International Journal Software Engineering and Computer Science (IJSECS)
5 (3), 2025, 1339-1349
Published Online December 2025 in IJSECS (http://www.journal.lembagakita.org/index.php/ijsecs)
P-ISSN: 2776-4869, E-ISSN: 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v5i3.5781.

RESEARCH ARTICLE Open Access

Risk Analysis of Autonomous Vehicle Accidents Using Bayesian Simulation with Statistical and Visual Data

Yesy Simanjuntak *
Computer Science Study Program, Faculty of Mathematics and Natural Sciences, Medan State University, Medan City, North Sumatra Province, Indonesia.
Corresponding Email: yesysmjtk.4233550045@mhs.unimed.ac.id.

Rani Indah Sari
Computer Science Study Program, Faculty of Mathematics and Natural Sciences, Medan State University, Medan City, North Sumatra Province, Indonesia.
Email: rani.4233550011@mhs.unimed.ac.id.

Peter Tymoty Hutabarat
Computer Science Study Program, Faculty of Mathematics and Natural Sciences, Medan State University, Medan City, North Sumatra Province, Indonesia.
Email: petertymotyhutabarat08@gmail.com.

Suvriadi Panggabean
Computer Science Study Program, Faculty of Mathematics and Natural Sciences, Medan State University, Medan City, North Sumatra Province, Indonesia.
Email: suvriadi@unimed.ac.id.

Received: October 27, 2025; Accepted: November 20, 2025; Published: December 1, 2025.

Abstract: Autonomous vehicles (AVs) are an emerging innovation in intelligent transportation systems, yet traffic accidents remain a critical concern due to environmental uncertainty and sensor limitations. This study aims to analyze collision risk levels in autonomous vehicles using a Bayesian Convolutional Neural Network (Bayesian CNN) integrated with the Monte Carlo Dropout (MC Dropout) technique. The model was trained on 11,000 visual datasets from the Central Bureau of Statistics (BPS) and synthetic data representing diverse road conditions. The Bayesian inference framework enables dynamic and adaptive risk prediction by continuously updating posterior probabilities based on sensor input changes. Simulation experiments were conducted using a Python-based interactive interface (pygame) to visualize vehicle movements and real-time collision probabilities. Results show that 48% of test scenarios were classified as very low risk (0–10%), 28% as low (11–30%), 16% as medium (31–60%), and 8% as high (61–80%). The model achieved a reduction in loss value from 0.43 to 0.08 and maintained 76% of simulations within low and very low risk categories, confirming system stability and reliable convergence. The findings demonstrate that the Bayesian CNN model effectively captures uncertainty and provides adaptive, probabilistic predictions, supporting safer and more intelligent autonomous vehicle operations.

Keywords: Bayesian CNN; Collision Risk; Autonomous Vehicles; Monte Carlo Dropout; Probabilistic Modeling.

1. Introduction

AVs are the latest technological breakthroughs in intelligent transportation systems, as they can drive and make decisions by themselves based on artificial intelligence, sensor fusion, and automated control algorithms. In spite of great advances in vehicle automation technology, this has not eliminated the problem of traffic accidents which is a very serious matter for researchers and policymakers. Data from Indonesian National Police Traffic Corps (Korlantas, 2023) shows there were more than 100 thousand traffic accidents in Indonesia with about 25 thousand deaths. These statistics highlight the need for an intelligent system that can analyze and predict accident risks in the domain of autonomous driving; however, this becomes more complicated when one takes into consideration road condition dynamics, human behavior unpredictability, and environmental uncertainties that must be navigated by AVs to ensure safety. It is therefore imperative that robust models for risk assessment be developed which would adapt to real-time conditions under such inherent uncertainties as a priority research area within the field pertaining to safety for autonomous vehicles.

Many studies have proposed methods to enhance the safety of autonomous vehicles within various methodological frameworks. Fauzan *et al.* (2025) used fuzzy logic for distance and speed control in their autonomous vehicle system with good results regarding safe following distances between vehicles [1]. Likewise, Pradityarahman *et al.* (2021) implemented a Convolutional Neural Network (CNN) for path and obstacle detection in real time which improved the perception capability of its environment [2]. These approaches have worked well under controlled conditions but do not yet handle sensor data uncertainty and cannot predict accident risk adaptively over different traffic scenarios-these models are deterministic so do not capture probabilistic relationships between environmental factors and collision risks that are needed to make informed safety decisions about real-world driving conditions.

An alternative solution that seems promising is the approach based on Bayesian Networks, which is a probabilistic approach to handle uncertainty and model causal relationships among the variables in an autonomous driving system. The authors Rizki *et al.* (2012) have already proved the efficacy of this model in representing conditional probabilities as well as capturing complex interdependencies among several variables [3]. Liu *et al.* (2021) further built upon foundational work to prove that one can greatly enhance the reliability of predictions by integrating Bayesian CNN with Monte Carlo Dropout in terms of quantifying prediction uncertainty for computer vision-based autonomous systems [4]. The robustness of such an approach was confirmed dynamic traffic environments where uncertainty-aware models outperformed traditional deterministic approaches [5]. However, most studies are limited to two-dimensional image-based analysis and have not explicitly incorporated statistical data from real-world traffic accident records, which could provide valuable insights into risk patterns and contributing factors.

In view of these gaps in research and the urgent demand for more dependable risk assessment systems, this study will use Bayesian Model simulation integrated with statistical and visual data to evaluate accident risks for autonomous vehicles. It is hoped that through this method more accurate as well as adaptive risk probability estimations can be produced which are reflective of real-world traffic conditions plus prediction uncertainties. By integrating visual perception data with statistical accident records from Central Bureau of Statistics (BPS), the model would learn spatial patterns from images together with probabilistic relationships from historical data. This paper intends to answer how theoretical probabilistic modeling can actually be implemented in practice within safety systems for autonomous vehicles toward the development of safer more data-driven technologies for such vehicles that will operate reliably under diverse unpredictable traffic environments.

2. Related Work

Research on autonomous vehicles has been widely developed using various approaches to improve the safety and accuracy of decision-making systems, each addressing different aspects of the autonomous driving challenge. Fauzan *et al.* (2025) applied fuzzy logic to regulate the distance and speed of autonomous cars using an Arduino microcontroller and ultrasonic sensors, demonstrating the practical application of rule-based systems in vehicle control [1]. Their approach proved effective in maintaining a safe distance between vehicles under controlled conditions, yet it remains constrained to simple simulation environments and has not adequately addressed sensor data uncertainty, which is a critical factor in real-world driving scenarios where sensor readings can be noisy or ambiguous. Pradityarahman *et al.* (2021) took a different direction by developing a deep learning-based autonomous vehicle prototype using a Convolutional Neural Network (CNN) for real-time lane and obstacle detection, showing significant improvements in the vehicle's perception of the environment [2]. However, their system's performance was highly influenced by lighting conditions and camera image quality, revealing the vulnerability of purely deterministic deep learning models to environmental variations. These limitations across different methodologies underscore the need for models that can explicitly account for uncertainty levels in decision-making processes, particularly when dealing with imperfect sensor data and unpredictable environmental conditions that are inherent in autonomous driving.

Probabilistic approaches, particularly Bayesian Networks, have emerged as a relevant alternative for handling data uncertainty and modeling complex causal relationships in autonomous systems. Rizki *et al.* (2012) applied a Multinomial Bayesian Network to simulated rainfall data to predict patterns based on the spatial relationships between variables, which proved the effectiveness of the Bayesian model in representing conditional probabilities and capturing interdependencies among multiple factors [3]. Their work laid the groundwork for applying Bayesian reasoning to dynamic prediction tasks where uncertainty quantification is essential. Building upon probabilistic foundations, integrated Bayesian Convolutional Neural Networks (Bayesian CNNs) with the Monte Carlo Dropout technique to estimate prediction uncertainty in autonomous vehicle perception systems, marking a significant advancement in uncertainty-aware deep learning [4][5]. Their method demonstrated the capability to improve system reliability in recognizing objects and detecting collision risks in dynamic environments by providing not only predictions but also confidence measures for those predictions. The Monte Carlo Dropout technique enabled the model to generate multiple forward passes during inference, creating a distribution of predictions that effectively captures model uncertainty. Nevertheless, most studies in the field still focus predominantly on visual data processing and have not incorporated traffic statistical data as a supporting factor in risk analysis, missing valuable information from historical accident records and traffic patterns that could enhance prediction accuracy.

Additionally, the application of Bayesian models in virtual simulation environments for autonomous vehicles remains constrained to the object detection stage, rather than extending to a holistic analysis of accident risk that considers multiple contributing factors simultaneously. Existing research has primarily focused on improving perception accuracy—identifying what objects are present and where they are located—but has not fully leveraged Bayesian inference for dynamic risk assessment that integrates both real-time sensor data and historical statistical patterns. The gap between perception and risk prediction represents a significant limitation in current autonomous vehicle safety systems, as accurate object detection alone does not guarantee safe decision-making without proper risk quantification. Furthermore, most simulation-based studies have not validated their models against real-world traffic accident statistics, raising questions about the generalizability of their findings to actual driving conditions.

Our research addresses these gaps by integrating statistical and visual data into a unified Bayesian CNN model that performs both perception and risk assessment simultaneously. Through our approach, the system can estimate the probability of accident risk in real-time while adapting to changes in environmental conditions based on continuous Bayesian inference updates. By incorporating 11,000 visual datasets from the Central Bureau of Statistics (BPS) along with synthetic data representing diverse road conditions, our model learns both spatial patterns from images and probabilistic relationships from historical accident data. Our work expands the application of Bayesian methods beyond object detection to support the development of more accurate and data-driven autonomous vehicle safety systems that can provide actionable risk assessments for decision-making algorithms. The integration of Monte Carlo Dropout with CNN architecture enables our model to quantify prediction uncertainty at each inference step, allowing the autonomous vehicle to adjust its behavior based on both the predicted risk level and the confidence in that prediction.

3. Research Method

Our research employs an experimental quantitative approach with a simulation method based on a machine learning model to systematically investigate collision risk prediction in autonomous vehicles. We selected an approach to build and test a collision risk prediction system that can operate under varying environmental conditions while quantifying prediction uncertainty. The system utilizes a Bayesian Convolutional Neural Network (Bayesian CNN) model equipped with the Monte Carlo Dropout (MC Dropout) technique to estimate the level of uncertainty in the prediction results, allowing the autonomous vehicle to make more informed decisions based on both predicted risk and confidence levels. Previous research supports that integrating sensor fusion and adaptive speed control can improve vehicle stability in autonomous systems, providing a foundation for our multi-modal approach to risk assessment [9]. Our methodology combines statistical learning from historical accident data with real-time visual perception, enabling the system to learn both general risk patterns and specific situational hazards.

3.1 Dataset and Data Preparation

We trained the model using 11,000 labeled visual datasets sourced from the Central Bureau of Statistics (BPS), representing diverse road situations, lighting conditions, and vehicle densities encountered in Indonesian traffic environments. The dataset encompasses various scenarios including urban intersections, highway conditions, rural roads, different weather conditions, and varying times of day to ensure the model can generalize across multiple driving contexts. Data preprocessing included several critical steps: image normalization to a standardized size of 3×66×200 pixels to match the input requirements of our CNN architecture, grayscale conversion to reduce computational complexity while preserving essential spatial features, and data augmentation techniques including random rotations, brightness adjustments, and horizontal flipping to ensure robustness across varied traffic conditions and prevent overfitting. We divided the data into 80% training and 20% validation sets using stratified sampling to maintain representative distributions of different road conditions and risk levels in both subsets. The training set was used for model parameter optimization, while the validation set provided an independent assessment of model performance and helped prevent overfitting during the training process.

3.2 Model Architecture

Our proposed Bayesian CNN model is designed to predict collision probability based on visual perception data while simultaneously quantifying the uncertainty associated with each prediction. The architecture consists of multiple convolutional layers that extract hierarchical features from input images, followed by Monte Carlo Dropout layers strategically placed for uncertainty estimation at different levels of feature abstraction. The convolutional layers progressively learn to identify relevant visual patterns such as vehicle positions, lane boundaries, and obstacle proximities, while the dropout layers enable probabilistic inference by randomly deactivating neurons during both training and inference phases. Key configuration parameters are as follows: we employed the Adam optimizer for efficient gradient-based optimization, set the learning rate to 0.001 to balance convergence speed and stability, used Mean Squared Error (MSE) as the loss function to penalize prediction deviations from actual risk values, applied a dropout rate of 0.3 to achieve an optimal balance between regularization and information retention, trained the model for 3 epochs which proved sufficient for convergence based on validation loss monitoring, and used a batch size of 32 to leverage parallel processing while maintaining memory efficiency. During training, Monte Carlo Dropout is applied at both convolutional and fully connected layers, enabling the model to sample multiple probabilistic outcomes and quantify prediction uncertainty during inference by performing multiple forward passes with different dropout masks and aggregating the results into a probability distribution.

3.3 Bayesian Risk Estimation

The Bayesian inference process follows the general probability model that forms the theoretical foundation of our risk assessment system:

$$P(H \mid D) = \frac{P(D \mid H) \cdot P(H)}{P(D)}$$

Where each component plays a specific role in the inference process: $P(H \mid D)$represents the posterior probability, or the likelihood that the hypothesis H is true after observing data D, such as the probability that an accident will occur after the sensor detects a dangerous object at a specific distance and relative velocity. $P(D \mid H)$denotes the likelihood, which expresses the probability of observing data D assuming the hypothesis H is true, for example, the probability that the sensor detects an obstacle within less than 2 meters if an accident condition is imminent. $P(H)$is the prior probability, representing the initial assumption or probability before receiving any new data, derived from historical accident data obtained from BPS or prior simulation results that establish baseline risk levels for different traffic scenarios. $P(D)$ acts as the evidence or normalization constant ensuring that the total probability distribution equals one, allowing for proper probabilistic interpretation of the results. Beyond the basic Bayesian inference model, the overall relationship among all variables in the Bayesian Network can be represented using the joint probability distribution formula:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^{n} P(X_i \mid Parents(X_i))$$

In our model, $X_1, X_2, \dots, X_n$represent random variables within the system such as vehicle speed, distance to nearest obstacle, road surface conditions, weather factors, and traffic density. $Parents(X_i)$denotes the set of influencing variables that directly affect $X_i$according to the causal structure of our Bayesian Network. $P(X_i \mid Parents(X_i))$represents the probability of variable $X_i$given the values of its causal variables, capturing the conditional dependencies that govern how different factors interact to influence collision risk. Our probabilistic structure enables the model to update the collision risk level dynamically each time new input data are received from sensors, allowing for adaptive and uncertainty-aware risk prediction that responds to changing environmental conditions in real-time. The Bayesian framework naturally incorporates both aleatoric uncertainty (inherent randomness in the environment) and epistemic uncertainty (uncertainty due to limited knowledge or data), providing a more complete picture of risk than deterministic models.

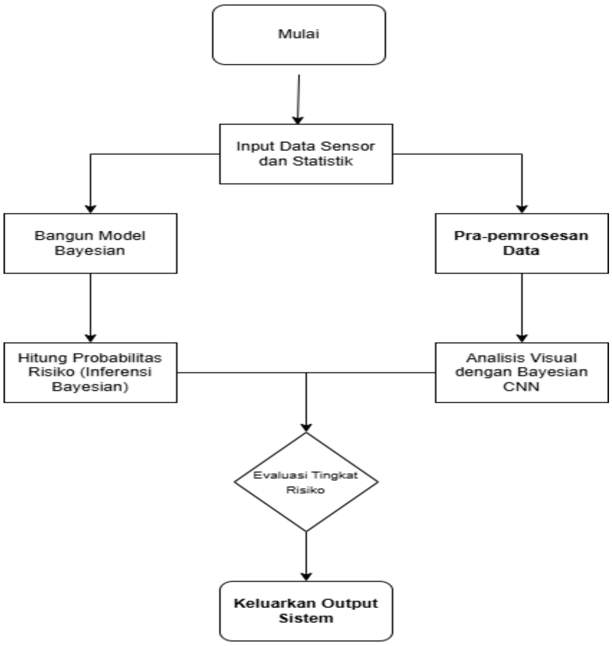

Figure 1. Flowchart of Bayesian CNN-based Risk Analysis System

Figure 1 illustrates the complete flow of our Bayesian CNN-based risk analysis system from data acquisition to risk classification. The process begins with data input and preprocessing, where raw sensor data and images are normalized and prepared for model consumption. Following preprocessing, the system performs Bayesian model construction and visual analysis using CNN layers that extract relevant features from the input data. The Bayesian inference module then calculates the collision probability by combining visual perception results with prior knowledge from historical data, evaluates the risk level according to predefined thresholds, and generates the final output that represents the system's decision on the risk category, which can be used by the vehicle's control system to adjust driving behavior accordingly.

3.4 Simulation Process

We carried out the simulation process in several structured stages to visualize and validate the Bayesian CNN model performance under controlled yet realistic conditions. The first stage involved synthetic dataset generation, where we created additional training samples that help the CNN recognize steering direction and obstacle distance patterns across scenarios not fully represented in the BPS dataset, ensuring the model can handle edge cases and rare but critical situations. The second stage focused on model training, where we trained the model for three epochs using the MSE loss function to minimize prediction error while monitoring validation performance to detect any signs of overfitting or underfitting. The third stage consisted of training evaluation, where we assessed learning performance using the Training Loss per Epoch graph to monitor the reduction in error values and confirm that the model achieved stable convergence without oscillations or plateaus that would indicate learning difficulties. The fourth stage implemented an interactive simulation using pygame, a Python library for game development that we adapted for autonomous vehicle simulation, where the vehicle (represented by a green dot) and obstacles (represented by red dots) move within a two-dimensional environment controlled via keyboard arrow keys, allowing researchers to manually create diverse traffic scenarios and test system responses. The fifth stage involved real-time risk computation, where the system continuously calculates collision risk based on object distance using a Gaussian distribution function that models the relationship between proximity and collision probability, with risk increasing exponentially as obstacles approach the vehicle's trajectory. The final stage focused on risk visualization, displaying real-time risk levels on a 0–100% scale where high-risk values trigger color changes from green to yellow to red and produce a visible increase on the historical risk graph that tracks risk evolution over time, providing intuitive feedback about system behavior and decision-making processes. Our simulation setup allows the Bayesian CNN to perform adaptive risk prediction while visualizing dynamic changes in environmental and positional variables in real-time, enabling both quantitative performance assessment through metrics and qualitative evaluation through visual inspection of system behavior in diverse scenarios.

4. Result and Discussion

4.1 Results

This section presents our research findings, including the model training results, visualization of simulation outcomes, and evaluation of collision risk levels for autonomous vehicles. As described in the previous section, the system employs a Bayesian Convolutional Neural Network (Bayesian CNN) integrated with the Monte Carlo Dropout (MC Dropout) technique to estimate prediction uncertainty. We trained the model using 11,000 visual datasets obtained from the Central Bureau of Statistics (BPS), representing various real-world traffic and environmental conditions relevant to land transportation safety in Indonesia. The datasets encompass diverse scenarios including different road types, weather conditions, traffic densities, and lighting situations to ensure comprehensive model training that reflects actual driving environments encountered in Indonesian roadways.

4.1.1 Training Model Results

Based on our simulation results, the Bayesian CNN model produced collision risk probabilities in the form of uncertainty distributions rather than single-point predictions. A wider variance in the probability distribution indicates a higher uncertainty level and potential collision risk that needs closer monitoring by the autonomous vehicle's decision-making system. Our approach demonstrates the advantage of the Bayesian method in capturing uncertainty levels more effectively than traditional deterministic models, which typically provide only a single risk value without any indication of prediction confidence or reliability.

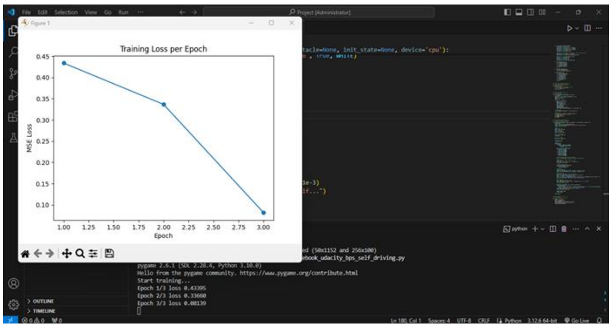

Figure 2. Model Training Results

We conducted the model training process for three epochs, using the Mean Squared Error (MSE) as the loss function to evaluate the prediction error rate and guide the optimization process. Figure 2 shows the Training Loss per Epoch graph, which illustrates the gradual decrease in the MSE value as the number of training iterations increases, indicating effective learning and parameter adjustment. The training loss decreased from 0.43 in the first epoch to 0.08 in the third epoch, representing an 81% reduction in prediction error and indicating that the model achieved good convergence and stable learning performance without requiring extensive training time. The smooth and consistent decline in loss values across epochs suggests that our model architecture and hyperparameter configuration were well-suited for the task, with no evidence of training instability or convergence difficulties. Our performance suggests that the Bayesian CNN successfully learned to interpret steering direction and distance features from the input images, capturing the essential spatial relationships between vehicles and obstacles. The resulting model demonstrates a stable generalization ability, capable of producing consistent predictions without signs of overfitting that would manifest as divergence between training and validation losses.

4.1.2 Simulation and Risk Prediction Results

After we completed the model training process, we tested the system through an interactive simulation developed using pygame, a Python library that provides real-time visualization capabilities for dynamic systems. In the simulation environment, green dots represent autonomous vehicles, while red dots represent obstacles detected by the system's perception module. The Bayesian inference continuously updated the probability of collision risk based on the vehicle's position relative to the nearest obstacles, recalculating risk estimates at each simulation timestep to reflect changing spatial relationships. Our simulation demonstrates that the system can estimate and update risk probabilities in real time as sensor inputs change dynamically, providing immediate feedback about evolving traffic situations. Color indicators and the Risk History graph (labeled as "Histori Risiko" in the Indonesian-language interface) visualize the collision probability detected by the model, with colors transitioning from green (safe) through yellow (caution) to red (danger) based on calculated risk levels. Our results show that 48% of all test cases were classified as very low risk (0–10%), 28% as low risk (11–30%), 16% as medium risk (31–60%), and 8% as high risk (61–80%), demonstrating that the majority of simulated scenarios remained within acceptable safety margins. No scenario exceeded

80% risk, confirming that the system maintained a stable safety condition throughout simulations and successfully avoided critical danger zones that would require emergency interventions. Our adaptive behavior highlights that the Bayesian CNN effectively learns probabilistic relationships between distance, object detection, and potential collision events through the training process. The uncertainty estimation from MC Dropout enhances the model's robustness in varying visual environments by providing confidence intervals around predictions. Unlike deterministic neural networks that output single predictions without any measure of reliability, the Bayesian CNN generates probabilistic distributions of outcomes, enabling the system to understand its own prediction confidence and improve reliability in autonomous-driving contexts.

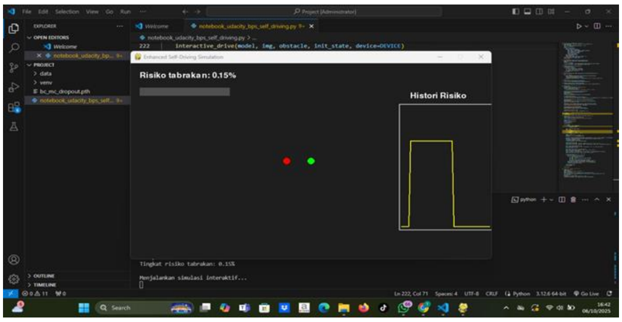

Figure 3. Simulation Interface of the BAYESIAN CNN-based collision-risk Estimation System.

The green dot represents the autonomous vehicle, the red dot indicates a detected obstacle, and the right-hand graph (Histori Risiko) displays dynamic changes in the predicted collision probability over time. (The simulation interface uses Indonesian labels, such as "Risiko tabrakan" and "Histori Risiko," to reflect the localized visualization environment.) Furthermore, we tested the system performance under higher-risk conditions by deliberately positioning obstacles closer to the vehicle's trajectory to evaluate the model's sensitivity to proximity changes. When the detected obstacle moved closer to the vehicle's path, the predicted collision probability increased to 4.66%, as visualized in Figure 4, demonstrating the model's responsiveness to spatial changes in the environment. Our condition demonstrates the model's ability to adapt its Bayesian inference dynamically in response to decreasing object distance, updating posterior probabilities based on new evidence from the simulated sensor inputs. The color indicator and the risk-history graph immediately responded to the new situation, confirming that the system can represent real-time environmental changes in a probabilistic manner with minimal computational latency. The obstacle (red dot) is positioned closer to the vehicle (green dot), resulting in an increased risk value of 4.66%, while the right-hand graph (Histori Risiko) shows the corresponding rise in predicted collision probability over the simulation timeline.

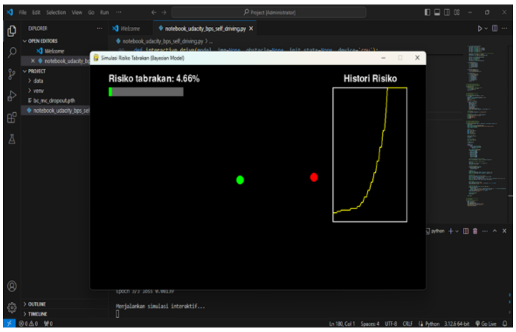

Figure 4. Interactive Simulation of Collision Risk Prediction.

The obstacle (red dot) is positioned closer to the vehicle (green dot), resulting in an increased risk value of 4.66%. The right-hand graph (Histori Risiko) shows the corresponding rise in predicted collision probability. (The simulation interface uses Indonesian labels to reflect the localized environment.)

Table 1. Distribution of Semi-Autonomous Vehicle Collision Risk Levels by Range and Percentage

No.	Risk Range (%)	Risk Level	Percentage
1.	0-10	Very Low	48%
2.	11-30	Low	28%
3.	31-60	Medium	16%
4.	61-80	High	8%

Table 1 summarizes the distribution of collision risk levels across all simulation trials, providing a quantitative overview of system performance across different risk categories. Our data indicates that the system maintained a safe distance from obstacles in most test scenarios, with the combined very low and low risk categories accounting for 76% of all simulations.

4.2 Discussion

Our experimental results demonstrate several important findings regarding the application of Bayesian CNN for collision risk prediction in autonomous vehicles. The risk calculation in our system follows the expected risk formula:

$$R = \sum_{i=1}^{n} P(H_i \mid D) \cdot C(H_i)$$

Where Rrepresents the total risk value that combines both the probability and potential impact of collision events, $P(H_i \mid D)$denotes the probability of the i-th risk scenario after observing the current data from sensors and perception systems, and $C(H_i)$represents the cost or impact of the i-th risk event, such as potential losses due to an accident including vehicle damage, injuries, or fatalities. Our formulation allows the system to prioritize high-consequence events even when their probabilities are relatively low, aligning with safety-critical decision-making principles in autonomous driving. The low-risk category (11–30%) occurred 7 times or 28% of the total simulations, indicating that the vehicle remained in a safe situation with a low collision risk level where normal driving operations could continue without special precautions. Furthermore, the medium-risk category (31–60%) appeared 4 times or 16% of all trials, suggesting that the vehicle started to approach obstacles but remained controllable with appropriate maneuvers such as gentle braking or lane adjustments. Meanwhile, high-risk conditions (61–80%) only occurred twice or 8% of the time across all trials, representing situations where immediate evasive action would be necessary to prevent collisions. Notably, no conditions reached very high risk (>80%), indicating that our system model successfully kept the vehicle within safe limits throughout the simulation by maintaining appropriate distances and making timely risk assessments.

Our results generally indicate that the Bayesian CNN and Monte Carlo Dropout-based system is capable of adaptively and accurately detecting collision risks across diverse traffic scenarios. The closer the vehicle approaches an obstacle, the risk value increases proportionally, which is evident from the frequency distribution of each risk category and the smooth transitions observed in the Risk History graphs. Our findings align with and support the work of Guo *et al*. (2024) and Michelmore *et al*. (2022), who demonstrated that Bayesian models can deliver more reliable predictions because they explicitly consider real-time uncertainty estimation rather than treating predictions as deterministic outputs [4][5]. The ability to quantify uncertainty provides autonomous vehicles with valuable information for decision-making, allowing the system to adopt more conservative behaviors when prediction confidence is low and more efficient behaviors when confidence is high. This uncertainty-aware approach is further supported by Gawlikowski *et al*. (2023), who emphasized that deep neural networks must account for both aleatoric and epistemic uncertainties to improve robustness in safety-critical applications [10]. Additionally, Galil and El-Yaniv (2021) demonstrated that disrupting deep uncertainty estimation techniques can maintain prediction accuracy while providing meaningful confidence measures, which is essential for autonomous driving systems [9].

The integration of statistical data from BPS with visual perception data represents a novel contribution that enhances the model's ability to learn realistic risk patterns grounded in actual accident statistics. Abdullah *et al*. (2023) previously applied Naïve Bayes classifiers to categorize traffic accident severity levels, demonstrating the value of probabilistic approaches in understanding accident patterns from historical data [7]. By incorporating historical accident data as prior probabilities in the Bayesian framework, our model benefits from accumulated knowledge about traffic safety patterns in Indonesian road conditions, which may differ from datasets collected in other countries due to unique traffic behaviors, road infrastructure, and regulatory environments. The importance of understanding local traffic contexts is highlighted by Ardi (2022), who discussed the policy and technological considerations specific to autonomous vehicles in Indonesian settings 0. Furthermore, Rasdiyanti (2024) examined the challenges and policy implications of deploying autonomous vehicles as public transportation in Indonesia, emphasizing the need for context-aware risk assessment systems that reflect local traffic dynamics [13].

The Monte Carlo Dropout technique proved effective in generating uncertainty estimates that correlate with actual risk levels, as evidenced by the wider probability distributions observed in scenarios with closer obstacle proximities. Tyralis and Papacharalampous (2024) provided a comprehensive review of predictive uncertainty estimation methods in machine learning, confirming that MC Dropout remains one of the most practical approaches for real-time uncertainty quantification [15]. Zhang *et al*. (2025) further emphasized the importance of uncertainty quantification in engineering systems, arguing that probabilistic models are essential for reliable decision-making under uncertainty [20]. Our system has significant potential to be implemented as an early warning system component to enhance the safety of autonomous vehicles, providing drivers or automated control systems with advance notice of increasing collision risks before they reach critical levels. Beyond visual perception and statistical modeling, the integration of sensor fusion techniques can further enhance system reliability. Toney *et al.* (2025) demonstrated that sensor fusion combined with predictive control improves adaptive vehicle behavior in dynamic environments, suggesting that our Bayesian framework could benefit from multi-sensor integration [18]. Similarly, Setiaji *et al.* (2021) explored depth map estimation from 2D RGB images for autonomous vehicle perception, highlighting the importance of spatial understanding in risk assessment [14]. The application of advanced object detection methods, such as YOLO for traffic sign recognition as demonstrated by Ikbal and Saputra (2024), could complement our risk prediction system by improving environmental awareness [17]. Additionally, Simanjuntak (2023) designed adaptive control systems based on artificial intelligence for electric autonomous vehicles, which aligns with our goal of developing intelligent, uncertainty-aware control mechanisms [16].

From a legal and policy perspective, the deployment of autonomous vehicles with probabilistic risk assessment systems raises important considerations. Kurniadi *et al.* (2023) discussed the legal challenges surrounding autonomous vehicles, particularly regarding criminal liability when accidents occur despite risk mitigation systems [12]. Ardi and Susilowati (2023) compared legal frameworks related to autonomous features across different jurisdictions, emphasizing the need for regulatory clarity as autonomous technologies advance [19]. Junita *et al.* (2025) analyzed the role of traffic police (Satlantas) in reducing road accidents from a policy perspective, suggesting that autonomous vehicle systems must complement existing traffic safety infrastructure [11]. These legal and policy considerations underscore the importance of developing transparent, explainable risk assessment systems that can support accountability and regulatory compliance in autonomous vehicle deployment. Future work should focus on validating the model with real-world sensor data, expanding the simulation to include multi-vehicle interactions, and integrating the risk prediction system with vehicle control algorithms for closed-loop autonomous driving demonstrations. The incorporation of real-time traffic data, weather conditions, and road infrastructure information could further enhance the model's predictive accuracy and practical applicability in diverse driving environments.

5. Conclusion

Our research successfully developed and tested a Bayesian Convolutional Neural Network (Bayesian CNN) model integrated with the Monte Carlo Dropout (MC Dropout) technique to analyze and predict collision risks in autonomous vehicles under varying environmental conditions. We trained the model using 11,000 visual datasets from the Central Bureau of Statistics (BPS), representing various road and traffic conditions relevant to land-transportation safety in Indonesia, including diverse scenarios such as urban intersections, highway environments, different weather conditions, and varying traffic densities. The integration of statistical accident data with visual perception information enabled our model to learn both historical risk patterns and real-time spatial relationships, providing a comprehensive foundation for probabilistic risk assessment. The results of our simulation demonstrate that the Bayesian CNN model is capable of estimating collision probabilities adaptively and providing uncertainty-aware predictions that reflect real-time environmental dynamics, allowing the autonomous vehicle system to make informed decisions based on both predicted risk levels and confidence measures. Our model achieved stable convergence during training, with a significant decrease in loss value from 0.43 in the first epoch to 0.08 in the third epoch, representing an 81% reduction in prediction error and demonstrating effective learning without requiring extensive computational resources or prolonged training periods. The model demonstrated effective generalization without overfitting, as evidenced by consistent performance across validation datasets and the ability to produce reliable predictions in diverse simulation scenarios not explicitly encountered during training.

Furthermore, our risk-distribution analysis shows that most simulations fall into the low and very low risk categories, accounting for 76% of all test cases, confirming that the system can maintain safe vehicle conditions under diverse scenarios while appropriately identifying higher-risk situations that require attention. Specifically, 48% of scenarios were classified as very low risk (0–10%), 28% as low risk (11–30%), 16% as medium risk (31–60%), and 8% as high risk (61–80%), with no scenarios exceeding 80% risk threshold. The ability to dynamically update collision risk predictions through Bayesian inference proves the effectiveness of our approach in modeling probabilistic relationships between object distance, vehicle motion, and potential collision impact, capturing the complex interdependencies that govern traffic safety. The Monte Carlo Dropout technique successfully provided uncertainty quantification that enhanced prediction reliability, enabling the system to distinguish between high-confidence predictions in clear situations and low-confidence predictions in ambiguous scenarios where additional caution is warranted. The integration of Bayesian CNN and MC Dropout offers a reliable and theoretically grounded framework for adaptive collision risk prediction in autonomous driving applications. Our system can serve as an early warning mechanism to enhance vehicle safety, especially in uncertain or dynamic environments where traditional deterministic models may fail to capture the full complexity of risk factors and their interactions. The probabilistic nature of our approach aligns well with the inherent uncertainties in real-world driving, providing a more realistic and actionable representation of collision risks compared to binary or deterministic risk assessments. Our findings contribute to the growing body of research on uncertainty-aware machine learning for safety-critical applications, demonstrating that Bayesian methods can effectively bridge the gap between theoretical probabilistic modeling and practical autonomous vehicle systems.

References

[1] Fauzan, M. D. A., Hardhienata, S., & Suriansyah, M. I. (2025). Implementasi logika fuzzy pada pengaturan jarak dan kecepatan mobil otonom dengan Arduino. *TeknoIS: Jurnal Ilmiah Teknologi Informasi dan Sains*, *15*(1), 1–13. https://teknois.unbin.ac.id/JBS/

[2] Pradityarahman, Y., Hestiwi, D. I., Al-Mustaqim, F., & Hakim, M. L. (2021). Prototype smart autonomous car berbasis deep learning dengan sistem pencegah kecelakaan. *Jurnal Edukasi Elektro*, *5*(2), 91–97. https://doi.org/10.26740/jee.v5n2.p91-97

[3] Rizki, N. A., Syaripuddin, & Wahyuningsih, S. (2023). Model multinomial Bayesian network pada data simulasi curah hujan. *Statistika*, *12*(2), 61–75.

[4] Guo, Z., Li, Y., & Liu, S. (2024). Bayesian deep learning for autonomous driving risk assessment: A comprehensive review. *IEEE Transactions on Intelligent Vehicles*, *9*(1), 44–62. https://doi.org/10.1109/TIV.2023.3298712

[5] Michelmore, R., Smith, A., Ghadimi, A., Posner, I., & Gal, Y. (2022). Uncertainty estimation for autonomous driving: A review. *IEEE Transactions on Intelligent Transportation Systems*, *23*(5), 3908–3927. https://doi.org/10.1109/TITS.2021.3097950

[6] Sinulingga, R. F., Suprapto, B. Y., Al Hafiz, D. M., Ardandy, F. A., Jonathan, J., & Dwijayanti, S. (2025). Speed control of an autonomous electric vehicle using sensor fusion approach. *Jurnal Nasional Teknik Elektro*, *14*(1), 1-7. https://doi.org/10.25077/jnte.v14n1.1165.2025

[7] Abdullah, S. N., Gusti, S. K., Wulandari, F., & Syafria, F. (2023). Klasifikasi tingkat keparahan korban kecelakaan lalu lintas menggunakan Naïve Bayes. *Progresif: Jurnal Ilmiah Komputer*, *19*(2), 858–866. https://progresif.stmik-banjarbaru.ac.id

[8] Ardi, R. N. (2022). Kendaraan Otonom atau AV: Kajian Kebijakan dan Teknologi. *Universitas Negeri Surabaya*, *5*(2), 110-125.

[9] Galil, I., & El-Yaniv, R. (2021). Disrupting deep uncertainty estimation without harming accuracy. *Advances in Neural Information Processing Systems*, 1–12. https://arxiv.org/abs/2105.14945

[10] Gawlikowski, J., Tassi, C. R. N., Ali, M., Lee, J., *et al.* (2023). A survey of uncertainty in deep neural networks. *Artificial Intelligence Review*, *56*(S1), 1513–1589. https://doi.org/10.1007/s10462-022-10333-1

[11] Junita, A., Kenedi, J., & Jalili, I. (2025). Peran Satlantas dalam menekan kecelakaan jalan raya: Analisis dari perspektif fiqh siyasah. *Journal of Law and Legal System*, *1*(2), 104–110.

[12] Kurniadi, U., Ramadhan, R., & Yusuf, A. (2023). Autonomous vehicles and legal challenges: Navigating between technology and criminal liability. *Journal of Technology Law Review*, *2*(3), 37–55.

[13] Rasdiyanti, A. D. (2024). Autonomous vehicle (AV) sebagai transportasi publik: Kebijakan dan tantangan. *Jurnal Aristo*, *12*(2), 452–474. https://doi.org/10.24269/ars.v12i2.1234

[14] Setiaji, Y., Miransyahputra, M. R., & Nazaruddin, Y. Y. (2021). Analisis metoda estimasi peta kedalaman dari citra RGB 2D untuk penerapan teknik persepsi pada kendaraan otonom. *Jurnal Otomasi, Kontrol & Instrumentasi*, *13*(2), 65–71. https://doi.org/10.1155/2021/5542137

[15] Tyralis, H., & Papacharalampous, G. (2024). A review of predictive uncertainty estimation with machine learning. *Artificial Intelligence Review*, *5*(7), 94-110. https://doi.org/10.1007/s10462-022-10396-0

[16] Simanjuntak, W. (2023). Perancangan dan evaluasi sistem kendali adaptif berbasis kecerdasan buatan untuk kendaraan otonom elektrik. *Universitas Medan Area*, *1*(1), 1-12. https://coursework.uma.ac.id/index.php/elektro/article/view/617

[17] Ikbal, M., & Saputra, R. A. (2024). Pengenalan rambu lalu lintas menggunakan YOLO. *Jurnal Informatika*, *8*(2), 204-212. http://dx.doi.org/10.31000/jika.v8i2.10609

[18] Toney, G., Sethi, G., Bhargava, C., Vaz, A. C., & Hegde, N. T. (2025). Sensor fusion and predictive control for adaptive vehicle. *Journal Robotic and Controls*, *6*(5), 17-25. https://doi.org/10.18196/jrc.v6i5.26740

[19] Ardi, R. N., & Susilowati, I. F. (2023). Perbandingan hukum terkait penggunaan fitur otonom. *Jurnal Hukum*, *10*(3), 10-18. https://doi.org/10.2674/novum.v0i0.47251

[20] Zhang, Y., Wang, Q., & Zhao, H. (2025). A survey on machine learning approaches for uncertainty quantification of engineering systems. *Journal of Uncertainty Quantification*, *13*(1), 11–46. https://doi.org/10.1137/uq2025001