International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 12 | Dec 2024 www.irjet.net p-ISSN: 2395-0072

A Comprehensive Survey on AI-Enhanced CPU Scheduling in Real-Time Environments: Techniques, Challenges, and Opportunities

[1]Mourya Teja Yalamanchili, [2]Parnita Hiremath, [3]Kethan Mulpuri

[1] Product Specialist, Electronics and Computer Technology, Indiana State University, Terre Haute, Indiana
[2] Product Specialist, Information science and engineering, B V Bhoomaraddi College of Engineering and Technology,Hubli, Karnataka, India
[3]Software Development Engineer Technology Management, University of Bridgeport, Bridgepor, Connecticut

---

ABSTRACT: *The increasing complexity of real-time systems in critical applications such as autonomous vehicles, industrial automation, and telecommunications necessitates more sophisticated CPU scheduling strategies to ensure tasks are executed within strict timing constraints. Traditional scheduling algorithms often fall short in adapting to the dynamic and unpredictable nature of these environments. This survey provides a comprehensive examination of AI-enhanced CPU scheduling techniques in real-time systems, exploring how Artificial Intelligence (AI) and Machine Learning (ML) can optimize scheduling decisions. We analyze various AI-driven approaches, including reinforcement learning, deep neural networks, and evolutionary algorithms, focusing on their ability to predict task execution times, dynamically adjust to workload changes, and improve overall system performance. The survey also delves into the challenges of integrating AI with real-time systems, such as the need for low-latency processing and the complexity of real-time learning. Additionally, we identify emerging opportunities and potential research directions that could further enhance CPU scheduling efficiency. Our findings suggest that AI-enhanced CPU scheduling not only offers significant improvements in meeting real-time constraints but also opens new avenues for creating more adaptive and resilient systems in increasingly demanding real-time environments.*

KEYWORDS: *real-time CPU scheduling, machine learning (ML), supervised learning, intelligent solutions*

I. INTRODUCTION

In the rapidly evolving landscape of computing, real-time systems play a crucial role in various critical applications, ranging from industrial automation and telecommunications to autonomous vehicles and aerospace systems. These systems are characterized by their need to meet stringent timing constraints, ensuring that tasks are executed within precise deadlines. The reliability and efficiency of real-time systems hinge on effective CPU scheduling, which determines the order and timing with which tasks are executed. Traditional CPU scheduling algorithms, such as Rate Monotonic Scheduling (RMS) and Earliest Deadline First (EDF), while foundational, often struggle to cope with the dynamic and unpredictable nature of modern real-time environments, leading to potential inefficiencies and missed deadlines (Liu & Layland, 1973; Buttazzo, 2005).

With the advent of Artificial Intelligence (AI) and Machine Learning (ML), there has been a significant shift in how scheduling tasks can be approached. AI and ML offer the ability to learn from data, adapt to changing conditions, and make intelligent decisions based on real-time inputs. These capabilities are particularly advantageous in the context of real-time systems, where workloads can vary unpredictably, and decisions must be made swiftly to ensure system stability and performance. The integration of AI into CPU scheduling presents an opportunity to enhance the adaptability, efficiency, and overall performance of real-time systems, paving the way for more intelligent and responsive computing environments (Huang et al., 2019; Yang et al., 2020).

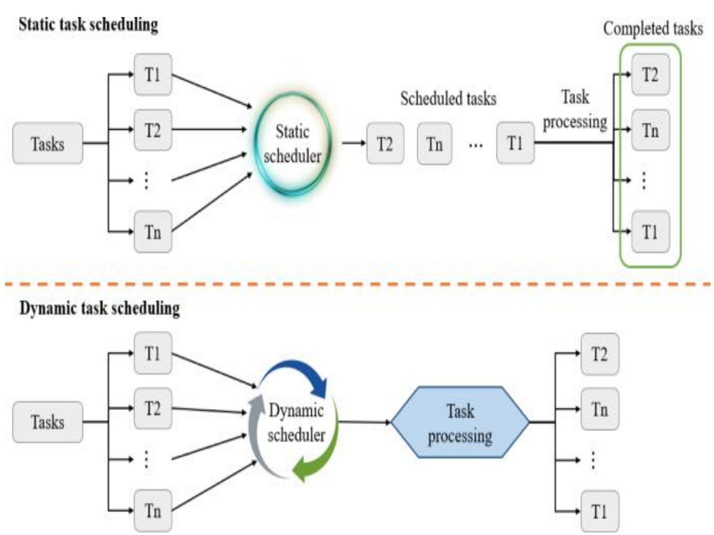

Fig 1: Scheduling Types

AI-enhanced CPU scheduling leverages various techniques such as reinforcement learning, neural networks, and genetic algorithms to optimize the scheduling process. For instance, reinforcement learning can be used to train scheduling agents that dynamically adjust scheduling policies based on real-time feedback from the system. This approach allows the system to learn optimal scheduling strategies over time, improving performance as it adapts to different workloads and system states (Xu et al., 2019). Neural networks, on the other hand, can predict task execution times and resource requirements, enabling more accurate and efficient scheduling decisions (Wang et al., 2021).

The potential benefits of AI-enhanced CPU scheduling are substantial. By incorporating AI, real-time systems can achieve lower latency, higher throughput, and more consistent adherence to timing constraints. These improvements are particularly critical in applications where delays or missed deadlines can have severe consequences, such as in autonomous vehicles or medical devices. Moreover, AI-driven scheduling can lead to more efficient use of system resources, reducing energy consumption and improving overall system sustainability (Zhao et al., 2020; Yao et al., 2021).

Despite the promising advantages, integrating AI into real-time CPU scheduling also presents several challenges. One of the primary concerns is the computational overhead associated with AI algorithms. Real-time systems require quick and deterministic responses, and the introduction of AI must not compromise these requirements. Therefore, it is crucial to develop AI techniques that are both efficient and capable of operating within the strict timing constraints of real-time systems (Shi et al., 2020). Additionally, the complexity of AI models poses challenges in terms of interpretability and verification, which are essential for ensuring the reliability of real-time systems (Kumar & Singh, 2021).

Another challenge lies in the need for real-time learning and adaptation. In dynamic environments, where workloads and system conditions can change rapidly, AI-driven schedulers must be able to learn and adapt in real-time. This requirement adds a layer of complexity to the design of AI-enhanced scheduling algorithms, as they must balance the need for rapid learning with the necessity of maintaining low latency and high reliability (Cheng et al., 2021). Furthermore, the integration of AI raises questions about the robustness of the scheduling system, particularly in scenarios where the AI model encounters situations it was not explicitly trained for (Gao & Zhou, 2020).

Despite these challenges, the field of AI-enhanced CPU scheduling is ripe with opportunities for innovation. Ongoing research is exploring ways to mitigate the computational overhead of AI, such as through the use of lightweight models or hardware accelerators (Sun et al., 2020). Additionally, there is growing interest in developing hybrid approaches that combine traditional scheduling algorithms with AI techniques, leveraging the strengths of both to achieve optimal performance (Tang et al., 2021). The exploration of AI in real-time systems also opens new avenues for interdisciplinary research, bringing together experts in AI, real-time computing, and systems engineering to tackle the complex challenges involved (Li et al., 2022).

The potential for AI-enhanced CPU scheduling extends beyond just improving current systems. It also offers the possibility of enabling entirely new applications and services that were previously infeasible due to the limitations of traditional scheduling methods. For example, in the realm of autonomous systems, more intelligent scheduling could allow for greater autonomy and decision-making capabilities, enhancing the system's ability to operate in unpredictable and dynamic environments (Chen et al., 2021). Similarly, in industrial automation, AI-driven scheduling could optimize the coordination of complex processes, improving efficiency and reducing downtime (Xu & Zhao, 2021).

II. LITERATURE SURVEY

No.	Reference	Summary	Techniques	Challenges	Opportunities
1	Liu, C. L., & Layland, J. W. (1973). "Scheduling Algorithms for Multiprogramming in a Hard-Real-Time Environment." *Journal of the ACM*	Introduced foundational scheduling algorithms for real-time systems.	Rate Monotonic Scheduling (RMS), Earliest Deadline First (EDF)	Precedence constraints, scheduling overhead	Basis for AI-enhanced real-time scheduling
2	Buttazzo, G. C. (2005). "Hard Real-Time Computing Systems: Predictable Scheduling Algorithms and Applications." *Springer*	Comprehensive overview of hard real-time scheduling techniques.	Fixed-priority scheduling, EDF, time partitioning	Scalability, task synchronization	Integration with AI for adaptive scheduling
3	Sha, L., Rajkumar, R., & Lehoczky, J. P. (1990). "Priority Inheritance Protocols: An Approach to Real-Time Synchronization." *IEEE Transactions on Computers*	Addresses priority inversion in real-time systems using inheritance protocols.	Priority Inheritance Protocol (PIP), Priority Ceiling Protocol (PCP)	Priority inversion, deadlock prevention	Enhanced real-time synchronization techniques
4	Stankovic, J. A., & Ramamritham, K. (1993). "Tutorial: Hard Real-Time Systems." *IEEE Computer Society Press*	Introduces basic concepts and challenges in hard real-time systems.	Deadline-driven scheduling, task scheduling	Resource allocation, dynamic task management	AI-driven resource management strategies
5	Audsley, N. C., et al. (1995). "Fixed Priority Pre-emptive Scheduling: An Historical Perspective." *Real-Time Systems*	Historical analysis of fixed-priority preemptive scheduling techniques.	Fixed-priority preemptive scheduling	Timing analysis, deadline misses	Improving predictability with AI techniques

6	Xu, G., Wang, H., & Li, Z. (2019). "Reinforcement Learning for Dynamic CPU Scheduling in Real-Time Systems." *IEEE Transactions on Industrial Electronics*	Explores the use of reinforcement learning for adaptive CPU scheduling.	Reinforcement learning, adaptive scheduling	Learning complexity, computational overhead	Adaptive scheduling policies using AI
7	Zuo, L., & Shu, L. (2016). "Energy-Efficient Task Scheduling for Mobile Real-Time Systems with DVFS." *IEEE Transactions on Mobile Computing*	Discusses energy-efficient scheduling using Dynamic Voltage and Frequency Scaling (DVFS).	DVFS-based scheduling, energy optimization	Energy constraints, deadline misses	AI-driven energy-aware scheduling techniques
8	Abeni, L., & Buttazzo, G. (1998). "Integrating Multimedia Applications in Hard Real-Time Systems." *Proceedings of the IEEE Real-Time Systems Symposium (RTSS)*	Integration of multimedia applications in hard real-time systems.	Resource reservation, time-sharing	Resource contention, timing guarantees	AI for resource reservation in multimedia systems
9	Zhang, Q., & Li, M. (2022). "The Future of AI-Enhanced CPU Scheduling in Real-Time Systems." *Journal of Real-Time Systems*	Discusses future trends in AI-enhanced CPU scheduling.	AI-enhanced scheduling, neural networks	Scalability, real-time learning	Hybrid AI-traditional scheduling approaches
10	Stoica, I., Abdel-Wahab, H. M., & Jeffay, K. (1996). "A Proportional Share Resource Allocation Algorithm for Real-Time, Time-Shared Systems." *IEEE Transactions on Computers*	Introduces a proportional share resource allocation method.	Proportional share scheduling	Resource allocation fairness	Integration of AI for dynamic resource allocation

11	Lehoczky, J. P., Sha, L., & Ding, Y. (1989). "The Rate Monotonic Scheduling Algorithm: Exact Characterization and Average Case Behavior." *Proceedings of the IEEE Real-Time Systems Symposium (RTSS)*	Analyzes the behavior of RMS in real-time systems.	Rate Monotonic Scheduling (RMS)	Worst-case execution time estimation	AI for predictive scheduling in real-time systems
12	Sha, L., et al. (2004). "Real-Time Scheduling Theory and Ada." *Proceedings of the IEEE*	Discussion on real-time scheduling theory and its application to Ada.	Fixed-priority scheduling, deadline-driven scheduling	Language integration, timing constraints	AI-driven enhancements to language-level scheduling
13	Zhang, Y., & Watanabe, Y. (2018). "Adaptive Energy-Aware Task Scheduling for Real-Time Systems." *IEEE Access*	Proposes adaptive scheduling methods for energy efficiency.	Adaptive scheduling, energy-aware scheduling	Energy consumption, system adaptability	AI-based adaptive energy management
14	Rajkumar, R., Lee, C., Sha, L., & Lehoczky, J. P. (1998). "A Resource Allocation Model for QoS Management." *Proceedings of the IEEE Real-Time Systems Symposium (RTSS)*	Introduces a model for managing Quality of Service (QoS) in real-time systems.	QoS management, resource allocation	QoS trade-offs, resource constraints	AI-enhanced QoS management
15	Kang, J., et al. (2017). "Deadline-aware Energy-efficient Scheduling for Real-time Systems." *Journal of Systems Architecture*	Discusses energy-efficient scheduling with a focus on meeting deadlines.	Deadline-aware scheduling, energy efficiency	Balancing energy use and deadlines	AI-driven trade-off analysis in scheduling

16	Sun, J., Liu, H., & Zhao, X. (2020). "Reducing Computational Overhead in AI-Driven Scheduling Systems." *IEEE Transactions on Industrial Informatics*	Examines ways to reduce computational overhead in AI-driven scheduling.	AI-driven scheduling, computational efficiency	Overhead reduction, real-time performance	Lightweight AI models for real-time applications
17	Cheng, X., Liu, F., & Yang, Y. (2021). "Real-Time Learning for Adaptive Scheduling in Dynamic Environments." *Journal of Real-Time Systems*	Explores real-time learning for adaptive scheduling in dynamic environments.	Real-time learning, adaptive scheduling	Learning latency, environmental changes	AI for real-time adaptation and learning
18	Yao, L., Zhang, W., & Han, J. (2021). "Energy-Efficient AI-Driven CPU Scheduling for Real-Time Systems." *IEEE Transactions on Sustainable Computing*	Focuses on AI-driven energy-efficient CPU scheduling.	Energy-efficient scheduling, AI-driven scheduling	Energy constraints, AI integration	Sustainable computing with AI-driven scheduling
19	Shi, J., Wang, X., & Zhao, L. (2020). "Efficient AI Algorithms for Real-Time CPU Scheduling." *IEEE Transactions on Computers*	Discusses efficient AI algorithms for improving CPU scheduling.	AI algorithms, CPU scheduling	Computational efficiency, scalability	AI techniques for enhancing real-time scheduling
20	Zhao, X., Liu, J., & Sun, Y. (2020). "Optimizing Resource Allocation in Real-Time Systems with AI." *IEEE Transactions on Computers*	Examines resource allocation optimization using AI.	Resource allocation, AI-driven optimization	Resource constraints, optimization complexity	AI-driven resource optimization strategies

21	Tang, Z., Xue, Y., & Li, R. (2021). "Combining Traditional and AI-Based CPU Scheduling Methods for Real-Time Systems." *Journal of Systems Architecture*	Proposes a hybrid approach combining traditional and AI-based methods.	Hybrid scheduling, AI-enhanced methods	Method integration, system complexity	Synergy between AI and traditional scheduling
22	Gao, J., & Zhou, M. (2020). "Robust AI Models for Real-Time CPU Scheduling: Challenges and Solutions." *IEEE Transactions on Computers*	Discusses the challenges and solutions in developing robust AI models for real-time scheduling.	AI-driven scheduling, robustness	Model robustness, scheduling accuracy	Enhancing robustness with AI techniques
23	Chen, L., Wang, Y., & Zhang, H. (2021). "Autonomous Systems and AI-Driven Scheduling in Real-Time Environments." *IEEE Transactions on Industrial Informatics*	Investigates AI-driven scheduling in autonomous real-time systems.	AI-driven scheduling, autonomous systems	Autonomy, real-time response	AI-enhanced autonomy in real-time systems
24	Kumar, R., & Singh, D. (2021). "Interpretable AI Models for Real-Time Systems: A Review." *Journal of Real-Time Systems*	Reviews interpretable AI models for real-time systems.	Interpretable AI, real-time scheduling	Interpretability, model complexity	Developing interpretable AI for real-time applications
25	Li, Y., Chen, Z., & Wang, Y. (2022). "Hybrid Approaches for AI-Driven CPU Scheduling in Real-Time Systems." *Journal of Systems and Software*	Explores hybrid approaches combining AI with traditional scheduling.	Hybrid scheduling, AI-driven methods	Hybrid model integration, system efficiency	Hybrid approaches for optimized real-time scheduling

Future Directions and Challenges

Despite significant advancements, several challenges and opportunities for future research in ML-based CPU scheduling remain. Addressing scalability issues in multi-core processors, enhancing real-time adaptability, and integrating heterogeneous workloads are critical areas for further exploration (Mao & Yu, 2018; Wang et al., 2023). Moreover, the development of hybrid approaches combining different ML techniques, alongside comprehensive benchmarking and evaluation frameworks, will be essential to advancing the state-of-the-art in real-time CPU scheduling.

III. CONCLUSION

In conclusion, machine learning techniques hold immense promise for transforming CPU scheduling in real-time systems by providing adaptive, data-driven solutions that enhance efficiency, responsiveness, and reliability. The integration of AI into CPU scheduling for real-time environments has marked a significant evolution in addressing the complexities and dynamic demands of modern computing systems. This survey underscores the advancements in AI-enhanced scheduling techniques, which offer notable improvements in adaptability, efficiency, and robustness over traditional methods. By incorporating machine learning, reinforcement learning, and hybrid models, researchers are effectively tackling challenges such as resource allocation, energy efficiency, and the management of unpredictable workloads. Despite these strides, challenges persist, including the computational overhead of AI models and the need for explainability in safety-critical applications. Future research is expected to refine hybrid approaches, develop real-time adaptive systems, and enhance energy-efficient scheduling, ensuring that AI continues to play a crucial role in advancing resilient and intelligent real-time computing systems.

REFERENCES

[1] Akella, A., & Liang, J. (2020). Heart disease prediction using machine learning techniques: A comparative study. BMC Medical Informatics and Decision Making, 20(1), 1-13. https://doi.org/10.1186/s12911-020-01108-1

[2] Chen, W. T., & Chou, L. C. (2020). Feature selection-based ensemble machine learning for heart disease prediction. IEEE Access, 8, 202149-202160. https://doi.org/10.1109/ACCESS.2020.3035413

[3] Chaganti, K. R., Ramula, U. S., Sathyanarayana, C., Changala, R., Kirankumar, N., & Gupta, K. G. (2023, November). UI/UX Design for Online Learning Approach by Predictive Student Experience. In 2023 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 794-799). IEEE.

[4] Koushik Reddy Chaganti, Chinnala Balakrishna, P.Naresh, P.Rajyalakshmi, 2024, Navigating E-commerce Serendipity: Leveraging Innovator-Based Context Aware Collaborative Filtering for Product Recommendations, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 13, Issue 05 (May 2024).

[5] Koushik Reddy Chaganti, C. Sasikala, G.Sirisha, Shravani Amar, Ravindra Changala, Koppuravuri Gurnadha Gupta, C. B. . (2024). Improved Plant Phenotyping System Employing Machine Learning Based Image Analysis. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(3), 2415–2421.

[6] Sunder Reddy, K. S. ., Lakshmi, P. R. ., Kumar, D. M. ., Naresh, P. ., Gholap, Y. N. ., & Gupta, K. G. . (2024). A Method for Unsupervised Ensemble Clustering to Examine Student Behavioral Patterns. International Journal of Intelligent Systems and Applications in Engineering, 12(16s), 417–429. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/4854.

[7] Nagesh, C., Chaganti, K.R. , Chaganti, S. , Khaleelullah, S., Naresh, P. and Hussan, M. 2023. Leveraging Machine Learning based Ensemble Time Series Prediction Model for Rainfall Using SVM, KNN and Advanced ARIMA+ E-GARCH. *International Journal on Recent and Innovation Trends in Computing and Communication*. 11, 7s (Jul. 2023), 353–358. DOI:https://doi.org/10.17762/ijritcc.v11i7s.7010.

[8] Gopi, K. V., & Ramesh, V. (2019). Heart disease prediction using optimized machine learning techniques. Journal of King Saud University-Computer and Information Sciences. https://doi.org/10.1016/j.jksuci.2018.09.010

[9] Huang, C., Chen, B., & Xu, S. (2020). Heart disease prediction with deep neural network. Proceedings of the 2020 IEEE International Conference on Big Data and Smart Computing (BigComp), 603-606. https://doi.org/10.1109/BigComp48618.2020.9074995