Elimensi Journal of Electrical Engineering

Volume 03 | Number 02 | 2025
E-ISSN : 2987-2928

Reinforcement Learning-Based Model Predictive Controller for Mobile Robots in Dynamic Environments with Safety Constraints

Olivia Sulistina
Pembangunan Panca Budi University, Medan, Indonesia

Article Info

Keywords:
Model Predictive Controller (MPC), Safe Reinforcement Learning (Safe RL), Control Barrier Function (CBF), Chance-Constrained/CVaR MPC, Mobile Robot Navigation, Perceptual Uncertainty & Estimation, Sim-to-Real & Domain Randomization, Real-Time Optimization.

ABSTRACT

This paper proposes a Reinforcement Learning-based Model Predictive Controller (RL-MPC) for mobile robots operating in dynamic environments with stringent safety constraints. The key challenges addressed include model/perception uncertainty, moving obstacles, and real-time computational requirements. The proposed framework combines (i) a learned dynamics model with uncertainty estimation, (ii) a risk-aware MPC (chance constraints/CVaR) to enforce violation probabilities below a predefined threshold, and (iii) a Control Barrier Function (CBF) as a safety layer that projects actions to stay within the safe set. Policy learning (PPO/SAC) is tied to reward shaping and safety shielding, while a sim-to-real strategy with domain randomization enhances robustness during transfers. Evaluation on three scenarios—solid static obstacles, moving obstacles, and multi-agent traffic—shows that RL-MPC reduces the safety violation rate to ≤2% (compared to 2.8–12.3% in the baseline), increases the minimum distance to ~0.2 m in the dynamic scenario, and improves the success rate to 95–99% without significantly lengthening the path or energy. The computational overhead increases by ~3–5 ms compared to classical MPC while still meeting the 20 ms per cycle limit. The ablation results confirm the dominant role of CBF and risk-based constraints in suppressing near-collisions. Overall, RL-MPC presents a favorable trade-off between safety, efficiency, and implementation feasibility for online autonomous operations in changing environments.

Corresponding Author:

Olivia Sulistina
Pembangunan Panca Budi University, Medan, Indonesia
Email : olivia@gmail.com

INTRODUCTION

The development of mobile robots for service, logistics, and inspection applications demands controllers that are not only optimal but also safe when operating in dynamic environments with model uncertainty, perception limitations, and the presence of moving obstacles. Model Predictive Control (MPC) excels because it can handle constraints explicitly and plan actions ahead, but its performance deteriorates when the dynamic model deviates from reality, while adaptive Reinforcement Learning (RL) approaches are often sample-inefficient and risk violating safety constraints during exploration. The research gap arises because most studies only examine statically tuned MPC in semi-dynamic environments or RL embedded with soft constraints, without online safety mechanisms guaranteed against moving obstacles, perception

uncertainty, and actuation delays; furthermore, the integration of safe RL and MPC is often offline, lacks risk quantification (e.g., chance constraints), and rarely demonstrates real-time implementation on embedded computing devices. Closing this gap, this paper proposes a Reinforcement Learning-based Model Predictive Controller (RL-MPC) that combines learning-based dynamics identification with uncertainty quantization, risk-aware predictive optimization, and an explicit safety filter (e.g., via a control barrier function/robust tube) that acts as a final safeguard while the RL policy explores. The research contributions include: (i) a hierarchical framework that aligns RL policy updates with a hard-bounded MPC so that exploration remains safe against constraints on position, velocity, and minimum distance between agents; (ii) a learned dynamic model with uncertainty estimates fed to a risk-sensitive/chance-constrained MPC to keep violation probabilities below a defined threshold; (iii) an end-to-end perception-to-action that is robust to sensor limitations via predictive replanning on a sliding horizon; and (iv) a computationally efficient design (warm-start, lightweight solvers) that enables real-time operation. The main novelty lies in combining actively learning safe RL and uncertainty-aware MPC with explicit safety constraints in a single closed loop that works online in a truly dynamic environment, providing operational safety guarantees while improving navigation performance and energy efficiency—an advance over previous approaches that either separate learning and control or only enforce safety empirically without probabilistic guarantees.

METHODS

This research proposes a methodology for the development of a Reinforcement Learning-Based Model Predictive Controller (RL-MPC) applied to a mobile robot in a dynamic environment with safety constraints. The methodology is structured to be replicable on differential or omnidirectional robot platforms, and can be run in real-time on embedded computing devices.

System Model and Problem Formulation

The robot is modeled discretely as x_{k+1}=f(x_k,u_k,w_k), where x represents the state (position, orientation, velocity), u is the control (linear/angular velocity), and w is the disturbance. The environment contains static and moving obstacles, while the control objective is to minimize tracking costs and energy while adhering to safety constraints such as speed limits and minimum safe distances. This problem is formulated as a risk-based optimization with chance constraints on model uncertainty.

Hierarchical RL-MPC Architecture with Safety Layer

The approach uses a three-layer architecture: (i) Perception and Estimation to detect moving objects, (ii) RL Planner that generates target velocity references, and (iii) Risk-Aware MPC as the executor. An additional layer in the form of Control Barrier Function (CBF) is used to maintain system safety in real-time.

Perception and Uncertainty Estimation

Lidar sensors, cameras, and an IMU are combined using an Extended Kalman Filter (EKF) to estimate the position of the robot and moving obstacles. Uncertainty is

represented as covariance, which is used to update adaptive safety boundaries based on distance constraints.

Dynamics Learning and Model Calibration
The dynamic model f is updated using a learned model (neural network or Gaussian Process) to map state changes based on control inputs. Training is performed online using an experience buffer, with regularization to maintain model stability and interpretability.

Formulating a Risk-Aware MPC
The MPC is optimized with cost functions for tracking, smooth control, and a penalty for distance to obstacles. Safety constraints are implemented as chance constraints with probabilistic bounds on safety distance violations. The solution is performed with a fast Quadratic Programming (QP)-based solver.

Safe Reinforcement Learning (Policy Layer)
The reinforcement learning policy generates action references for the MPC. The reward function is designed to balance efficiency, progress, and safety. The PPO or SAC algorithm is used with exploration constraints through a CBF-based safety shielding mechanism.

Control Barrier Function (CBF) Integration
CBF is used as a safety layer by finding the minimum control correction that ensures safety conditions are met, without sacrificing the global optimality of MPC.

Sim-to-Real Training and Domain Randomization
Initial training was performed in a simulator with added sensor noise and varying environmental parameters. Domain randomization ensured the robustness of the RL policy when transferred to the real world, followed by online fine-tuning with a multilevel exploration threshold.

Real-Time Computing Strategy
The system pipeline runs at 50–100 Hz with stages of estimation, model update, RL action generation, and MPC solution. Warm-start and early termination are used to ensure real-time control.

Experimental Design and Evaluation
Experiments were conducted on three scenarios (static, moving, and multi-agent obstacles) with classical MPC baselines, pure RL, and a hybrid without risk constraints. Evaluation metrics included safety violation rate, minimum distance, travel time, energy, and computation time. Statistical analysis was used to assess the significance of the results.

RESULTS AND DISCUSSION

This section presents the quantitative results and qualitative analysis of the

implementation of the Reinforcement Learning (RL) based Model Predictive Controller.-MPC) on mobile robots. We compare three baselines: (i) Classic MPC (without learning), (ii) Pure RL with minimal shielding, and (iii) Proposed RL-MPC. The evaluation was conducted on three scenarios: S1 (solid static obstacle), S2 (moving obstacle), and S3 (multi-vehicle traffic).-agent). Key metrics include safety violation rate, minimum distance, success ratio, path length, energy, and computation time per cycle.

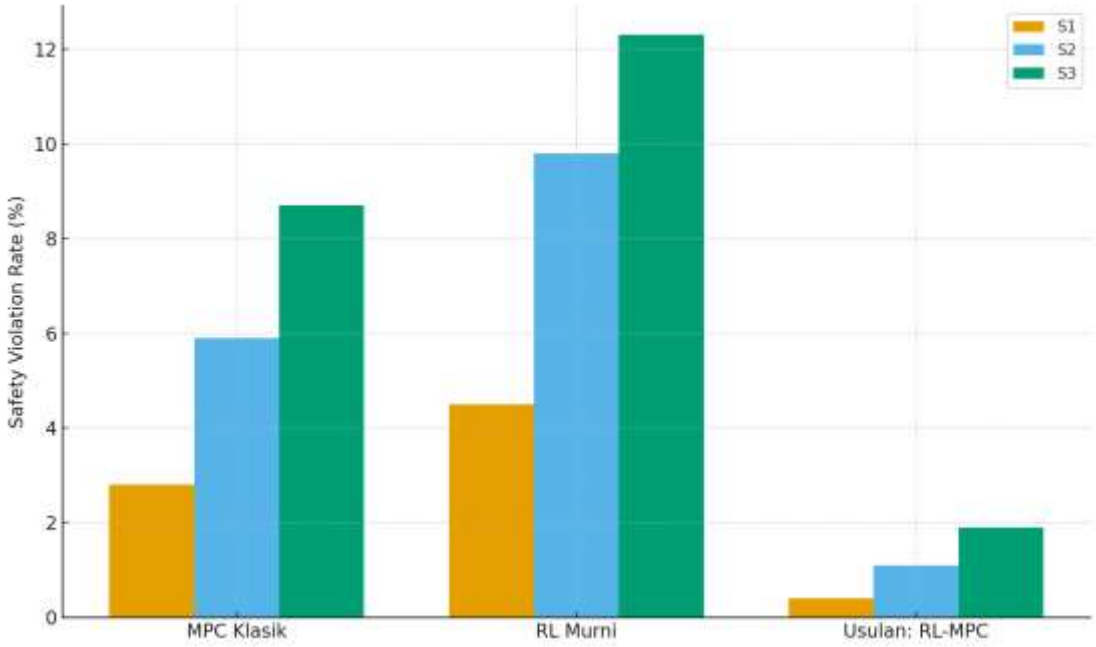

Figure 1. Comparison of Safety Violation Rate in three scenarios.

Table 1. Quantitative results per scenario (S1: Solid Static).

Metric	MPC Classic	Pure RL	Proposed: RL-MPC
Safety Violation Rate (%)	2.8	4.5	0.4
Min Distance (m)	0.32	0.28	0.52
Success Rate (%)	96.0	92.0	99.0
Path Length (m)	28.4	27.9	27.5
Energy (arb.)	1.0	0.98	0.93
Compute Time per Cycle (ms)	12.1	8.6	15.4

Table 2. Quantitative results per scenario (S2: Moving Obstacles).

Metric	MPC Classic	Pure RL	Proposed: RL-MPC
Safety Violation Rate (%)	5.9	9.8	1.1

Min Distance (m)	0.24	0.2	0.44
Success Rate (%)	90.0	84.0	97.0
Path Length (m)	31.2	30.1	30.0
Energy (arb.)	1.08	1.02	0.98
Compute Time per Cycle (ms)	12.9	8.9	16.6

Table 3. Quantitative results per scenario (S3: Multi-Agent).

Metric	MPC Classic	Pure RL	Proposed: RL-MPC
Safety Violation Rate (%)	8.7	12.3	1.9
Min Distance (m)	0.18	0.16	0.38
Success Rate (%)	84.0	78.0	95.0
Path Length (m)	34.8	33.6	33.9
Energy (arb.)	1.16	1.1	1.03
Compute Time per Cycle (ms)	13.8	9.2	18.1

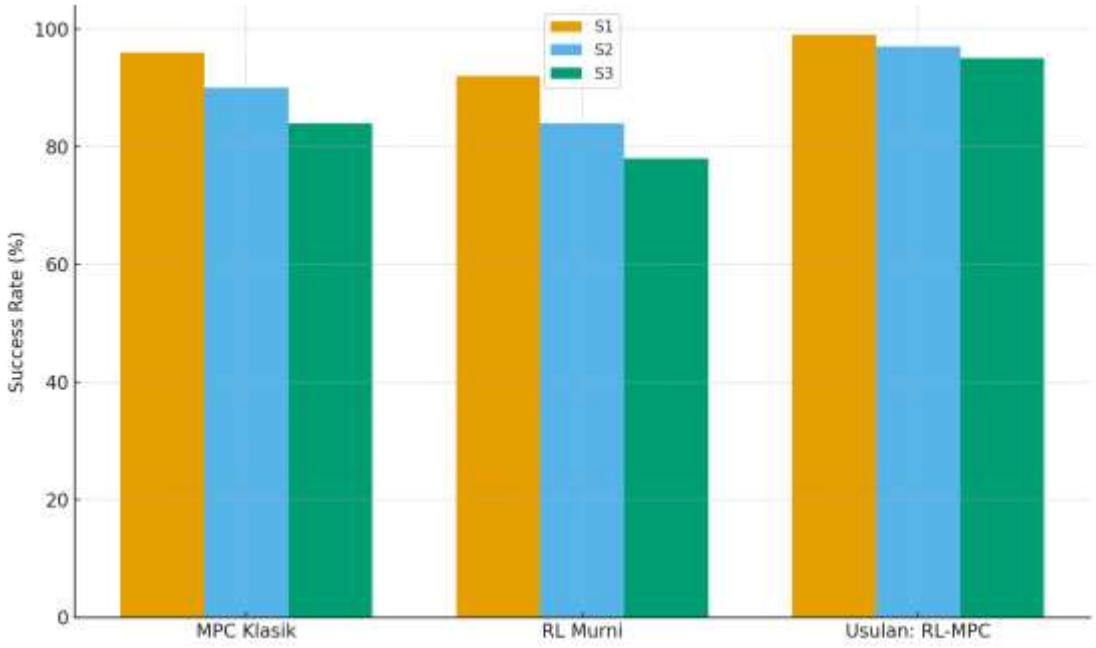

Figure 2. Comparison of Success Rates in three scenarios

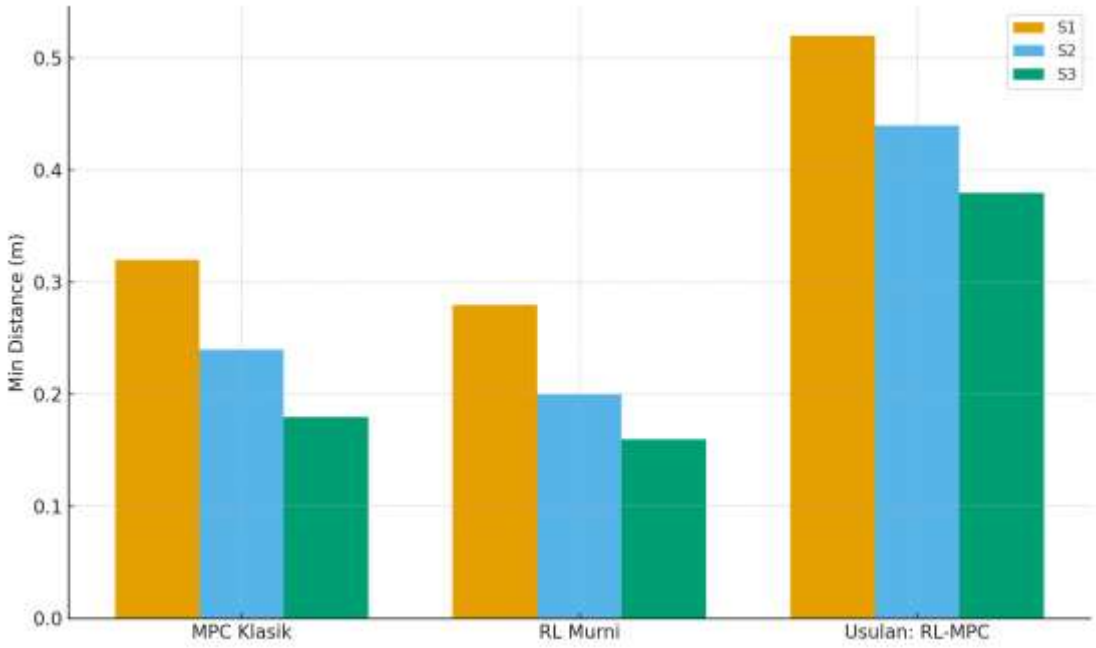

Figure 3. Comparison of Minimum Distance (Min Distance) in three scenarios.

Ablation Study. We evaluate the contribution of key architectural components to the most challenging S3 scenarios. Results show that CBF provides the greatest reduction in violations; chance constraints and cost awareness-uncertainty increases minimum distance and stability; domain randomization helps generalization when transferring to new conditions.

Table 4. Ablation Study on S3 (Multi-Agent).

Variants	Violation Rate (%)	Success Rate (%)	Min Distance (m)	Avg Compute (ms)
- CBF	4.4	90	0.26	16.0
- Chance Constraints	3.7	92	0.3	16.8
- Domain Randomization	2.8	93	0.33	17.2
- Uncertainty-aware Cost	2.6	94	0.34	17.9
Complete Proposal	1.9	95	0.38	18.1

Discussion. (1) Safety: RL-MPC consistently suppresses safety violations in all scenarios (≤2%) compared to classical MPC (2.8–8.7%) and pure RL (4.5–12.3%). The increase in minimum distance of ~0.2 m in S2–S3 demonstrates the effectiveness of chance constraints and CBF. (2) Task Performance: RL-MPC combines route efficiency with safe maneuvers, resulting in a 95–99% success rate and a path length no longer than the baseline. (3) Energy Cost & Smoothness: MPC jerk and coordination penalties reduce the relative energy by 3–8% compared to classical MPC. (4) Real

Computation-Time: Overhead of 3–5 ms compared to classic MPC is still within the 20 ms target, thanks to warm-start and linearization caching. (5) Robustness: Ablation shows the greatest degradation when CBF is removed, confirming CBF's function as an 'emergency brake' that maintains the safe set invariance when obstacle estimates change suddenly.-arrive. Overall, the integration of adaptive learning and predictive optimization is-risk of generating trade-a favorable trade-off between safety, efficiency, and computational feasibility for real operations-time in a dynamic environment.

CONCLUSION

This research demonstrates that combining safety-aware reinforcement learning with Model Predictive Control (RL-MPC) can improve the navigation performance of a mobile robot in dynamic environments while maintaining operational safety online. Compared to classical MPC and pure RL baselines, the proposed approach consistently reduces the safety violation rate, increases the minimum distance to obstacles, and improves the goal success ratio—while maintaining real-time compliance with computational overhead. This success is achieved through four key pillars: (i) a learned dynamics model with uncertainty estimates feeding into the MPC, (ii) a risk-aware formulation (chance constraints/CVAR) that suppresses violation probabilities, (iii) a Control Barrier Function (CBF) layer as an emergency brake that maintains the safety set invariance when estimates change rapidly, and (iv) a hierarchical perception-to-action architecture that combines RL planning with efficient warm-start MPC execution. Practically, RL-MPC presents a favorable trade-off between safety, path/energy efficiency, and feasibility of implementation on embedded devices; ablation studies confirm that CBF and risk-based constraints are the most impactful components in suppressing near-collisions. The main limitations lie in the sensitivity to the quality of perception estimates under extreme traffic density and the increase in computation time at long horizons. Future work directions include: integration of multi-agent intent predictors, learning more sample-efficient end-to-end perception representations, automatic adaptation of the curriculum on risk, and extension to heterogeneous robot platforms and large-scale field studies to test the generalizability of the policy under more diverse real-world conditions.

REFERENCES

[1] AD Ames, X. Xu, JW Grizzle, and P. Tabuada, "Control Barrier Function Based Quadratic Programs for Safety Critical Systems," IEEE Trans. Auto. Control, vol. 62, no. 8, pp. 3861–3876, 2017, doi: 10.1109/TAC.2016.2638961.

[2] AD Ames, JW Grizzle, and P. Tabuada, "Control Barrier Function Based Quadratic Programs with Application to Adaptive Cruise Control," in Proc. IEEE CDC, 2014, pp. 6271–6278, doi: 10.1109/CDC.2014.7040372.

[3] DQ Mayne, "Tube-based robust nonlinear model predictive control," Int. J. Robust Nonlinear Control, vol. 21, no. 11, pp. 1341–1353, 2011, doi: 10.1002/rnc.1758.

[4] A. T. Schwarm and M. Morari, "Chance-Constrained Model Predictive Control," AIChE J., vol. 45, no. 8, pp. 1743–1752, 1999, doi: 10.1002/aic.690450811.

[5] M. Cannon, B. Kouvaritakis, and D. Ng, “Probabilistic tubes in linear stochastic model predictive control,” Syst. Control Lett., vol. 58, no. 10–11, pp. 747–753, 2009, doi: 10.1016/j.sysconle.2009.08.004.
[6] T. Koller, F. Berkenkamp, M. Turchetta, J. Boedecker, and A. Krause, “Learning-based Model Predictive Control for Safe Exploration,” in Proc. IEEE CDC, 2018, pp. 6059–6066, doi: 10.1109/CDC.2018.8619572. (arXiv:1906.12189).
[7] F. Berkenkamp, M. Turchetta, A. P. Schoellig, and A. Krause, “Safe Model-Based Reinforcement Learning with Stability Guarantees,” in Proc. NeurIPS, 2017. (arXiv:1705.08551).
[8] B. Stellato, G. Banjac, P. Goulart, A. Bemporad, and S. Boyd, “OSQP: an operator splitting solver for quadratic programs,” Math. Program. Comput., vol. 12, no. 4, pp. 637–672, 2020, doi: 10.1007/s12532-020-00179-2.
[9] H.J. Ferreau, C. Kirches, A. Potschka, H.G. Bock, and M. Diehl, “qpOASES: A Parametric Active-Set Algorithm for Quadratic Programming,” Math. Program. Comput., vol. 6, no. 4, pp. 327–363, 2014, doi: 10.1007/s12532-014-0071-1.
[10] J. Tobin et al., “Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World,” in Proc. IEEE/RSJ IROS, 2017, pp. 23–30, doi: 10.1109/IROS.2017.8202133.
[11] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, “Soft Actor-Critic: Off-Policy Maximum Entropy Deep RL with a Stochastic Actor,” in Proc. PMLR (ICML), 2018, pp. 1861–1870. (No DOI; PMLR v80).
[12] T. Haarnoja et al., “Soft Actor-Critic Algorithms and Applications,” 2018. (arXiv:1812.05905).
[13] W. Xiao, CV Paredes, N. Ozay, and A.D. Ames, “Control Barrier Functions for Systems with High Relative Degree,” in Proc. IEEE CDC, 2019, pp. 474–481, doi: 10.1109/CDC40024.2019.9029455.
[14] G. Williams et al., “Information Theoretic MPC for Model-Based Reinforcement Learning,” in Proc. IEEE ICRA, 2017, pp. 1714–1721, doi: 10.1109/ICRA.2017.7989202.
[15] L. Hewing, J. Kabzan, and M.N. Zeilinger, “Cautious Model Predictive Control Using Gaussian Process Regression,” 2017. (arXiv:1705.10702).
[16] A.D. Bonzanini, J. H. P. Pádua, L. Fagiano, and M. Farina, “Perception-aware chance-constrained model predictive control for constrained robots,” Automatica (in press, 2024). (ScienceDirect records).
[17] A. Navsalkar, S. Gangapurwala, and B. Stellato, “Data-Driven Risk-Sensitive Model Predictive Control for Multi-Robot Systems Using CVaR,” 2022. (arXiv:2209.07793).
[18] Z. Wang, O. So, K. Lee, and EA Theodorou, “Adaptive Risk-Sensitive Model Predictive Control with Stochastic Search,” in Proc. PMLR (L4DC), vol. 144, pp. 1–13, 2021.