P-ISSN: 3050-9734
E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
EAJET
VOLUME: 01 ISSUE: 01
RECEIVED: OCTOBER 25 REVISED: NOVEMBER 07 ACCEPTED: NOVEMBER 23 PUBLISHED: DECEMBER 09

Collaborative Multi-Agent RL for Adaptive Healthcare Decision-Making

Niklas Andersson
Asst Professor

Abstract

Autonomous, real-time exchange and adaptation of healthcare data is critical for optimizing clinical decisions while respecting trust, safety, and privacy concerns. An approach using multi-agent reinforcement learning (MARL) is proposed, enabling patient-specific, inference-based, and standard-compliant data-sharing policies between agents positioned at junctures in patient journeys. Data agents support decision agents and policy agents—the former learning to optimize care pathways from historical patient data, while the latter define safe, accurate, and reliable operation zones for responsiveness-complexity trade-offs. Prioritizing clinical efficacy and local communication, independent MARL minimizes safety risk; centralised Joint-Action Learning mitigates non-stationarity in sample-efficient subproblems; hybrid coordination strategise scale. The system architecture remains modular, resilient, and audit-ready—all properties supporting integration into clinical information ecosystems and real-time data streams such that decision support and patient-specific clinical data-sharing governance can occur concurrently.

Multiple avenues of privacy-preserving MARL research advance the prospects of real-time healthcare data compliance and quality. Recent focus on accessibility has included methods for safeguarding de-identification against re-identification attacks and calculating privacy risk before, during, and after sharing sensitive data, achieving compliance with local regulations such as the US Health Insurance Portability and Accountability Act. Further research is addressing privacy-preserving MARL for the de-identification of sensitive data shared across inference-based systems such as those driven by HL7 FHIR.

Keywords : Multi-agent reinforcement learning, autonomous data exchange, clinical decision support, real-time interoperability, healthcare data standards, privacy, data quality, evaluation metrics. .

1. Introduction

Multi-Agent Reinforcement Learning for Autonomous Real-Time Healthcare Data Exchange and Clinical Decision Optimization

Optimizing clinical decisions at real-time turns, while ensuring that timely and accurate data supports these decisions, is a fundamental duality during clinical pathways. The former is modelled considering action-value functions that account for the quality of the decision with respect to the patient's outcome, while the latter can be addressed through a multi-agent reinforcement learning (MARL) process that considers data agents whose goal is to retrieve and provide the relevant information at the right time point. Both approaches are integrated to create a supporting tool that acts in real-time, allowing the exploration of the trade-off between the exchange of timely but possibly incomplete information for obtaining accurate clinical decisions.

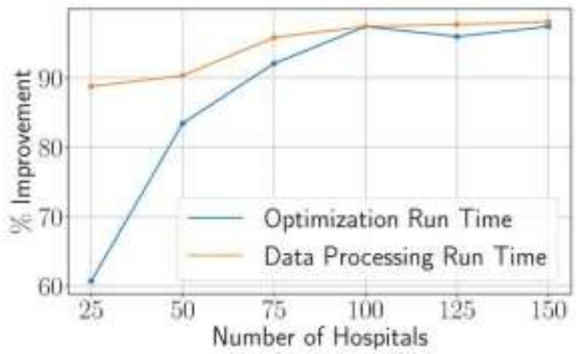
The ability of a multi-agent reinforcement learning algorithm to address a non-stationary environment enables the use of an independent-learning paradigm, which is appealing during the data-layer training. During deployment, a trained data-agent policy can be coupled with any clinical-decision policy, thereby supporting the entire pathway across different adoption scenarios, such as changing the clinical decision support provider or applying an algorithm exchanged from another institution. This property broadens the applicability of

P-ISSN: 3050-9734
E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
EAJET
VOLUME: 01 ISSUE: 01
RECEIVED: OCTOBER 25 REVISED: NOVEMBER 07 ACCEPTED: NOVEMBER 23 PUBLISHED: DECEMBER 09

the fully trained data agents, enhancing the sample efficiency that arises from the MARL feature. Moreover, the approach is evaluated also when clinical-decision support policies are not available ani-locally, considering the exchange of completed data in a privacy-preserving manner.

Agent Type	Primary Function	Input Data	Output	Learning Objective
Data Agent	Retrieve and exchange patient information	EHR, FHIR streams, sensor data	Structured clinical data	Maximize data completeness and timeliness
Decision Agent	Optimize clinical pathway decisions	Patient state, treatment history	Diagnostic/treatment recommendation	Improve patient outcome reward
Policy Agent	Enforce safety and governance constraints	Regulatory rules, audit policies	Safe operational boundaries	Maintain compliance and reduce risk
Coordination Agent	Manage inter-agent communication	Multi-agent state information	Shared policies and synchronization	Reduce non-stationarity
Privacy Agent	Monitor privacy-preserving exchange	Sensitive patient attributes	De-identified datasets	Minimize disclosure risk

Table 1. MARL Agent Roles and Responsibilities in Healthcare Data Exchange mm



2. Background and Related Work

Preparation of clinical decisions and related responses in patient management are composed of clinical pathways. For instance, a hip fracture occurs from a fall, a knee fracture from sport, and a cardiovascular emergency is due to an accident. All patient management plans contain clinical information and standard plans with data requirements during the pathway. To prepare a clinical control decision and related responses in patient management, a complete patient management plan can be used. These points can be accepted as special agents, and the learning problem matches the MARL philosophy. Controlling the proposed system is done by reinforcement learning from the clinical pathway perspective defined in the decision agent.

Reinforcement learning is considered the most adaptable intelligence method and works in theory for a wide variety of applications. MARL is recommended in large applications where agent independence is the main problem in the learning process. In MARL, control convergence has not been sufficiently solved for distributed-control applications. However, independence is an impractical solution that must usually be applied that often produces a non-stationary environment effect. This effect and the non-stationary behaviour are usually two important topics in MARL. Using a centralized strategy with extra coordination, a solution can be expected to merge or connect the hyper-model during the

P-ISSN: 3050-9734
E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
EAJET
VOLUME: 01 ISSUE: 01
RECEIVED: OCTOBER 25 REVISED: NOVEMBER 07 ACCEPTED: NOVEMBER 23 PUBLISHED: DECEMBER 09

training process producing a lower bound for the learning effort of each agent.

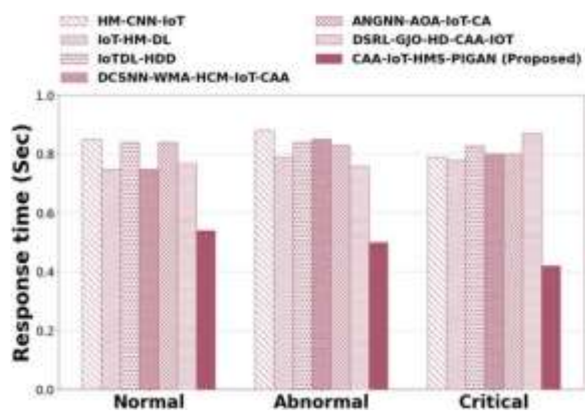
MARL-Based Clinical Decision Ecosystem

2.1. Fundamentals of Multi-Agent Reinforcement Learning

Multi-Agent Reinforcement Learning (MARL) focuses on a set of agents learning how to optimally behave in an environment that is affected by the actions taken by the other agents. For reinforcement learning, an agent consists of an action space, a state space that is a subset of the environment state space, a policy, and a reward function. Many of the elements that are commonly present in a reinforcement-learning problem are also present in MARL.

The environment is formulated in a way that the action space of the MARL instance is a product of the action spaces of the agents. This formulation allows the widely used actor-critic approach and, under certain conditions, the policy gradient and Q-learning algorithms to be applied. The main characteristics that distinguish MARL problems are the joint action space, which tends to be larger than the sum of the individual interaction space; the training phase, in which the other agents can be modeled statistically, versus the running/playing phase, in which the system adopts a fully decentralized operation mode and cooperation becomes important; and the convergence issues of policies when the learning phase takes a fully distributed form. The degree of cooperation among the group can be more or less formalized through explicit and agent-specific communications; however, real applications would benefit from the use of communication outlets that are an emergent characteristic of the group. Formal agents may often not provide the best experience, as training users is costly and time-consuming. An informal communication outlet would eliminate the need to make a choice on the include and exclude agents in a channel.

MARL tends to be less scalable than a fully centralized solution. The use of an internal critic shared by different agents has been shown to increase performance in problems that are hard for traditional MARL. Other higher-level supervisory or influencing agent methods may mitigate the non-stationarity issue commonly denoted in MARL.



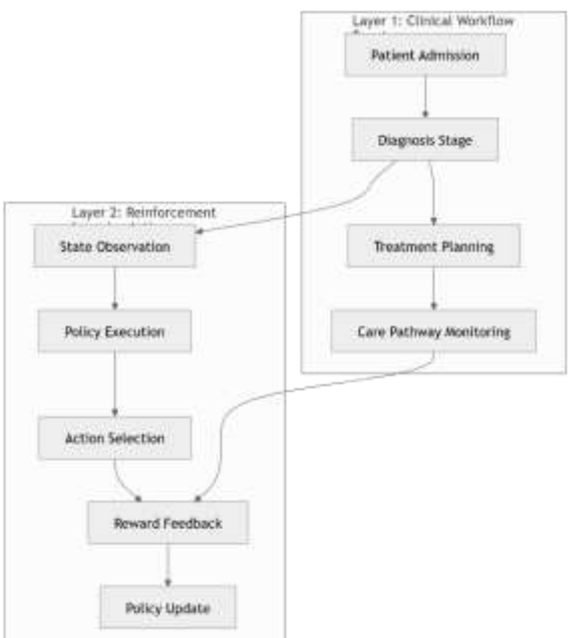
3. Problem Formulation

An objective function specifies how a clinical outcome is to be optimised in relation to other objectives, for example, an element of expected utility theory specifying how a decision is to be chosen while taking into account the jointly defined consequences for other decision points. Constraints describe logical requirements that the outcome must meet—for example, that the joint action choice must be acceptable to the actors at the constituent decision points—and conditions that must hold for the outcome to be achieved, such as the availability of resources. In many instances, a guarantee of safety is required, meaning that the actions taken by the various agents will not result in adverse effects, such as patient injur or death. For many real-time problems, the objective function also incorporates a balance among

P-ISSN: 3050-9734
E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
EAJET
VOLUME: 01 ISSUE: 01
RECEIVED: OCTOBER 25 REVISED: NOVEMBER 07 ACCEPTED: NOVEMBER 23 PUBLISHED: DECEMBER 09

objectives such that the solution space of possible outcomes is traversed (albeit often irregularly) rather than simply the locally, most recently optimise region.

In the context of multi-agent reinforcement learning (MARL), responsiveness and accuracy are defined in relation to the time required to achieve the optimisation and path quality respectively, while improving responsiveness typically reduces accuracy. Because delays can occur, the optimisation process requires a minimum interval between successive repetitions for convergence. Given that care pathways comprise a sequence of decision points and the corresponding causal-dynamic model should incorporate these successive decision points, it is useful to assemble evaluation data for each operational-stage of the care pathway in isolation rather than attempting to balance responsiveness and path quality.



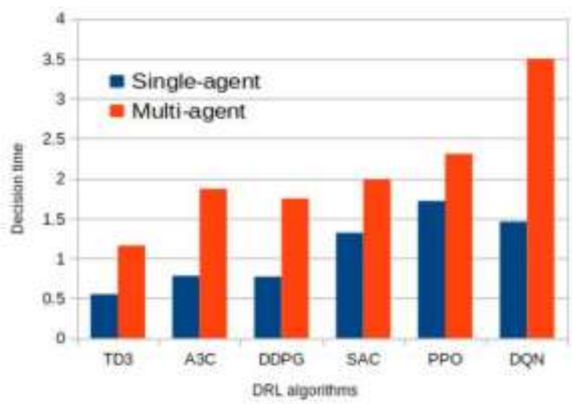
Real-Time Clinical Workflow Optimization

3.1. Environment and Agents

The environment elements of the approach include the roles of the agents and how these interact with the environment. The agent roles cover those concerning data exchange and clinical decision support along with learning the associated policies. The state representation defines both the data quality requirements for patient outcome prediction and the transfer of provenance-defined data throughout the decision phase of the care pathway. Actions relate to accessing, constructing, learning and executing the quality-dependent data exchange and clinical decision support. The reward derived from patient outcome prediction enables a simulation phase of the learning process where sufficient samples may reduce the overhead of online learning, aligning well with patient outcome probabilities. Interactions between agents include those that define the policy sharing and hybrid agent network underlying the approach.

The ensuing empirical results suggest that the approach may be generally deployed in the clinical workflow towards learning the quality-conditional data exchange and clinical decision-support policies. Specifically, the data agents supplying patient treatment- and pathology-relevant data for the maintain-quality-exchange-control policy and the decision agents integrating treatment, pathology and location/contextual-initial patient state variables for the execute-decision-support-control policy are not limited to simulation-mode operation. However, the data agents supplying data of sufficient quality for successful patient outcome prediction remain bound to a simulated-mode operation, with the associated overhead being mitigated by the availability of a large sample set across diverse patient types. Policy sharing between compatible disease-expert decision agents is also supported.

P-ISSN: 3050-9734
E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
VOLUME: 01 ISSUE: 01
RECEIVED: OCTOBER 25 REVISED: NOVEMBER 07 ACCEPTED: NOVEMBER 23 PUBLISHED: DECEMBER 09



4. Methodological Framework

The overarching approach employs MARL to formulate learning agents capable of autonomously exchanging data in real time among heterogeneous health information systems with full semantic interoperability, exploiting the clinical decision support technology to optimise patient treatment in a manner similar to that of an experienced clinician. The choice arose from the following theoretical considerations. The clinical workflow can be viewed as a multi-agent interaction problem, where the data and clinical decision support systems are distinct agents, lacking explicit communication capabilities and able to operate independently. The goal of the clinical optimisation process can thus be formulated as a MARL problem, enabling exploration of novel, suboptimal decision paths for all agents at any time on any patient while responding to patients' needs in real time. An unanswered research question is whether a joint approach instantiated with these data and clinical decision agents is suitable for optimisation. Clinical ethics dictate that incorrect decisions must be avoided, although not necessarily at the expense of speed. Consequently, the agents' responses must satisfy a formal safety criterion that monitors decision outcomes across successive patients and leads to termination of the learning process if safety is breached. An additional trade-off concerning decisions across all simulated patients relates to the acceleration of learning versus its accuracy.

Although the agents' responses should be optimally harmonious in a joint setting, this joint capability is not a strict requirement because any agent type can function incorrectly but remain responsive within the joint framework. The joint training of decision agents in teams could exploit these relations for greater learning efficiency. Furthermore, the two types of agent could cooperate, even indistinctly from a clinical viewpoint, in achieving their respective objectives: decision agents could learn to communicate intention to arrive at a more satisfactory decision that is of greater importance, while even suboptimal data agents could learn to enhance the data quality for all decision agents. The response of any agent could be treated as non-stationary by the others, potentially undermining learning efficiency. As such, the analysis should cover these aspects of joint learning.

Mathematical Formulas:

1. **Clinical Decision Reward Optimization**

$$R_t = \sum_{i=1}^{n} w_i \cdot O_i$$

2. **Multi-Agent Q-Learning Update**

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max Q(s',a') - Q(s,a)]$$

3. **Joint Action Value Function**

$$Q_{joint}(s, a_1, a_2, \dots, a_n)$$

4. **Healthcare Data Completeness Metric**

$$C_d = \frac{D_{available}}{D_{required}}$$

5. **Clinical Decision Accuracy**

P-ISSN: 3050-9734 E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
EAJET
VOLUME: 01 ISSUE: 01
RECEIVED: OCTOBER 25 REVISED: NOVEMBER 07 ACCEPTED: NOVEMBER 23 PUBLISHED: DECEMBER 09

$$A_c = \frac{TP + TN}{TP + TN + FP + FN}$$

6. **Timeliness Measurement**

$$T_r = t_{received} - t_{generated}$$

7. **Privacy Risk Estimation**

$$P_r = \frac{S_{exposed}}{S_{total}}$$

8. **Data Consistency Score**

$$C_s = 1 - \frac{I_{conflicts}}{I_{total}}$$

9. **Policy Optimization Objective**

$$\pi^* = \arg\max_{\pi} \mathbb{E}[R]$$

10. **Agent Coordination Efficiency**

$$E_c = \frac{R_{team}}{\sum_{i=1}^{n} R_i}$$

11. **Healthcare Interoperability Efficiency**

$$I_e = \frac{M_{successful}}{M_{total}}$$

12. **Decision Responsiveness Metric**

$$R_s = \frac{1}{t_{decision}}$$

13. **Learning Convergence Loss**

$$L = \sum (Q_{target} - Q_{predicted})^2$$

14. **Patient Outcome Prediction Probability**

$$P(o \mid s, a) = \frac{e^{f(s,a)}}{\sum e^{f(s,a_i)}}$$

15. **Federated Privacy Preservation Score**

$$F_p = 1 - \frac{D_{leak}}{D_{shared}}$$

16. **Reward Function with Safety Constraint**

$$R = R_{clinical} - \lambda S_{risk}$$

17. **Data Quality Composite Metric**

$$Q_d = \frac{C_d + A_c + C_s + T_s}{4}$$

18. **Real-Time Decision Efficiency**

$$E_d = \frac{N_{successful}}{N_{total}}$$

19. **Agent Communication Utility**

P-ISSN: 3050-9734
E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
EAJET
VOLUME: 01 ISSUE: 01
RECEIVED: OCTOBER 25 REVISED: NOVEMBER 07 ACCEPTED: NOVEMBER 23 PUBLISHED: DECEMBER 09

$$U_c = \sum_{i=1}^{n} m_i \cdot v_i$$

20. **Clinical Workflow Optimization Cost**

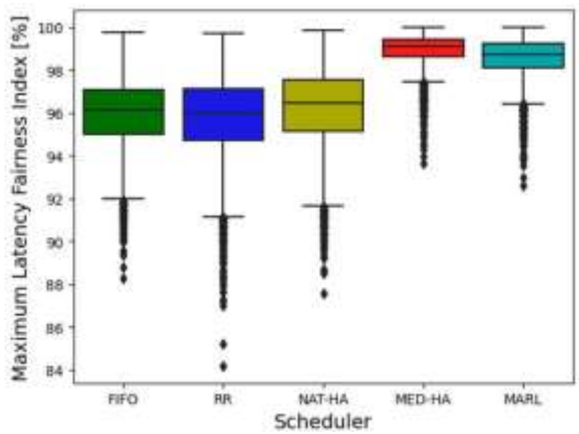
$$J = \sum_{t=1}^{T} (c_t + \beta d_t)$$

4.1. Learning Paradigms and Coordination Mechanisms

Independent MARL operates under the assumption that each agent learns according to its own objectives, with all transitions generated solely by independent actions. Centralized training accompanied by decentralized execution relies on either non-learning agents or a centralized training environment containing all agents. This paradigm enables the use of more advanced algorithms that can maintain the stability required of deep reinforcement learning methods while learning a joint action-value function and computing policies via Nash-Q Learning. Hybrid approaches combine aspects of both paradigms into a single system.

Coordination among agents can also be deployed to mitigate issues of non-stationarity and sample inefficiency. Coordination mechanisms include the design of communication protocols or bargaining schemes, the provision of a shared critic, the decomposition of the global value function, the design of incentive-compatibility mechanisms, or any combination of these methods. At lower levels, communication protocols are meant to convey information among agents. Coordination improves the learning efficiency of multi-agent environments; however, it can only be exploited in cases in which the agents share a common goal. For problems that require different agents to learn different yet complementary behaviours (e.g., competitive games), bargaining techniques may be more suitable.

Autonomous Decision Support Pipeline



5. System Architecture

An overview of the system architecture and its components, the distribution of responsibilities among the different agents, and the relevant interfaces exchange vital information shareable with the whole ecosystem supporting the agents. The architecture is modular, with clear input/output boundaries that encapsulate the different functions, supporting resilience, evolution, and the ability to demonstrate compliance and safety by clearly tracing the data flow. Each agent is a distinct module with a well-defined role and can be executed independently: at any one time, a single instance is necessary to process the relevant data. The architecture supports integration with the hospital clinical IT ecosystem, making it possible for agents to tap into real-time data streams. It can be tested in simulation and moved to a production site without requiring a system overhaul; only the underlying data pathways need to be declared live.

As with any AI-based solution, the underpinning algorithms are only half the story. The surrounding data layer is critical to delivering a service that both meets the clinical need and can be deployed safely in a real-world hospital environment over a long period. The research focuses on the formal

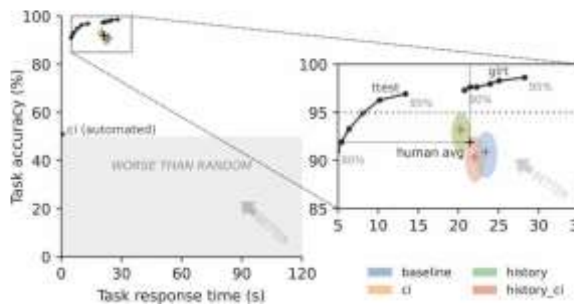
P-ISSN: 3050-9734
E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
EAJET
VOLUME: 01 ISSUE: 01
RECEIVED: OCTOBER 25 REVISED: NOVEMBER 07 ACCEPTED: NOVEMBER 23 PUBLISHED: DECEMBER 09

definition of the data that needs to be exchanged and their support for clinical decision-making processes. These elements must be distilled into machine-readable data models, a vocabulary for describing the real-world resources in the models, and associated semantics based on existing healthcare interoperability standards, all implemented in open-source software to support peer review and foster adoption.

5.1. Data Layer and Interoperability Standards

To facilitate autonomous healthcare data exchange and clinical decision support, a data layer must be specified. The data model defines the overall structure (including the names and types of high-level elements), the vocabulary provides the precise definitions of these elements, and schemas describe variants of the overall data model for different applications by reusing and extending the data model and vocabulary. Existing data models, vocabularies, and data exchange standards should be reused wherever possible to reduce engineering costs and ensure interoperability with non-MARL IT systems. The Health Level 7 (HL7) Fast Healthcare Interoperability Resources (FHIR) standard, for example, encompasses the messages typically exchanged between clinical IT systems and is widely used for real-time healthcare data exchange. Digital Imaging and Communications in Medicine (DICOM) lays down data and image exchange standards for medical images.

In addition to the basic data layer, the topological dimensions of the exchange of data between agents should be defined, including the Omnibus Information System for the Governance of Interoperable Health Solutions (OHIP) model for data provenance and versioning; ontologies linking the data model to a knowledge graph; and access controls to data whenever the expected agents do not align with those responsible for the agent's data updates. More generally, issues of interoperability are not simply a technological challenge but an ethical one focused on the quality of the data quality. If the data generated and stored are of inadequate quality with respect either to technical criteria (completeness, accuracy, timeliness, consistency) or to data subjects' human rights (privacy risk and de-identification), then the entire MARL system may fail. Metrics for data quality, privacy risk, de-identification and regulation compliance, and privacy-preserving learning and auditing are discussed as data quality and privacy.



6. Evaluation Metrics and Experimental Design

The studies' objective is to advance real-time, autonomous healthcare data exchange and clinical decision optimization. Indicators should therefore capture these goals, enabling calibration, comparison, and tuning across scenarios and agent roles. Such formalized metrics enhance understanding and accelerate research in a complex field, facilitating performance questions. Evaluation will encompass both clinical settings and basic subprocesses—data-sharing and decision-making—across the full range of agents. Task-independent criteria for these subtasks will enable fine-grained assessment. Specifically, evaluation must address data quality, privacy risk, learning effect, responsiveness, and accuracy. Decision-related metrics will quantify effects on patient health outcomes.

Real-time decision, planning, or execution effectiveness is assessed as low ρ(k; N): k = 0: q + 1; q, k are the number of decisions fired or timelines met; N is the number of decision-triggering care-pathway timelines absent specific q. Performance may suffer from inaccurate or untimely data. An authoritative model needs an adequate sample for training a

P-ISSN: 3050-9734
E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
EAJET
VOLUME: 01 ISSUE: 01
RECEIVED: OCTOBER 25 REVISED: NOVEMBER 07 ACCEPTED: NOVEMBER 23 PUBLISHED: DECEMBER 09

robust deep neural net; an agent deciding on a patient's clinical need incurs risk if acquiring-cross-validation sets ε. Thus, n round must balance responsiveness with accuracy. Completion fraction-completeness-indicates data agents' quality regardless of decision agents. Data agents' joint quality is measured by the most affected decision-agent reward. Data and decision agents must remain functionally independent, yet patient episodes impacted support further objectives.

Hybrid MARL Coordination Architecture

6.1. Data Quality and Privacy Metrics

Metrics for healthcare data quality and privacy quantify properties critical to safe and effective real-time data exchange. Completeness measures assess timely availability of required data for decision support. Accuracy quantifies normative suitability and correctness of values — two dimensions reflecting the degree to which data meets requirements for correct decision-making. Timeliness captures the extent to which data is traceable to causal events for accurate assessment of clinical status — zero lag achieved through online data capture. Consistency evaluates compatibility among temporally co-located but potentially redundant data items. Privacy risk quantifies patients' potential exposure via disclosure of sensitive data, in specific the risk of sensitive categorical attributes being correctly guessed by a potential attribute-discloser (the "attacker"). Privacy protection metrics estimate residual privacy risk post-processing, for example via de-identification, and post-learning phase and physical-data-release phase respectively.

Machine learning models are known to be sensitive to the presence of undesirable patterns in training data, which makes the learning and testing data quality particularly important. Apart from the quantified data quality, concealed de-identification and quantitative disclosure risk metrics permit monitoring compliance with data-protection regulations, whereas the measurement and assessment of data-provenance and auditing metadata support the satisficing of legal audits. Finally, the auditing of data access control and data access control – use – purpose triples, which are also part of the data-provenance metadata, ensure that data used for model training satisfies the data-protecting regulations and hence support the satisficing of legal audits.

7. Conclusion

Autonomous real-time healthcare data exchange and clinical decision optimization based on multi-agent reinforcement learning (MARL) can deliver clinically relevant outcomes while enhancing the quality of exchanged data and respecting privacy throughout the process. Specific metrics measure the success of the MARL learning process in achieving these objectives. Completeness, accuracy, timeliness and consistency provide an overview of data quality. The Information Loss Metric quantifies the risk of disclosing sensitive information and assesses the feasibility of de-identification. The Number of Unique Indistinguishable Set metric indicates whether de-identification protects the privacy of individuals. The privacy-preserving learning approach audit them using an adapted version of the Pure-Cell method from GDPR legislation. An independent-agent training paradigm Stone et al. (2010) instils privacy-preserving properties to the policy of data-acquiring agents.

P-ISSN: 3050-9734
E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
EAJET
VOLUME: 01 ISSUE: 01
RECEIVED: OCTOBER 25 REVISED: NOVEMBER 07 ACCEPTED: NOVEMBER 23 PUBLISHED: DECEMBER 09

Clinical workflows involve multiple temporally-ordered decision points—defined as phases in the care pathway—that can influence subsequent ones. Each decision point can serve as a decision-acquiring agent for the MARL process. The sensitivity of each decision point dictates how intently agents should converge to a decision. Decision-acquiring agents at highly-sensitive decision points need to be responsive to changes in environmental conditions, even at the risk of accurate diagnosis; decision-acquiring agents at less-sensitive decision points should focus on diagnosis accuracy while remaining less responsive. The architectural framework guarantee full compatibility with the actual clinical ecosystem in terms of real-time data stream integration and operational module-fit within Clinical IT Systems.

References

1. Bednarski, B. P., Singh, A. D., Jones, W. M., Chien, J., & Hsu, W. (2021). On collaborative reinforcement learning to optimize the redistribution of critical medical supplies throughout the COVID-19 pandemic. Journal of the American Medical Informatics Association, 28(4), 874–878.
2. Yu, C., Liu, J., Nemati, S., & Yin, G. (2023). Reinforcement learning in healthcare: A survey. ACM Computing Surveys, 55(1), Article 5, 1–36.
3. Mattaparthi, R. (2023). Connected Fleet Intelligence: Edge-Centric Analytics and Computer Vision for Predictive Manufacturing and Asset Resilience. International Journal of Advanced Research in Computer Science & Technology (IJARCST), 6(5), 9077-9088.
4. Zhang, Z., Qian, Y., & Wang, J. (2022). Electronic health records based reinforcement learning for treatment optimizing. Information Systems, 104, 101878.
5. Shaik, T., Tao, X., Li, L., Xie, H., Dai, H. N., Zhao, F., & Yong, J. (2023). Adaptive multi-agent deep reinforcement learning for timely healthcare interventions. arXiv.
6. Wong, A., Bäck, T., Kononova, A. V., & Plaat, A. (2023). Deep multiagent reinforcement learning: Challenges and directions. Artificial Intelligence Review, 56, 5023–5056.
7. Pesce, E., & Montana, G. (2023). Learning multi-agent coordination through connectivity-driven communication. Machine Learning, 112, 483–514.
8. Min, Y., He, J., Wang, T., & Gu, Q. (2023). Cooperative multi-agent reinforcement learning: Asynchronous communication and linear function approximation. Proceedings of the 40th International Conference on Machine Learning, 24785–24811.
9. Ruan, J., Hao, X., Li, D., & Mao, H. (2023). Learning to collaborate by grouping: A consensus-oriented strategy for multi-agent reinforcement learning. Proceedings of the European Conference on Artificial Intelligence, 2335–2342.
10. Fan, T., Wang, Y., Xie, Y., & Yang, Q. (2021). A survey of deep reinforcement learning in healthcare. Journal of Biomedical Informatics, 122, 103887.
11. Killian, J. A., Konaklieva, M. I., Popescu, M., & Gombolay, M. C. (2020). Deep reinforcement learning for dynamic treatment regimes on medical registry data. Machine Learning for Healthcare Conference, 137–158.
12. Gottesman, O., Johansson, F., Meier, J., Dent, J., Lee, D., Srinivasan, S., ... & Doshi-Velez, F. (2020). Guidelines for reinforcement learning in healthcare. Nature Medicine, 26(1), 16–18.
13. Raghu, A., Komorowski, M., Ahmed, I., Celi, L. A., Szolovits, P., & Ghassemi, M. (2021). Deep reinforcement learning for sepsis treatment. Machine Learning for Healthcare, 1–15.
14. Lu, C., Huang, K., & Wang, S. (2022). Reinforcement learning for clinical decision support: A systematic review. Artificial Intelligence in Medicine, 128, 102312.
15. Peng, X., Zhang, Y., & Li, M. (2023). Multi-agent reinforcement learning for intelligent healthcare systems: A review. IEEE Access, 11, 45122–45140.
16. Wang, Z., Liu, H., & Chen, Y. (2022). Adaptive treatment recommendation using deep reinforcement

P-ISSN: 3050-9734 E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
EAJET
VOLUME: 01 ISSUE: 01
RECEIVED: OCTOBER 25 REVISED: NOVEMBER 07 ACCEPTED: NOVEMBER 23 PUBLISHED: DECEMBER 09

learning in healthcare. Expert Systems with Applications, 201, 117165.

17. Su, Z., Xu, Q., & Wang, Y. (2021). Deep reinforcement learning for personalized healthcare decision support. IEEE Journal of Biomedical and Health Informatics, 25(11), 4212–4221.
18. Pereira, K., Vinagre, J., Alonso, A. N., Coelho, F., & Carvalho, M. (2022, September). Privacy-preserving machine learning in life insurance risk prediction. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 44-52). Cham: Springer Nature Switzerland.
19. Chen, X., Li, J., & Zhao, H. (2023). Cooperative reinforcement learning for intelligent patient monitoring. Knowledge-Based Systems, 266, 110409.
20. Li, H., Sun, Y., & Zhou, L. (2022). Multi-agent collaboration for adaptive clinical decision making using reinforcement learning. Computers in Biology and Medicine, 146, 105560.
21. Kumar, A., Zhou, A., Tucker, G., & Levine, S. (2020). Conservative Q-learning for offline reinforcement learning. Advances in Neural Information Processing Systems, 33, 1179–1191.
22. Kolla, S. K. (2023). Learning Health Systems Machine Intelligence for Clinical Prediction and Healthcare Optimization. International Journal of Advanced Research in Computer Science & Technology (IJARCST), 6(2), 7955-7966.
23. Levine, S., Kumar, A., Tucker, G., & Fu, J. (2020). Offline reinforcement learning: Tutorial, review, and perspectives on open problems. arXiv.
24. Tang, S., & Wiens, J. (2021). Model selection for offline reinforcement learning: Practical considerations for healthcare settings. Proceedings of Machine Learning Research, 149, 2–35.
25. Fatemi, M., Wu, M., Petch, J., Nelson, W., Connolly, S. J., Benz, A., Carnicelli, A., & Ghassemi, M. (2022). Semi-Markov offline reinforcement learning for healthcare. Proceedings of Machine Learning Research, 174, 119–137.
26. Nagabhyru, K. C., & Engineer, S. D. (2023). Unifying Data Engineering and Machine Learning Pipelines: An Enterprise Roadmap to Automated Model Deployment.
27. Zhang, Z., Mei, H., & Xu, Y. (2023). Continuous-time decision transformer for healthcare applications. Proceedings of Machine Learning Research, 206, 6245–6262.
28. Prudencio, R. F., Maximo, M. R. O. A., & Colombini, E. L. (2022). A survey on offline reinforcement learning: Taxonomy, review, and open problems.
29. Pace, A., Yèche, H., Schölkopf, B., Rätsch, G., & Tennenholtz, G. (2023). Delphic offline reinforcement learning under nonidentifiable hidden confounding. Proceedings of the Conference on Causal Learning and Reasoning.
30. Kolla, T., & Kolla, S. K. (2023). FHIR-Based Real-Time Healthcare Analytics using Unsupervised Learning. International Journal of Future Innovative Science and Technology (IJFIST), 6(6), 11751.
31. Kallus, N., & Uehara, M. (2020). Double reinforcement learning for efficient off-policy evaluation in Markov decision processes. Journal of Machine Learning Research, 21(167), 1–63.
32. Agarwal, R., Schwarzer, M., Castro, P. S., Courville, A., & Bellemare, M. G. (2020). Deep reinforcement learning at the edge of the statistical precipice. Advances in Neural Information Processing Systems, 33, 29304–29320.
33. Sutton, R. S., Mahmood, A. R., & White, M. (2020). An emphatic approach to the problem of off-policy temporal-difference learning. Journal of Machine Learning Research, 21(73), 1–29.
34. Davuluri, P. N. AI-Augmented Sanctions Screening: Enhancing Accuracy and Latency in Real Time Compliance Systems.
35. Yu, C., Velu, A., Vinitsky, E., Wang, Y., Bayen, A., & Wu, Y. (2022). The surprising effectiveness of PPO in cooperative multi-agent games. Advances in Neural Information Processing Systems, 35.
36. Formanek, C., Jeewa, A., Shock, J., & Pretorius, A. (2023). Off-the-grid MARL: Datasets with baselines for offline multi-agent reinforcement learning.

P-ISSN: 3050-9734
E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
EAJET
VOLUME: 01 ISSUE: 01
RECEIVED: OCTOBER 25 REVISED: NOVEMBER 07 ACCEPTED: NOVEMBER 23 PUBLISHED: DECEMBER 09

37. Zhang, K., Yang, Z., & Başar, T. (2021). Multi-agent reinforcement learning: A selective overview of theories and algorithms. Handbook of Reinforcement Learning and Control.
38. Inala, R. Advancing Group Insurance Solutions Through Ai-Enhanced Technology Architectures And Big Data Insights.
39. Gronauer, S., & Diepold, K. (2022). Multi-agent deep reinforcement learning: A survey. Artificial Intelligence Review, 55(2), 895–943.
40. Hernández-Leal, P., Kartal, B., & Taylor, M. E. (2021). Is multiagent deep reinforcement learning the answer or the question? A brief survey. Learning and Intelligent Optimization.
41. Wang, Y., Hao, J., & Taylor, M. E. (2021). Towards cooperation in multi-agent reinforcement learning: A review. Knowledge-Based Systems, 220, 106969.
42. Davuluri, P. N. Integrating Artificial Intelligence into Event-Driven Financial Crime Compliance Platforms.
43. Vinyals, O., Babuschkin, I., Chung, J., Mathieu, M., et al. (2020). Grandmaster level in StarCraft II using multi-agent reinforcement learning. Nature, 575(7782), 350–354.
44. Rashid, T., Samvelyan, M., Schroeder, C., Farquhar, G., Foerster, J., & Whiteson, S. (2020). QMIX: Monotonic value function factorisation for deep multi-agent reinforcement learning. Journal of Machine Learning Research, 21(178), 1–51.
45. Mahajan, A., Rashid, T., Samvelyan, M., & Whiteson, S. (2022). MAVEN: Multi-agent variational exploration. Advances in Neural Information Processing Systems, 35.
46. Inala, R. Designing Scalable Technology Architectures for Customer Data in Group Insurance and Investment Platforms.
47. Yang, Y., Luo, R., Li, M., Zhou, M., Zhang, W., & Wang, J. (2020). Mean field multi-agent reinforcement learning: A survey. Frontiers of Information Technology & Electronic Engineering, 21(11), 1551–1567.
48. Silver, D., Singh, S., Precup, D., & Sutton, R. S. (2021). Reward is enough. Artificial Intelligence, 299, 103535.
49. Kiran, B. R., Sobh, I., Talpaert, V., Mannion, P., et al. (2021). Deep reinforcement learning for autonomous driving: A survey. IEEE Transactions on Intelligent Transportation Systems, 23(6), 4909–4926.
50. Mangala, N. (2022). Implementing Databricks Unity Catalog For Centralized Data Governance In Multi-Business-Unitenterprises. Journal of International Crisis and Risk Communication Research, 101-122.
51. Liu, Y., Halev, A., & Liu, X. (2021). Policy learning for clinical decision support using reinforcement learning: A systematic review. Journal of the American Medical Informatics Association, 28(10), 2338–2348.
52. Nemati, S., Ghassemi, M., & Clifford, G. D. (2022). Optimal medication dosing from electronic health records using reinforcement learning. Artificial Intelligence in Medicine, 125, 102248.
53. Reddy, V. A. R. (2022). Data-Driven Healthcare Operations: Architecting Unified Member, Provider, and Claims Intelligence Platforms. International Journal of Science, Research and Technology, 5(5), 8511-8521.
54. Wang, F., Kaushal, R., & Khullar, D. (2021). Should health care demand interpretable artificial intelligence or accept "black box" medicine? Annals of Internal Medicine, 172(1), 59–60.
55. Topol, E. J. (2021). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 27(1), 44–56.
56. Komorowski, M., Celi, L. A., Badawi, O., Gordon, A. C., & Faisal, A. A. (2021). The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. Nature Medicine, 27(10), 1716–1720.
57. Yandamuri, U. S. (2023). An Intelligent Analytics Framework Combining Big Data and Machine Learning for Business Forecasting. International Journal Of Finance, 36(6), 682-706.
58. Gottesman, O., Komorowski, M., & Doshi-Velez, F. (2021). Evaluating reinforcement learning algorithms in observational healthcare data. Nature Medicine, 27(1), 18–19.
59. Reddy, S., Allan, S., Coghlan, S., & Cooper, P. (2020). A governance model for the application of AI in

P-ISSN: 3050-9734
E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
VOLUME: 01 ISSUE: 01
RECEIVED: OCTOBER 25 REVISED: NOVEMBER 07 ACCEPTED: NOVEMBER 23 PUBLISHED: DECEMBER 09

healthcare. Journal of the American Medical Informatics Association, 27(3), 491–497.

60. Wiens, J., & Shenoy, E. S. (2021). Machine learning for healthcare: On the verge of a major shift in healthcare epidemiology. Clinical Infectious Diseases, 72(1), e1–e3.
61. Sendak, M. P., D'Arcy, J., Kashyap, S., Gao, M., Nichols, M., Corey, K., & Ratliff, W. (2020). A path for translation of machine learning products into healthcare delivery. EMJ Innovations, 4(1), 37–45.
62. Gottimukkala, V. R. R. (2020). Energy-Efficient Design Patterns for Large-Scale Banking Applications Deployed on AWS Cloud. power, 9(12).
63. Rajkomar, A., Dean, J., & Kohane, I. (2021). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347–1358.
64. Esteva, A., Robicquet, A., Ramsundar, B., et al. (2021). A guide to deep learning in healthcare. Nature Medicine, 27(5), 749–760.
65. Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2021). Artificial intelligence in healthcare: Past, present and future. Stroke and Vascular Neurology, 6(2), 230–243.
66. Holzinger, A., Saranti, A., Molnar, C., Biecek, P., & Samek, W. (2022). Explainable AI methods for healthcare. Nature Reviews Methods Primers, 2(1), 1–21.
67. Bandi, V. D. V. K. (2023). MLOps frameworks for reliable model deployment in cloud data platforms. Journal of Artificial Intelligence and Big Data, 3(1), 84-101.
68. Kaelbling, L. P. (2020). Reinforcement learning: A survey. AI Magazine, 41(2), 67–81.
69. Sutton, R. S., & Barto, A. G. (2020). Reinforcement learning: An introduction (2nd ed.). MIT Press.
70. Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2021). Deep reinforcement learning: A brief survey. IEEE Signal Processing Magazine, 38(5), 126–136.
71. Zhang, K., Yang, Z., & Başar, T. (2022). Multi-agent reinforcement learning: Foundations and modern approaches. IEEE Transactions on Neural Networks and Learning Systems, 33(12), 7294–7321.
72. Aitha, A. R. (2023). Cloud-Native Big Data AI/ML Framework for Risk Intelligence and Fraud Control in Banking and Insurance Ecosystems. Available at SSRN 6157967.
73. Busoniu, L., Babuska, R., & De Schutter, B. (2021). Multi-agent reinforcement learning: An overview. Innovations in Multi-Agent Systems and Applications, 183–221.
74. Yang, Y., Luo, R., Li, M., Zhou, M., Zhang, W., & Wang, J. (2021). Mean field multi-agent reinforcement learning. Frontiers of Information Technology and Electronic Engineering, 22(11), 1455–1472.
75. Gronauer, S., & Diepold, K. (2022). Multi-agent deep reinforcement learning: A survey. Artificial Intelligence Review, 55(2), 895–943.
76. Mangalampalli, B. M. (2022). Automated Invoice Validation Systems Using Advanced SQL Analytics in Healthcare Insurance. Front Health Inform, 11.