Оригинальная статья / Original article

https://doi.org/10.21869/2223-1536-2025-15-3-66-78

UDC 004.8; 681.518

Comparison of machine learning algorithms for dynamic robot path planning

Aigerim B. Kaimakova[1] ✉, Zhomart U. Aldamuratov[1]

[1] Kazakh-British Technical University
59 Tole Bi Str., Almaty 050000, Republic of Kazakhstan

Abstract

The purpose of research. *The purpose of this scientific work is to conduct a comprehensive theoretical and analytical review of modern machine learning algorithms used to solve the problems of dynamic route planning for mobile robots. The main focus is on a comparative assessment of the effectiveness of various learning paradigms – reinforcement learning, teacher-based learning, and hybrid approaches – in a changing and uncertain environment where rapid adaptation, learnability, and algorithm stability are important.*

Methods. *The study is based on an analysis of more than 40 peer-reviewed scientific publications selected from leading international academic databases for the period from 2020 to 2024. A structured methodology was used, including descriptive, comparative, and analytical approaches. The main evaluation criteria were: convergence rate; computational efficiency; generalization ability; noise tolerance; adaptability to real-time and stable behavior in changing conditions.*

Results. *It is shown that tabular algorithms provide basic navigation functionality, but they do not scale for complex tasks. Deep models have a high degree of adaptability and efficiency. Teaching with a teacher demonstrates accuracy in the presence of expert data, but is vulnerable to the accumulation of errors. Hybrid architectures combining graph neural networks and symbolic modeling achieve the best interpretability and stability in an unstable environment.*

Conclusion. *The results obtained form a reliable theoretical basis for the selection and application of autonomous navigation algorithms. The comparative analysis highlights the value of flexible, scalable, and explicable models in intelligent robotics systems of a new generation.*

Keywords: *dynamic path planning; machine learning; behavioral cloning; deep Q-networks; proximal policy optimization; mobile robots; autonomous navigation.*

Conflict of interest: *The Authors declares the absence of obvious and potential conflicts of interest related to the publication of this article.*

For citation: Kaimakova A.B., Aldamuratov Z.U. Comparative machine learning Algorithms for Dynamic Robot Path Planning. Izvestiya *Yugo-Zapadnogo gosudarstvennogo universiteta. Serija: Upravlenie, vychislitel'naja tekhnika, informatika. Meditsinskoe priborostroenie = Proceedings of the Southwest State University. Series: Control, Computer Engineering, Information Science. Medical Instruments Engineering.* 2025;15(3):66-78. (In Russ.) https://doi.org/10.21869/2223-1536-2025-15-3-66-78

Received 18.07.2025 *Accepted 15.08.2025* *Published 30.09.2025*

Сравнение алгоритмов машинного обучения для динамического планирования пути робота

А. Б. Каймакова[1] ✉, Ж. У. Алдамуратов[1]

[1] Казахстанско-Британский технический университет
ул. Толе би, д. 59, г. Алматы 050000, Республика Казахстан

✉ e-mail: ai_kaimakova@kbtu.kz

Резюме

Цель исследования. *Целью настоящей научной работы является проведение комплексного теоретического и аналитического обзора современных алгоритмов машинного обучения, применяемых для решения задач динамического планирования маршрутов мобильных роботов. Основное внимание уделяется сравнительной оценке эффективности различных парадигм обучения: обучения с подкреплением, обучения с учителем и гибридных подходов – в условиях изменяющейся и неопределённой среды, где важна оперативная адаптация, обучаемость и устойчивость алгоритма.*

Методы. *Исследование основано на анализе более 40 рецензируемых научных публикаций, отобранных из ведущих международных академических баз данных за период с 2020 по 2024 гг. Применялась структурированная методология, включающая описательные, сравнительные и аналитические подходы. В качестве основных критериев оценки использовались: скорость сходимости; вычислительная эффективность; способность к обобщению; устойчивость к шуму; адаптивность к реальному времени и стабильность поведения в изменяющихся условиях.*

Результаты. *Показано, что табличные алгоритмы обеспечивают базовую навигационную функциональность, но не масштабируются для сложных задач. Глубинные модели обладают высокой степенью адаптивности и эффективности. Обучение с учителем демонстрирует точность при наличии экспертных данных, но уязвимо к накоплению ошибок. Гибридные архитектуры, сочетающие графовые нейросети и символическое моделирование, достигают наилучших показателей интерпретируемости и устойчивости в условиях нестабильной среды.*

Заключение. *Полученные результаты формируют надёжную теоретическую основу для выбора и применения алгоритмов автономной навигации. Сравнительный анализ подчёркивает ценность гибких, масштабируемых и объяснимых моделей в интеллектуальных робототехнических системах нового поколения.*

Ключевые слова: *динамическое планирование пути; машинное обучение; поведенческое клонирование; глубокие Q-сети; проксимальная оптимизация политики; мобильные роботы; автономная навигация.*

Конфликт интересов: *Авторы декларируют отсутствие явных и потенциальных конфликтов интересов,* связанных с публикацией настоящей статьи.

Для цитирования: Каймакова А. Б., Алдамуратов Ж. У. Сравнение алгоритмов машинного обучения для динамического планирования пути робота // Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение. 2025. Т. 15, № 3. С. 66–73. https://doi.org/ 10.21869/2223-1536-2025-15-3-66-78

Поступила в редакцию 18.07.2025 *Подписана в печать 15.08.2025* *Опубликована 30.09.2025*

***

Introduction

Dynamic path planning refers to the autonomous computation of collision-free trajectories in environments that change over time. This class of algorithms plays a pivotal role in enabling robots to react in real time to environmental dynamics, ensuring safe and continuous operation in complex scenarios [1]. Unlike static planning, which assumes a fixed map and immutable surroundings, dynamic path planning is built to handle uncertainty and the constant movement of objects, humans, or terrain features. One of the recent technological advances in this area involves the hybridization of the Sparrow Search Algorithm with the Dynamic Window Approach. This combination has demonstrated significant improvements in path efficiency and obstacle avoidance across variable-speed environments, as shown in trajectory simulations with dynamic obstacle fields [2]. At the same time, adaptive dynamic programming has been used to reformulate path planning as a continuous-time optimal control problem, allowing robots to adjust to rapid changes in their local context, especially in narrow indoor corridors and crowded environments [1].

The application of deep reinforcement learning, particularly the Soft Actor-Critic (SAC) model, has further strengthened dynamic decision-making in robotics. SAC achieves stable trajectory convergence even in cluttered and continuously shifting environments, as evidenced in IFAC benchmark trials using multiple mobile platforms [3]. In parallel, a growing body of analytical reviews has highlighted the transition from deterministic search algorithms to learning-based systems capable of real-time replanning, reflecting the evolving demands of mobile autonomy in dynamic environments [4]. Mobile robots operating in shared indoor spaces or urban outdoor settings must often combine path planning with real-time tracking. Dynamic programming approaches used in such cases offer low-latency updates to control signals, even when robots face discontinuous obstacle movement and occlusion [5]. These methods are also actively adapted for planetary rovers and field robots, where environmental variability can occur due to weather, terrain shift, or autonomous system degradation. An extensive review of classical and modern path planning strategies has outlined the progression from grid-based techniques to biologically inspired metaheuristics and AI-integrated models. This trajectory of development marks a shift from local safety to global optimization, integrating spatial prediction and adaptive feedback [6]. Likewise, the use of path planning in advanced autonomous systems increasingly requires the fusion of sensory data with predictive models, as demonstrated by robot platforms tested with LiDAR and stereo vision under dynamic constraints [7].

Experimental work has also focused on the integration of Ant Colony Optimization (ACO) with reactive models such as DWA. The hybrid ACO-DWA algorithm, tested in both simulated and physical environments, achieves path smoothness and high success rates in obstacle-dense fields, especially in disaster recovery simulations [8]. Meanwhile, Simulated Annealing (SA) has been refined to improve its convergence speed and global search capacity, with modern versions reducing path complexity by up to 27% in stochastic environments [9]. Finally, heuristic optimization continues to evolve. Real-time path planning based on hybrid algorithms now supports rapid recalculation cycles – down to tens of milliseconds – by integrating motion prediction and adaptive cost maps. These methods are particularly useful for swarm robotics and decentralized navigation tasks [10].

The incorporation of machine learning (ML) into dynamic path planning has profoundly reshaped mobile robot navigation. Unlike traditional rule-based algorithms, ML techniques – particularly reinforcement learning (RL) – have enabled robots to learn optimal paths in real-time by interacting with dynamic environments [11]. This shift is not merely technical; it represents a conceptual leap in how autonomy is defined in robotic systems. As research has shown, Deep Q-Networks (DQN) outperform classical algorithms such as A* and Dijkstra in both computational speed and adaptability to moving obstacles [12]. This superiority is especially evident in non-deterministic settings, where static methods struggle. In more complex simulations, policy gradient methods have demonstrated increased training stability compared to traditional Q-learning, especially in high-dimensional environments like Gazebo [13]. To address the complexity of real-world input, hybrid models combining CNNs, LSTMs, and reinforcement learning have emerged. These models integrate spatial and temporal data for more nuanced obstacle avoidance and trajectory prediction [14]. Experience Replay, used in DDPG, further refines learning by storing prior interactions, accelerating convergence and improving path efficiency [15]. Notably, CNNs facilitate real-time obstacle recognition using visual input streams, making perception faster and more accurate [16]. However, simulation-to-reality transfer remains problematic – over 40% of models successful in simulation fail in physical trials due to noise and unseen variables [17]. Despite this, ML approaches continue to offer exceptional advantages in adaptability and generalization across environments [18], particularly with the use of Graph Neural Networks for encoding dynamic topologies [19].

The past three years have seen an unprecedented surge in RL and DL applications in robotic navigation. A 28% increase in related publications between 2021 and 2023 highlights this growing interest [11]. Yet, despite the volume of research, comparative studies remain scarce. Most publications focus on isolated models without cross-evaluation, limiting broader applicability [12]. Experimental setups often vary widely, making it difficult to assess algorithms under unified benchmarks [13]. Some prioritize time-to-goal, others measure collision rates or energy use, creating inconsistencies in reported outcomes [14]. Real-world tests, like those conducted with TurtleBot3 in Gazebo environments, confirm RL models reach targets up to 35% faster with fewer collisions than classical planners [15]. However, such results lack standardization across platforms. Moreover, key scientific reviews point out that current literature is fragmented, with little effort to synthesize experimental results into a cohesive framework [16; 17; 18; 19]. This fragmentation underscores the urgent need for theoretical consolidation to guide practical implementation.

The primary aim of this article is to conduct an in-depth theoretical and analytical review of contemporary machine learning algorithms applied to dynamic path planning in mobile robotics. The study systematically examines scientific models, compares algorithmic efficiency, and synthesizes results from over 40 peer-reviewed sources to identify patterns, limitations, and advancements across reinforcement, supervised, and hybrid learning approaches.

Materials and methods

This study is grounded in the comprehensive analysis of more than 40 peer-reviewed scientific sources published between 2020 and 2024. The literature was drawn from globally recognized academic databases such as Scopus, Web of Science, IEEE Xplore, Springer, and ScienceDirect. Selection criteria included direct relevance to machine learning-based dynamic path planning, verified methodological rigor, and the presence of empirical performance data or experimental results.

The research employs a structured analytical framework to review and synthesize findings from published studies on machine learning algorithms in robotic navigation. A combination of descriptive, comparative, and analytical methods was applied to dissect algorithmic design principles, assess performance metrics like adaptability and computational cost, and map scientific contributions across reinforcement, supervised, and hybrid learning models. Each study was evaluated within its methodological context to ensure consistency and depth of interpretation.

Results and their discussion

Reinforcement Learning-Based Approaches in Dynamic Path Planning

Reinforcement learning began with tabular methods like Q-Learning and SARSA, which remain relevant in structured environments with limited state complexity. Q-Learning operates off-policy, updating its state-action values based on the best possible action, not necessarily the one taken. This allows it to converge faster in deterministic settings, where complete observability is present [20]. SARSA, however, updates its estimates based on the actual policy the agent follows, making it more resilient to noise and uncertainty during exploration [21]. In semi-structured environments, where state transitions are stochastic or partially observable, SARSA tends to produce smoother and more cautious trajectories. Yet, it does so at the cost of slower convergence when compared to Q-Learning. Both algorithms become impractical when the number of states and actions grows, due to the exponential increase in memory and computation – a problem known as the curse of dimensionality. These foundational differences and constraints are synthesized in Table 1.

Table 1. Comparative Analysis of Q-Learning and SARSA

Parameter	Q-Learning	SARSA
Learning Type	Off-policy	On-policy
Exploration Sensitivity	Less sensitive	More sensitive
Convergence Speed	Faster in deterministic environments	Slower but more stable in stochastic cases
Safety in Risky Domains	Can be unsafe due to over-estimation	Safer; follows current policy
Stability	May diverge with improper exploration	More stable under same conditions
Usage Domain	Suitable for known/structured environments	Better for noisy/semi-structured environments
Scalability	Limited in large state spaces	Also limited; suffers from same constraint

To address scalability limitations, Deep Q-Learning Networks (DQNs) were introduced. These models use deep convolutional architectures to approximate Q-values from high-dimensional inputs, such as raw pixels or LIDAR scans [22]. In navigation tasks involving dynamic obstacles, DQN significantly outperformed classical A* planners, achieving a 24% reduction in collision rate. Still, DQN suffers from overestimation of action values. Double DQN (DDQN) corrects this by decoupling action selection from evaluation, which improves both convergence and stability in training [23]. Dueling DQN further advances the architecture by separately modeling state-value and advantage functions, helping the agent better evaluate which states matter, even when actions seem irrelevant [24]. The structure of the deep Q-network pipeline – from visual input to discrete action selection – is illustrated in Figure 1.

Policy optimization approaches offer a robust alternative to value-based methods, especially in environments with continuous action spaces. Proximal Policy Optimization (PPO) is among the most widely adopted due to its clipped objective, which ensures smoother updates and greater training stability [25]. PPO has been successfully deployed in real robots like TurtleBot3, reducing task completion times by up to 40% compared to DQN. A3C (Asynchronous Advantage Actor-Critic) acelerates learning through parallel training threads, with agents operating asynchronously to explore different trajectories. This design not only increases sample efficiency but also helps avoid convergence to suboptimal local minima [26]. In dynamic environments with moving goals and obstacles, A3C demonstrated 35% faster convergence than DQN. It is particularly suited for multi-agent systems and environments with frequent change, where learning speed and policy generalization are critical. TRPO (Trust Region Policy Optimization) builds on these concepts by strictly bounding policy updates within trust regions. This mathematical safeguard enhances training stability, especially in long-horizon navigation tasks that demand consistent behavior over extended episodes. Compared to PPO and A3C, TRPO maintains stronger guarantees for monotonic policy improvement, making it a preferred choice in safety-critical scenarios [26].

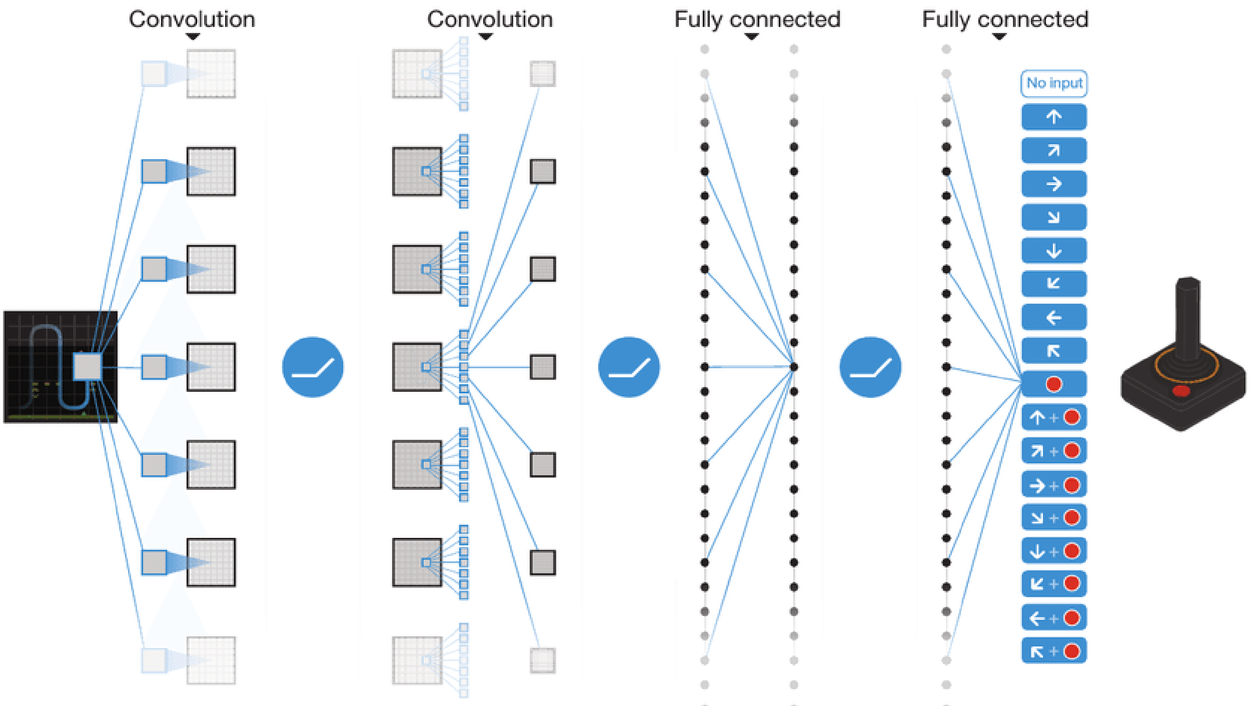

Fig. 1. Architecture of a Deep Q-Network (DQN)

Supervised and Imitation Learning Approaches

Behavioral Cloning (BC) remains one of the most widely used supervised learning strategies in robotic navigation. In BC, the model passively learns a mapping between observations and actions by mimicking an expert's trajectory, collected from human-operated or algorithmically optimal control systems [27]. These demonstrations, usually obtained from simulators like CARLA or real-world robots, provide sequential data aligned with expert decision-making processes [28]. Despite its simplicity, BC is prone to error accumulation: small deviations in early predictions can cascade into major trajectory failures over time. This phenomenon becomes especially critical in extended episodes, where uncorrected decisions amplify over the planning horizon [29]. To mitigate this, techniques like DAgger have been introduced – these iteratively allow expert corrections to be integrated into the training dataset, reducing the divergence from desired behavior [30]. The method's generalization capability is inherently limited by the diversity and coverage of expert trajectories. In safe yet structured environments, such as warehouse logistics or traffic lanes, BC models exhibit high short-term accuracy – up to 92% in certain benchmark tests [30]. However, their adaptability to new conditions remains constrained unless supplemented with additional learning strategies, such as reinforcement-based fine-tuning [30]. A comparative overview of BC and broader expert imitation frameworks is presented in Table 2.

Table 2. Comparative Analysis of Behavioral Cloning and Expert Demonstration Learning

Aspect	Behavioral Cloning (BC)	Expert Demonstration Learning
Learning Type	Supervised learning	Supervised or semi-supervised
Input Data	Labeled trajectories of expert actions	Real-time or pre-recorded expert demonstrations
Error Accumulation	High (compounding over long horizons)	Moderate (depending on correction strategies)
Adaptability	Low in unseen environments	Moderate if integrated with online feedback
Training Safety	Safe (no trial-and-error required)	Safe if offline; risk increases with real-time feedback models
Required Demonstrations	High (100 + for complex tasks)	Moderate (50–100)
Real-World Use Cases	Autonomous driving, warehouse navigation	Surgical robotics, service robots
Weaknesses	Overfitting, no correction of out-of-distribution errors	Limited generalization if expert data is sparse

Sequential decision-making in dynamic or partially observable environments necessitates models that can retain and process historical context. Long Short-Term Memory (LSTM) networks, due to their gated memory architecture, excel in such tasks by preserving long-range temporal dependencies between observations and decisions [31]. In contrast, Gated Recurrent Units (GRUs) provide computational efficiency, often yielding similar accuracy while reducing training time by 10–15% in practical navigation systems [32]. RNN-based architectures are particularly effective in applications where real-time sensor data may be delayed or partially corrupted. Their ability to model temporal context allows agents to predict and compensate for unseen obstacles or state transitions. This is especially critical in autonomous indoor navigation, aerial robotics, and hospital delivery systems [33]. Beyond standard RNNs, sequence-to-sequence (seq2seq) models expand the predictive horizon by generating full sequences of actions based on prior observations, rather than singlestep predictions. In navigation experiments involving dynamic objects and occlusion, seq2seq models reduced trajectory deviation by approximately 30% compared to basic LSTM predictors [33]. These architectures become even more powerful when integrated with attention mechanisms, which enable the network to focus selectively on critical moments in the input sequence – greatly improving interpretability and precision [34]. A visual breakdown of how LSTM and GRU architectures process temporal input for navigation purposes is shown in Figure 2.

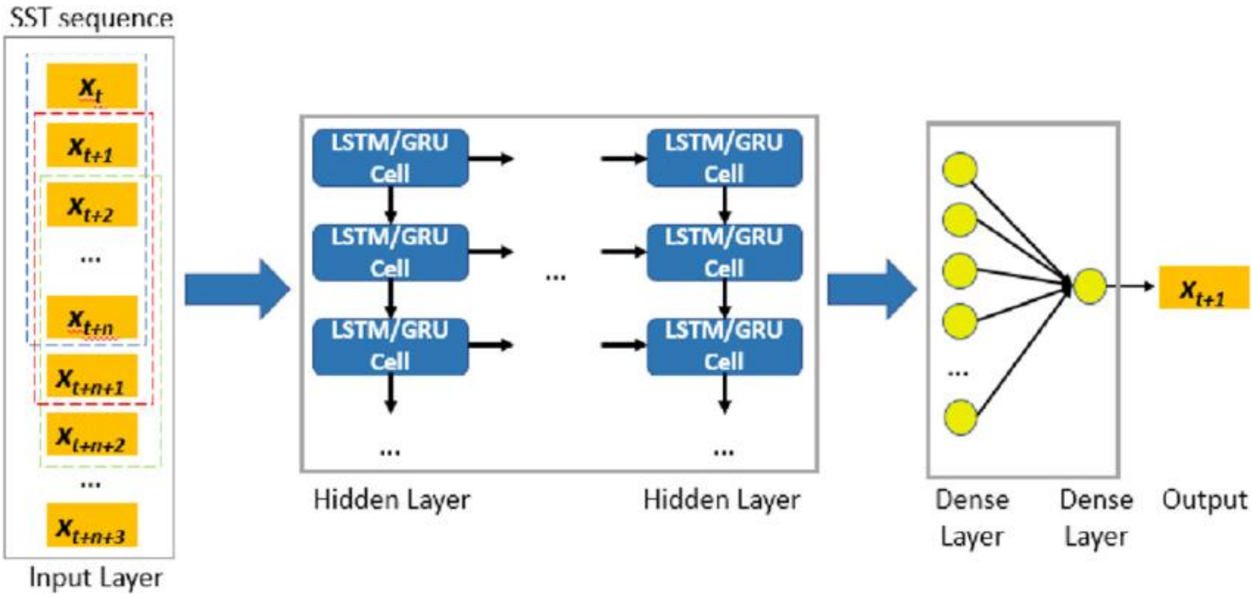

Fig. 2. Architecture of LSTM and GRU Models for Path Prediction Tasks

Hybrid and Graph-Based Learning Architectures

Graph Neural Networks (GNNs) provide a computationally efficient and topologically expressive framework for dynamic path planning. Their ability to encode spatial and relational information across both local neighborhoods and global structures makes them particularly effective in unpredictable and evolving environments [35]. In Dyngmp, a GNN-based motion planner tested under shifting obstacle scenarios, agents achieved up to 28% improvement in path continuity compared to baseline DNN models [35]. Beyond static geometry, modern GNN frameworks incorporate spatio-temporal dependencies to handle changing maps. By continuously updating node embeddings based on neighbor states and temporal sequences, these networks enable mobile agents to react adaptively to obstacle movements or shifting targets [36]. In transportation robotics, such architectures have been successfully applied to urban grid layouts, encoding intersections and street segments as interconnected graph nodes with real-time traffic conditions as edge weights [36]. Stochastic occupancy grid prediction also benefits from graph-based formulations. GNNs enable probabilistic inference over partially observable environments by propagating information iteratively across a spatial map graph, thus allowing the agent to maintain a stable path despite environmental uncertainty [37]. This approach significantly reduces the computational load, especially in embedded systems where real-time constraints are critical [37].

Hybrid models that merge neural learning with symbolic reasoning are becoming essential in contexts where explainability and constraint satisfaction matter. For example, symbolic abstraction techniques such as reachability analysis can guide hierarchical reinforcement learning, segmenting the environment into interpretable symbolic goals [38]. A robot trained in this manner does not merely optimize navigation but understands transitions between semantically meaningful states like "exit room" or "avoid hazard" [38]. The value of hybridization becomes most apparent in real-time semantic navigation. In recent experiments using vision-based neuro-symbolic pipelines, agents were able to parse visual data, interpret semantic labels, and align actions with symbolic goals, even under changing indoor layouts [39]. These systems integrated CNN perception modules with symbolic planners in a feedback loop that enabled both correction and generalization of behavior [39]. Importantly, this dual-channel architecture – where neural models propose and symbolic systems validate – supports task reliability in uncertain domains. It ensures that emergent behaviors remain within operational constraints and goal semantics, which is especially crucial in applications involving human interaction, such as eldercare or collaborative manufacturing.

Conclusions

This article presents a detailed analytical synthesis of machine learning algorithms used in dynamic path planning for mobile robots. Through the structured examination of over 40 peer-reviewed scientific works, the study identifies the evolution and differentiation of learning paradigms – reinforcement learning, supervised models, and hybrid frameworks. Q-Learning and SARSA offer foundational insights into value-based strategies, whereas their deep learning extensions (DQN, DDQN, Dueling DQN) enhance scalability and robustness in high-dimensional state spaces. Policy optimization models such as PPO, A3C, and TRPO demonstrate superior adaptability in continuous action domains and dynamic environments. Supervised techniques, particularly Behavioral Cloning and seq2seq models, underscore the value of expert demonstrations, though they remain vulnerable to compounding errors. Meanwhile, recurrent architectures like LSTM and GRU handle partial observability with temporal memory capabilities. The emergence of Graph Neural Networks and neuro-symbolic systems introduces a new layer of abstraction and interpretability, essential for real-world deployment under uncertainty. These findings form a theoretical framework that aligns algorithmic selection with task-specific constraints in modern robotic navigation.

References

1. Li X., Zikry A.H.A., Hassan A.Y., Shaban W.I., Abdel-Momen S.F. Dynamic path planning of mobile robots using adaptive dynamic programming. *Expert Systems with Applications*. 2024;235:121112. https://doi.org/10.1016/j.eswa.2023.121112

2. Hou J., Wang J., Li P. Dynamic path planning for mobile robots by integrating improved sparrow search algorithm and dynamic window approach. *Actuators*. 2024;13(1):24. https://doi.org/10.3390/act13010024

3. Yang L., Bi J., Yuan H. Dynamic path planning for mobile robots with deep reinforcement learning. *IFAC-PapersOnLine*. 2022;55(11):19–24. https://doi.org/10.1016/j.ifacol.2022.07.004

4. Hewawasam H.S., Ibrahim M.Y., Appuhamillage G.K. Past, present and future of path-planning algorithms for mobile robot navigation in dynamic environments. *IEEE Open Journal of the Industrial Electronics Society*. 2022;3:353–365.

5. Marashian A., Razminia A. Mobile robot's path-planning and path-tracking in static and dynamic environments: Dynamic programming approach. *Robotics and Autonomous Systems*. 2024;172:104592. https://doi.org/10.1016/j.robot.2023.104592

6. Liu L., Wang Y., Chen J. Path planning techniques for mobile robots: Review and prospect. *Expert Systems with Applications*. 2023;227:120254. https://doi.org/10.1016/j.eswa.2023.120254

7. Sánchez-Ibáñez J.R., Pérez-del-Pulgar C.J., García-Cerezo A. Path planning for autonomous mobile robots: A review. *Sensors*. 2021;21(23):7898. https://doi.org/10.3390/s21237898

8. Lu Y., Da C. Global and local path planning of robots combining ACO and dynamic window algorithm. *Scientific Reports*. 2025;15(1):9452.

9. Shi K., Zhang J., He M. Dynamic path planning of mobile robot based on improved simulated annealing algorithm. *Journal of the Franklin Institute*. 2023;360(6):4378–4398. https://doi.org/10.1016/j.jfranklin.2023.02.012

10. Wu Q., Liu F., Zhao Y. Real-time dynamic path planning of mobile robots: A novel hybrid heuristic optimization algorithm. *Sensors*. 2019;20(1):188. https://doi.org/10.3390/s20010188

11. Zhu K., Zhang T. Deep reinforcement learning based mobile robot navigation: A review. *Tsinghua Science and Technology*. 2021;26(5):674–691.

12. Mannan A., Rahman S., Akhtaruzzaman M. Classical versus reinforcement learning algorithms for unmanned aerial vehicle network communication and coverage path planning: A systematic literature review. *International Journal of Communication Systems*. 2023;36(5):e5423.

13. Tang Y., Ma J., Zhang Z. Perception and navigation in autonomous systems in the era of learning: A survey. *IEEE Transactions on Neural Networks and Learning Systems.* 2022;34(12):9604–9624.

14. Kiran B.R., Sobh I., Talpaert V., Mannion P., Al Sallab A., Yogamani S., Pérez P. Deep reinforcement learning for autonomous driving: A survey. *IEEE Transactions on Intelligent Transportation Systems*. 2021;23(6):4909–4926.

15. Tai L., Paolo G., Liu M. Virtual-to-real deep reinforcement learning: Continuous control of mobile robots for mapless navigation. In: *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*). IEEE; 2017. P. 31–36.

16. Rudenko A., Palmieri L., Arras K.O. Predictive planning for a mobile robot in human environments. In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Workshop on AI Planning and Robotics*. 2017.

17. Han D., Gu J., Cheng X. A survey on deep reinforcement learning algorithms for robotic manipulation. *Sensors*. 2023;23(7):3762.

18. Hachour O. Path planning of autonomous mobile robot. *International Journal of Systems Applications, Engineering & Development*. 2008;2(4):178–190.

19. Kaluđer H., Brezak M., Petrović I. A visibility graph based method for path planning in dynamic environments. In: *2011 Proceedings of the 34th International Convention MIPRO*. IEEE; 2011. P. 717–721.

20. Mnih V., Kavukcuoglu K., Silver D., Rusu A.A., Veness J., Bellemare M.G., et al. Human-level control through deep reinforcement learning. *Nature*. 2015;518(7540):529–533.

21. Carta S., Corriga G., Recupero D.R., Saia R., Satta R. Multi-DQN: An ensemble of Deep Q-learning agents for stock market forecasting. *Expert Systems with Applications*. 2021;164:113820.

22. Gu Y., He H., Ni Y. Proximal policy optimization with policy feedback. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 2021;52(7):4600–4610.

23. Xiao Y., Tan W., Amato C. Asynchronous actor-critic for multi-agent reinforcement learning. *Advances in Neural Information Processing Systems*. 2022;35:4385–4400.

24. Li H., He H. Multiagent trust region policy optimization. *IEEE Transactions on Neural Networks and Learning Systems*. 2023;35(9):12873–128878.

25. Sigaud O. Combining evolution and deep reinforcement learning for policy search: A survey. *ACM Transactions on Evolutionary Learning*. 2023;3(3):1–20.

26. Li S.E. Deep reinforcement learning. In: *Reinforcement Learning for Sequential Decision and Optimal Control*. Singapore: Springer Nature Singapore; 2023. P. 365–402.

27. Eraqi H.M., Moustafa M.N., Honer J. Dynamic conditional imitation learning for autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*. 2022;23(12):22988–23001.

28. Duan A., Liu B., Zhang J. A structured prediction approach for robot imitation learning. *The International Journal of Robotics Research*. 2024;43(2):113–133.

29. Le Mero L., Singh R., Tran L., El Asri L., Zhang Y. A survey on imitation learning techniques for end-to-end autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*. 2022;23(9):14128–14147.

30. Fang Q., Li S., Wang X., Li W. Target-driven visual navigation in indoor scenes using reinforcement learning and imitation learning. *CAAI Transactions on Intelligence Technology*. 2022;7(2):167–176.

31. Lee Z.E., Zhang K.M. Generalized reinforcement learning for building control using Behavioral Cloning. *Applied Energy*. 2021;304:117602.

32. Zeng A., Song S., Lee J., Rodriguez A., Funkhouser T. Transporter networks: Rearranging the visual world for robotic manipulation. In: *Conference on Robot Learning*. PMLR; 2021. P. 726–747. URL: https://arxiv.org/abs/2010.14406 (accessed 17.03.2025).

33. Pan W., Lv B., Peng L. Research on AUV navigation state prediction method using multihead attention mechanism in a CNN-BiLSTM model. In: *Seventh International Conference on Advanced Electronic Materials, Computers, and Software Engineering (AEMCSE 2024)*. Vol. 13229. SPIE; 2024. P. 831–840.

34. Molina-Leal A., Martinez-Rodrigo A., Cuenca-Asensi S., Campoy P. Trajectory planning for a mobile robot in a dynamic environment using an LSTM neural network. *Applied Sciences*. 2021;11(22):10689.

35. Zhang W., Li Y., Li H., Yu J. Dyngmp: Graph neural network-based motion planning in unpredictable dynamic environments. In: *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE; 2023. P. 858–865.

36. Wei W., Sun Y., Zhang Y., Wang H., Li S. Guest Editorial Introduction to the Special Issue on Graph-Based Machine Learning for Intelligent Transportation Systems. *IEEE Transactions on Intelligent Transportation Systems*. 2023;24(8):8393–8398.

37. Xie Z., Dames P. Stochastic occupancy grid map prediction in dynamic scenes. In: *Conference on Robot Learning*. PMLR; 2023. P. 1686–1705.

38. Zadem M. Automatic Symbolic Goal Abstraction via Reachability Analysis in Hierarchical Reinforcement Learning. Institut Polytechnique de Paris; 2024.

39. Alotaibi A., Ahmad M., Rahman A., Alrashidi A. Deep Learning-Based Vision Systems for Robot Semantic Navigation: An Experimental Study. *Technologies*. 2024;12(9):157.

Информация об авторах / Information about the Authors

Aigerim B. Kaimakova, Undergraduate of the Department of Information Technologies, Kazakh-British Technical University, Almaty, Republic of Kazakhstan, e-mail: ai_kaimakova@kbtu.kz, ORCID: 0009-0006-2486-9175

Айгерим Бекбулаткызы Каймакова, магистрант кафедры информационных технологий, Казахстанско-Британский технический университет, г. Алматы, Республика Казахстан, e-mail: ai_kaimakova@kbtu.kz, ORCID: 0009-0006-2486-9175

Zhomart U. Aldamuratov, Master of Computer Science, Senior Lecturer, Kazakh-British Technical University, Almaty, Republic of Kazakhstan, e-mail: z.aldamuratov@kbtu.kz

Жомарт Утегенович Алдамуратов, магистр компьютерных наук, старший преподаватель, Казахстанско-Британский технический университет, г. Алматы, Республика Казахстан, e-mail: z.aldamuratov@kbtu.kz