1. Introduction

* Discusses the importance of Predictive Maintenance (PdM) in Industry 4.0 and the challenges in adopting the right maintenance approach.
* Highlights the potential of Reinforcement Learning (RL) for PdM and the lack of RL-based maintenance frameworks in the literature.
* Proposes a novel multi-agent deep RL approach for maintenance policy learning, considering multiple machine failures, technician availability, and skills.

2. Maintenance policies and strategies 2.1 State-of-affairs and trends

* Reviews the literature on dynamic maintenance task scheduling, categorizing methods into metaheuristics, mathematical programming, and machine learning (ML).
* Notes the limitations of metaheuristics and mathematical programming in dealing with stochastic processes, and the growing interest in RL-based approaches.

2.2 RL-based maintenance

* Discusses existing RL-based maintenance strategies for sequential and parallel manufacturing processes, with centralized or distributed agent architectures.
* Identifies limitations in state-of-the-art studies, such as not considering technician availability or multiple failure types.

3. Multi-agent deep RL-based maintenance policy 3.1 System description

* Defines the manufacturing environment with M identical parallel machines, each with Cm components subject to failures modeled by Weibull distributions.
* Describes the three machine states (working, breakdown, maintenance) and the role of technicians T with varying repair times rtc for each component c.

3.2 Multi-agent models

* Formalizes the PdM problem as a Markov Game (MG), an extension of Markov Decision Process (MDP) for multiple agents.
* Defines the system state observed by each agent am, including machine states, technician availability, component states and lifespans, and remaining maintenance time (Eq. 1).
* Specifies the action space for each agent (Eq. 2) and the reward function guiding the agents to keep machines in working state and predict optimal maintenance timing (Eqs. 3-4).

1. Evaluation 4.1 Maintenance policy benchmarking

* Describes the implementation of three benchmark maintenance policies: Corrective Maintenance (CM), Preventive Maintenance (PM), and Random (RA).

4.2 Experimental setup

* Defines two scenarios with varying numbers of machines, technicians, and component failure distributions (Table 5).

4.3 Training of RL agents

* Details the training of RL agents using Proximal Policy Optimization (PPO) and Curiosity-driven exploration, with hyperparameters specified.
* Presents training curves for the 3-machine (3M) and 5-machine (5M) scenarios (Figs. 4a, 5a).

4.4 Maintenance policy comparison analysis

* Compares the RL policy against CM, PM, and RA policies using four metrics: percentage of time in working state, total reward, number of breakdowns, and number of maintenance actions.
* Demonstrates that the RL policy outperforms the benchmarks, keeping machines in working state for ≈60% of the time (Figs. 4b, 5b), achieving 20-35% higher cumulative reward (Figs. 4c, 5c), and reducing breakdowns by ≈75% (Figs. 4d, 5d), at the cost of more frequent maintenance actions (Figs. 4e, 5e).

4.5 Financial implication

* Analyzes the financial impact of the policies by calculating total profit (Eq. 5) under varying costs for breakdowns and maintenance actions.
* Illustrates that the RL policy leads to higher profits in most scenarios, except when maintenance action costs are very high (Fig. 6).

1. Discussion

* Highlights the advantages and challenges of the proposed multi-agent RL approach, such as scalability, non-stationarity, and the need for explainable AI (XAI).
* Suggests future research directions, including joint optimization with production activities, consideration of quality data, and integration of RL with metaheuristics.

1. Conclusion

* Summarizes the novel multi-agent deep RL system for PdM, considering multiple failure types, technician availability, and skills.
* Reports experimental results showing that the RL policy outperforms traditional maintenance policies, reducing breakdown time by up to 80% and failure prevention by up to 75%.
* Acknowledges the need for further economic analysis and validation of the proposed approach.

DEFINITIONS:

1. Predictive Maintenance (PdM): A proactive maintenance strategy that uses data analytics to predict when equipment is likely to fail, enabling just-in-time maintenance interventions to minimize downtime and costs.
2. Markov Decision Process (MDP): A mathematical framework for modeling decision-making in situations where outcomes are partly random and partly controlled by a decision-maker.
3. Markov Game (MG): An extension of MDP to multiple agents, where each agent has its own set of actions and rewards, and the state transitions depend on the joint actions of all agents.
4. Weibull Distribution: A continuous probability distribution commonly used to model equipment failures, with two parameters: scale (α) and shape (β).
5. Proximal Policy Optimization (PPO): A policy gradient method for RL that uses a surrogate objective function to ensure that the policy updates are conservative and stable.
6. Curiosity-Driven Exploration: An intrinsic motivation mechanism for RL agents to explore novel states by encouraging actions that lead to high prediction errors of the next state.