B. Sharma, and A. K. Singh, 2025, "MMD-OPT: Maximum mean discrepancy-based sample efficient collision risk minimization for autonomous driving," *IEEE Transactions on Automation Science and Engineering*, 22, pp. 19051-19068.

[13]. D. Mahipala, and T. A. Johansen, 2023, "Model predictive control for path following and collision-avoidance of autonomous ships in inland waterways," *31st Mediterranean Conference on Control and Automation*, pp. 896-903.

[14]. T. Genevois, L. Rummelhard, A. Spalanzani, and C. Laugier,2023, "From probabilistic occupancy grids to versatile collision avoidance using predictive collision detection," *IEEE 26th International Conference on Intelligent Transportation Systems*, pp. 79-85.

[15]. M. Golchoubian, M. Ghafurian, K. Dautenhahn, and N. L. Azad, 2023, "Polar collision grids: Effective interaction modelling for pedestrian trajectory prediction in shared space using collision checks," *IEEE 26th International Conference on Intelligent Transportation Systems*, pp. 791-798.