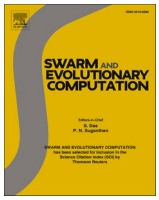
Swarm and Evolutionary Computation 91 (2024) 101769

Contents lists available at ScienceDirect

Swarm and Evolutionary Computation

journal homepage: www.elsevier.com/locate/swevo



ERLNEIL-MDP: Evolutionary reinforcement learning with novelty-driven exploration for medical data processing

Jianhui Lv [a,b,*], Byung-Gyu Kim [c], Adam Slowik [d], B.D. Parameshachari [e], Saru Kumari [f], Chien-Ming Chen [g], Keqin Li [h]

[a] The First Affiliated Hospital of Jinzhou Medical University, Jinzhou, PR China
[b] Peng Cheng Laboratory, Shenzhen, PR China
[c] Sookmyung Women's University, Seoul, Republic of Korea
[d] Koszalin University of Technology, Koszalin, Poland
[e] Nitte Meenakshi Institute of Technology, Bengaluru, Karnataka 560064, India
[f] Chaudhary Charan Singh University, Meerut, India
[g] Nanjing University of Information Science & Technology, PR China
[h] State University of New York, New Paltz, NY 12561, USA

ARTICLE INFO

Keywords:
Evolutionary reinforcement learning
Medical data processing
Novelty-driven exploration
Imitation learning

ABSTRACT

The rapid growth of medical data presents opportunities and challenges for healthcare professionals and researchers. To effectively process and analyze this complex and heterogeneous data, we propose evolutionary reinforcement learning with novelty-driven exploration and imitation learning for medical data processing (ERLNEIL-MDP) algorithm, including a novelty computation mechanism, an adaptive novelty-fitness selection strategy, an imitation-guided experience fusion mechanism, and an adaptive stability preservation module. The novelty computation mechanism quantifies the novelty of each policy based on its dissimilarity to the population and historical data, promoting exploration and diversity. The adaptive novelty-fitness selection strategy balances exploration and exploitation by considering policies' novelty and fitness during selection. The imitation-guided experience fusion mechanism incorporates expert knowledge and demonstrations into the learning process, accelerating the discovery of effective solutions. The adaptive stability preservation module ensures the stability and reliability of the learning process by dynamically adjusting the algorithm's hyperparameters and preserving elite policies across generations. These components work together to enhance the exploration, diversity, and stability of the learning process. The significance of this work lies in its potential to revolutionize medical data analysis, leading to more accurate diagnoses and personalized treatments. Experiments on MIMIC-III and n2c2 datasets demonstrate ERLNEIL-MDP's superior performance, achieving F1 scores of 0.933 and 0.928, respectively, representing 6.0 % and 6.7 % improvements over state-of-the-art methods. The algorithm exhibits strong convergence, high population diversity, and robustness to noise and missing data.

1. Introduction

Recently, the healthcare sector has experienced a notable increase in the generation and accumulation of medical data, propelled by the extensive implementation of electronic health records (EHRs), the expansion of medical imaging technologies, and the growing accessibility of wearable health monitoring devices [1–3]. This extensive collection of medical data offers both benefits and challenges for healthcare practitioners and researchers. The proliferation of data can transform healthcare by facilitating personalized medicine, enhancing disease diagnosis and prognosis, and refining treatment options [4–6]. Conversely, medical data's varied, high-dimensional, and frequently unstructured characteristics present considerable problems for conventional data processing and analysis methods [7,8]. Advanced computational methods are essential for efficiently leveraging medical data and extracting meaningful insights. Machine learning, especially deep learning, has emerged as a viable approach for tackling the complexity of medical data [9]. Convolutional neural networks (CNNs) and

* Corresponding author.
E-mail addresses: lvjh@pcl.ac.cn (J. Lv), bg.kim@sookmyung.ac.kr (B.-G. Kim), adam.slowik@tu.koszalin.pl (A. Slowik), paramesh@nmit.ac.in (B.D. Parameshachari), chienmingchen@ieee.org (C.-M. Chen), lik@newpaltz.edu (K. Li).

https://doi.org/10.1016/j.swevo.2024.101769
Received 8 July 2024; Received in revised form 27 September 2024; Accepted 25 October 2024
Available online 30 October 2024

recurrent neural networks (RNNs), sophisticated variants of deep learning algorithms, have demonstrated remarkable success across multiple medical domains. These encompass medical image analysis, electronic health record processing, and biological signal analysis [10–12]. Nevertheless, deep learning models sometimes necessitate significant amounts of labeled data for training, which can be costly and labor-intensive to obtain in the medical domain [13].

Reinforcement learning (RL) is acknowledged as an effective approach for agents to gain an understanding of the solution space through actions and interactions with the environment. This enables them to enhance their methodologies [14] perpetually. RL algorithms, including Q-learning and policy gradient methods, have been successfully applied in diverse domains such as robotics, gaming, and autonomous navigation [15,16]. Ongoing improvements in approaches considerably enhance RL capabilities through deep learning, enabling the acquisition of complex strategies directly from high-dimensional sensory inputs [17]. Deep RL (DRL), integrating DL and RL, has made substantial progress in various domains, such as mastering the game of Go and controlling robotic manipulators [18]. Even with the accomplishments of DRL in various applications, its potential in medical data processing still needs to be investigated. Medical data has distinct obstacles for DRL algorithms, including reward sparsity, confounding variables, and the necessity for interpretability [19]. Furthermore, DRL algorithms frequently experience sample inefficiency and instability, necessitating numerous interactions with the environment to acquire successful rules. Researchers have adopted evolutionary algorithms (EAs) as a supplementary method to DRL to tackle these issues [20].

Recent studies have been undertaken to enhance EAs by including RL approaches. This methodology is called RL-assisted EA (RL-EA) [21]. RL-EA employs acquired search information to improve solutions collaboratively, demonstrating its efficacy across various domains, including optimization, scheduling, and gaming. The use of RL in EAs enhances the efficiency of search space exploration, directing the evolutionary process toward attractive areas and expediting convergence. Additionally, certain studies seek to include EAs in RL, referred to as evolutionary RL (ERL). EA predominantly manages activities within this algorithmic framework, including hyperparameter optimization, policy search, exploration, and reward shaping. ERL has demonstrated capability in managing extensive and intricate RL tasks, including robotic control and autonomous driving. By implementing population-based search and diversity maintenance methods, ERL can proficiently navigate the policy space and surmount the constraints of conventional RL algorithms, including sparse rewards and local optima.

Despite the successful use of RL-EA and ERL across several areas, the theoretical examination of algorithms, benchmarks, training methodologies, and strategy formulation continues to be a vibrant field of research. The amalgamation of RL and EAs for medical data processing presents other challenges, such as the necessity for interpretability, noise, uncertainty, and the demand for durable and reliable solutions [22]. Therefore, it is essential to explore novel strategies to enhance algorithmic efficiency and tailor these approaches to address the particular requirements of medical data processing tasks.

ERLNEIL-MDP is distinguished by its distinctive amalgamation of novelty-driven exploration and imitation learning within the framework of ERL, specifically designed for medical data processing. This integration facilitates the effective investigation of intricate medical data landscapes while utilizing expert knowledge, tackling the essential challenge of reconciling innovation with conventional medical procedures. The primary aim of this project is to create a resilient and novel algorithm for medical data processing that tackles significant obstacles in the domain. Our objective is to develop a system proficient in managing various forms of medical data, encompassing structured electronic health records and unstructured clinical notes, while markedly enhancing the precision and efficacy of data analysis. We aim to facilitate more precise diagnoses and tailored treatment strategies through sophisticated data processing methodologies. We want to create an interpretable AI system that delivers explainable decisions, fulfilling the essential requirement for transparency in healthcare applications.

Therefore, this paper proposes a novel ERL algorithm incorporating novelty-driven exploration and imitation learning for processing complex medical data. The proposed algorithm addresses the exploration-exploitation dilemma in RL by introducing a novelty computation mechanism and a combination selection strategy. Furthermore, we employ an experience fusion imitation approach to enhance learning efficiency and a training stability module to ensure stable convergence. The architecture of the proposed algorithm is designed to handle the high-dimensional and heterogeneous nature of medical data effectively.

The main contributions of this paper can be summarized as follows:

- We introduce a novelty computation mechanism that quantifies the novelty of individuals based on their dissimilarity to the population and the historical data, promoting efficient search space exploration.
- We propose a combination selection strategy that balances the exploration and exploitation by considering individuals' novelty and fitness during the selection process.
- We employ an experience fusion imitation approach that preserves and propagates useful knowledge across generations, accelerating the learning process and improving overall performance.
- We design a training stability module that dynamically adjusts the learning rate and mutation strength to ensure stable convergence and avoid premature stagnation.

The remainder of this paper is organized as follows. Section 2 reviews the related works in EAs and medical data processing. Section 3 provides an overview of the fundamental concepts. Section 4 presents the proposed ERL algorithm with novelty-driven exploration and imitation learning. Section 5 describes the experimental setup and discusses the results. Section 6 gives the limitation and discussion of the work. Finally, Section 7 concludes the paper and outlines future research directions.

2. Related work

2.1. Evolutionary reinforcement learning

EAs are a category of optimization techniques grounded in the principles of natural evolution, encompassing reproduction, mutation, recombination, and selection. EAs have effectively addressed several challenges, including numerical optimization, combinatorial optimization, and machine learning. EAs enhance exploration, diversity, and robustness in the learning processes of ERL and medical data processing. The fundamental principle of EAs is to repeatedly generate a population of potential solutions, often depicted as chromosomes or genotypes, by applying various genetic operators over multiple generations. The assessment of each candidate solution's fitness relies on a predetermined objective function that measures the quality or performance of the solution inside the defined problem area. The process of evolution entails the continual selection of the most adapted individuals from the current population, the utilization of genetic operators to generate new children, and the substitution of less adapted individuals with newly created ones.

An established instance of an EA is the genetic algorithm (GA), which utilizes a binary or real-valued representation of prospective solutions [23,24]. Within the context of ERL, the prospective solutions in a GA can denote the parameters linked to a policy or value function in RL. The GA modifies the population of policies through selection, crossover, and mutation operators to produce new policies that retain advantageous characteristics from their predecessors while including specific alterations. The selection operator identifies the most appropriate individuals from the current population to serve as parents for the next generation. Two often employed selection strategies in EAs are tournament and roulette wheel selection. In tournament selection, individuals are randomly chosen to compete against one another based on their

fitness levels. In roulette wheel selection, individuals are selected with a probability proportionate to their fitness level. EAs have been employed in medical data processing for several objectives, including feature selection, parameter optimization, and model selection. GAs have been utilized to choose the most informative features from intricate medical datasets, including gene expression data and electronic health records, to improve the accuracy of predictive models [25]. EAs have been employed to optimize the hyperparameters of machine learning models, including support vector machines and deep neural networks, enhancing their capacity to generalize medical data [26].

ERL integrates EAs with RL to resolve the exploration-exploitation issue and enhance learning efficiency [27]. ERL algorithms sustain a population of policies and refine them through EA operations, simultaneously utilizing RL principles to modify the policies depending on environmental interactions. Hu et al. [28] introduced a DRL-assisted co-evolutionary differential evolution method for addressing limited optimization issues. Wu et al. [29] presented an innovative ERL algorithm utilizing particle swarm optimization to identify optimal action sequences. In quantitative trading, DRL agents have arisen to enhance decision-making across various market conditions, formulating lucrative trading strategies by integrating insights from historical data. Takara et al. [30] presented a novel trading system based on the deep Q-network. Parham et al. [31] presented an innovative deep clustering method termed automatic deep sparse clustering, which employs a dynamic population-based EA utilizing RL and transfer learning. Bora et al. [32] presented an enhanced version of the non-dominated sorting GA that integrates a parameter-free self-tuning method utilizing RL. Bora et al. [33] introduced an enhanced non-dominated sorting GA II that integrates a parameter-free self-tuning mechanism through RL.

Even with the efficacy of current ERL algorithms, they frequently encounter challenges due to medical data's high-dimensional and varied characteristics, resulting in sluggish convergence and unsatisfactory outcomes.

2.2. Medical data processing

Medical data processing has significantly advanced in recent years, with machine learning and artificial intelligence playing more vital roles in various medical applications. These advancements have improved diagnosis, tailored treatment alternatives, and superior patient care.

Numerous significant researches have arisen in the field of disease prediction and classification. Mostafa et al. [34] provided a thorough methodology for predicting hepatocellular carcinoma, evaluating the efficacy of various machine learning algorithms. Their research on feature reduction is especially pertinent to our study, as it tackles the problem of high-dimensional medical data, a prevalent concern in healthcare analytics. Likewise, Farghaly et al. [35] established a machine-learning framework to predict the Hepatitis C Virus among healthcare workers in Egypt, illustrating the applicability of these techniques in practical clinical environments.

Data preparation and balancing methodologies in medical data analysis are paramount. Khairy et al. [36] examined the efficacy of rebalancing approaches in mitigating class imbalance within cyberbullying datasets. Their findings offer significant insights applicable to medical datasets, which frequently have analogous imbalance challenges. Omar et al. [37] introduced a novel method for optimizing epileptic seizure recognition through deep learning models, utilizing feature scaling and dropout layers to improve performance.

Advanced machine learning methodologies have been utilized in numerous specialized medical domains. Hady et al. [38] used machine learning to predict abdominal fat composition following cavitation therapy, utilizing advanced hyperparameter optimization methods. Hady and Abd El-Hafeez [39] transformed core muscle analysis in female sexual dysfunction by machine learning, showcasing the applicability of these techniques in specific medical contexts.

Integrating deep learning models with medical imaging has advanced the frontiers of illness diagnosis and categorization. Eliwa et al. [40] suggested a method utilizing CNNs to classify monkeypox skin lesions, enhancing their model with the grey wolf optimizer algorithm. This study illustrates the capability of integrating sophisticated neural network topologies with evolutionary optimization methods, a notion that closely corresponds with our ERLNEIL-MDP framework.

Natural language processing techniques have demonstrated considerable utility in studying medical data. Hassan et al. [41] investigated transformer models and bidirectional long short-term memory networks for illness prediction based on symptom descriptions, underscoring the efficacy of language models in medical diagnosis. This study highlights the significance of managing unstructured textual data in healthcare, a difficulty our ERLNEIL-MDP method seeks to resolve.

Hady and Abd El-Hafeez [42] utilized machine learning to forecast female pelvic tilt and lumbar angle in instances of urine incontinence and sexual dysfunction. Their methodology of employing many scales rather than exclusively depending on imaging data illustrates the potential of machine learning in developing non-invasive diagnostic instruments.

In addition to conventional medical applications, machine learning has been utilized in associated domains that affect public health and safety. Shams et al. [43] introduced an innovative deep learning model that integrates a self-attention layer into a convolutional neural network for the detection of audio data in emergency vehicle sirens and road noise. This work demonstrates the adaptability of advanced machine learning algorithms in analyzing intricate sensory input, but it is not directly associated with medical diagnosis. This skill is particularly pertinent to medical signal processing.

These several studies demonstrate the rapidly expanding applicability of machine learning technology in the medical area. They elucidate the challenges of feature selection, data balance, model optimization, and the application of domain-specific knowledge to develop effective algorithms for medical data processing. The breadth of these enhancements inspires the creation of the ERLNEIL-MDP algorithm, which amalgamates elements of evolutionary strategies and RL to address the complexities inherent in medical datasets. The approach incorporates exploration strategies, including novelty-driven exploration and imitation learning, to enhance medical data analysis and eliminate previously established treatment boundaries, thereby facilitating improved diagnosis, targeted treatment, and optimized patient recovery.

3. Fundamental concepts

3.1. Reinforcement learning

RL is a specialized area in machine learning that specifically deals with teaching intelligent agents how to make a series of decisions in a given environment to maximize a cumulative reward. Within the RL framework, an agent engages with the environment by perceiving the present state, making decisions according to its policy, and receiving a reward and the subsequent state from the environment. The agent's objective is to acquire an optimal strategy that maximizes the anticipated total reward over time.

The RL problem is commonly expressed as a Markov decision process, represented by a tuple $(\mathscr{S},\mathscr{A},\mathscr{P},\mathscr{R},\gamma)$. Here, $\mathscr{S}$ denotes the state space, $\mathscr{A}$ represents the action space, $\mathscr{P}$ is the transition probability function, $\mathscr{R}$ stands for the reward function, and $\gamma \in [0,1]$ is the discount factor [44]. At each time step t, the agent perceives the current state $s_t \in \mathscr{S}$, selects an action $a_t \in \mathscr{A}$ based on its policy $\pi(a_t|s_t)$, and obtains a reward $r_t = \mathscr{R}(s_t,a_t)$ and the subsequent state $s_{t+1} \sim \mathscr{P}(\cdot|s_t,a_t)$.

The value function $V^{\pi}(s)$ denotes the anticipated total reward obtained by beginning from state s and adhering to policy π:

$$V^{\pi}(s) = \mathbb{E}\pi\left[\sum_{k=0}^{\infty}\gamma^{k}r_{t+k}|s_t = s\right] \tag{1}$$

where $\mathbb{E}_{\pi}$ represents the expectation calculated over the trajectories produced by policy π.

Similarly, the action-value function $Q^{\pi}(s,a)$ denotes the anticipated total reward when starting from state s, executing action a, and subsequently adhering to policy π:

$$Q^{\pi}(s,a) = \mathbb{E}\pi\left[\sum_{k=0}^{\infty}\gamma^{k}r_{t+k}|s_t = s, a_t = a\right] \tag{2}$$

The value function $V(s)$ and action-value function $Q(s,a)$ are defined as follows:

$$V^{*}(s) = \max_{\pi}V^{\pi}(s) \tag{3}$$

$$Q^{*}(s,a) = \max_{\pi}Q^{\pi}(s,a) \tag{4}$$

The objective of RL is to identify an optimal policy π^{*} that maximizes the predicted cumulative reward.

$$\pi^{*} = \mathrm{argmax}_{\pi}\mathbb{E}\pi\left[\sum_{t=0}^{\infty}\gamma^{t}r_t\right] \tag{5}$$

Several RL algorithms have been suggested to acquire the best policy. These algorithms estimate the value functions or explicitly optimize the policy using the observed rewards and state transitions. Nevertheless, conventional RL algorithms frequently encounter difficulties when dealing with state and action spaces with many dimensions, which is a regular occurrence in real-world applications like medical data processing. DRL integrates RL with deep neural networks (DNNs) to acquire intricate policies and value functions from unprocessed input data.

The deep Q-network (DQN) is a widely used DRL approach that estimates the action-value function by employing a deep neural network $Q_{\theta}(s,a)$, with θ denoting the network parameters [45]. The DQN is trained to minimize the error in the difference between predicted and actual values over time.

$$\mathscr{L}(\theta) = \mathbb{E}(s,a,r,s') \sim \mathscr{D}\left[\left(r + \gamma\max a' Q_{\theta'}(s',a') - Q_{\theta}(s,a)\right)^{2}\right] \tag{6}$$

where $\mathscr{D}$ refers to the replay buffer, which stores the agent's experiences as state-action-reward-next state tuples (s,a,r,s'). The parameter θ' represents the target network parameters, which are regularly changed to ensure the stability of the learning process.

Another notable DRL technique is proximal policy optimization (PPO) [46], an on-policy approach specifically designed to optimize the policy directly. PPO incorporates a surrogate goal function to restrict policy updates and avoid significant changes that could disrupt the learning process. The surrogate objective function is formally specified as:

$$\mathscr{L}^{CLIP}(\theta) = \mathbb{E}(s,a) \sim \pi\theta_{old}\left[\min\left(\frac{\pi_{\theta}(a|s)}{\pi_{\theta_{old}}(a|s)}A^{\pi_{\theta_{old}}}(s,a), \mathrm{clip}\left(\frac{\pi_{\theta}(a|s)}{\pi_{\theta_{old}}(a|s)}, 1-\epsilon, 1+\epsilon\right)A^{\pi_{\theta_{old}}}(s,a)\right)\right] \tag{7}$$

where π_{θ} refers to the current policy, $\pi_{\theta_{old}}$ refers to the previous policy, $A^{\pi_{\theta_{old}}}(s,a)$ represents the advantage function estimated using the previous policy, and ϵ is a hyperparameter that determines the range of clipping.

DRL algorithms have succeeded remarkably in diverse fields, including game-playing, robotics, and autonomous driving. Nevertheless, the use of machine learning algorithms in medical data processing tasks requires further enhancement due to the distinctive difficulties presented by medical data, including the limited availability of labeled data, the existence of noise and artifacts, and the want for solutions that are both interpretable and dependable. In order to tackle these difficulties, scientists have investigated the combination of EAs with DRL, resulting in the emergence of the discipline known as ERL. ERL utilizes the search and diversity maintenance skills of EAs to improve the exploration and resilience of DRL algorithms.

3.2. Imitation learning

Imitation learning (IL) is a machine learning approach that seeks to acquire a policy by emulating the actions of an expert or a collection of demonstrations [47]. Unlike RL, which relies on trial and error and a reward signal, imitation learning involves learning from an expert's activities. This approach reduces the need for lengthy exploration and speeds up learning. IL has proven effective in various applications, including robotics, autonomous driving, and game-playing.

In ERL and medical data processing, imitation learning can be crucial in guiding the learning process, providing a good starting point for the policy search, and incorporating domain knowledge from experts. By leveraging expert demonstrations, ERL algorithms can focus the search on promising regions of the policy space, reducing the computational cost and improving the sample efficiency of the learning process.

Moreover, imitation learning can be used to initialize the population of policies in ERL, providing a good starting point for the search process. By seeding the population with pre-trained policies using expert demonstrations, ERL can accelerate the learning process and focus the search on promising regions of the policy space. This is particularly useful in complex and high-dimensional environments, where random initialization may lead to policies far from the desired behavior.

In medical data processing, imitation learning can be used to learn from expert clinicians and incorporate their knowledge into decision-making. For example, in the task of sepsis treatment recommendation, an ERL algorithm can be initialized with a policy that mimics experienced clinicians' treatment decisions [48]. However, applying imitation learning to medical data processing tasks presents several challenges. First, expert demonstrations may be scarce or expensive, especially in critical care settings where clinicians need more time and resources. Second, expert decisions may be subject to bias or variability, requiring careful selection and preprocessing of the demonstration data. Third, the learned policies must be interpretable and reliable, as they can have significant implications for patient safety and well-being.

Researchers have proposed various techniques to address these challenges, such as active learning to efficiently collect expert demonstrations, domain adaptation to transfer knowledge from related tasks or populations, and interpretable machine learning to explain the learned policies. By incorporating these techniques into the ERL framework, imitation learning can be vital in guiding the learning process, incorporating domain knowledge, and improving the interpretability and reliability of the learned policies in medical data processing tasks.

3.3. Evolutionary reinforcement learning

The ERL method is a synergistic combination of EAs and RL, effectively addressing intricate sequential decision-making challenges. ERL algorithms combine the population-based search abilities of EAs with

RL's temporal credit assignment and value function approximation techniques. This integration allows ERL algorithms to efficiently explore state-action spaces that are large and have many dimensions, handle rewards that are sparse and delayed, and adapt to environments that change over time [49].

The main idea behind ERL is to use an EA to evolve a population of policies or value functions while using RL techniques to evaluate and update the individuals based on their performance in the environment. The evolutionary process operates at a higher level, searching for the optimal policy or value function parameters, while the RL process operates at a lower level, fine-tuning the parameters based on the agent's interactions with the environment.

One of the key advantages of ERL is its ability to maintain a diverse population of policies, which can help prevent premature convergence to suboptimal solutions and promote exploration of the policy space. By maintaining a pool of policies with different behaviors and characteristics, ERL algorithms can effectively balance the exploration-exploitation trade-off and adapt to environmental changes.

Another advantage of ERL is its flexibility in defining the fitness function used to evaluate the policies. In contrast to traditional RL algorithms, which typically rely on a single scalar reward signal, ERL algorithms can incorporate multiple objectives and constraints into the fitness function, such as the agent's performance, sample efficiency, and robustness. This multi-objective optimization capability allows ERL to find policies that satisfy multiple criteria and can be particularly useful in real-world applications, such as medical data processing, where the decision-making process must balance various factors, such as patient outcomes, resource utilization, and treatment costs.

ERL algorithms can be divided into two main categories: (1) algorithms that utilize EAs to optimize the parameters of a policy or value function with a fixed structure, and (2) algorithms that employ EAs to simultaneously develop both the structure and parameters of the policy or value function. ERL has demonstrated encouraging outcomes in many medical data processing applications, including dynamic therapy regimes, patient monitoring, and clinical trial design. ERL algorithms leverage the exploratory skills of EAs and the sequential decision-making powers of RL to uncover tailored treatment methods that enhance patient outcomes while minimizing adverse effects and expenses.

4. The proposed algorithm

This section presents the ERL algorithm with novelty-driven exploration and imitation learning for medical data processing (ERLNEIL-MDP). The key components of the algorithm include the architecture, novelty computation, combination selection strategy, experience fusion imitation, and training stability module.

The novelty computation strategy is driven by the need to thoroughly explore the complex policy space inherent in medical data processing tasks. In these domains, the optimal policy is often not immediately apparent due to the intricacy and high dimensionality of the data. The strategy quantifies the uniqueness of each policy by comparing it to the current population and historical policies. This approach encourages the discovery of innovative solutions by rewarding policies that exhibit distinct behaviors, thereby preventing premature convergence to suboptimal solutions and maintaining a diverse set of policies throughout the evolutionary process.

The combination selection strategy addresses the crucial balance between exploration and exploitation in the algorithm. It is motivated by the need to efficiently navigate the vast search space of potential policies while still focusing computational resources on promising areas. This strategy adaptively weighs the importance of novelty and fitness when selecting policies for reproduction, allowing the algorithm to shift its focus between exploring new solutions and refining existing ones as the search progresses.

The valuable role of expert knowledge in medical domains inspires experience fusion imitation. This strategy aims to accelerate learning and improve the quality of solutions by incorporating domain expertise into the evolutionary process. It combines parent policies' experiences with expert demonstrations, allowing offspring policies to benefit from evolutionary discoveries and established medical knowledge.

The training stability module is motivated by the need for reliable and consistent performance in medical applications, where stability and reproducibility are paramount. This module dynamically adjusts key algorithm parameters and preserves elite policies across generations. It aims to maintain a stable learning process while allowing for necessary adaptations to medical data's complex and often noisy nature.

Together, these strategies form a comprehensive approach to ERL in medical data processing. They work in concert to promote exploration, leverage domain knowledge, maintain diversity, and ensure stability - all crucial aspects for developing effective and reliable algorithms in the medical field. The synergy between these components allows ERLNEIL-MDP to navigate the unique challenges posed by medical data, potentially leading to more accurate diagnoses, personalized treatment plans, and improved patient outcomes.

4.1. Architecture of the ERLNEIL-MDP

The architecture of the ERLNEIL-MDP algorithm is designed to effectively address the challenges associated with processing complex and heterogeneous medical data while leveraging the strengths of EAs and RL. As depicted in Fig. 1, the ERLNEIL-MDP algorithm consists of two main components: the EA and RL modules.

The EA module maintains a population of policies, each represented by DNN. These policies are evolved through selection, crossover, and mutation operators, guided by their novelty scores and fitness values. The novelty scores are calculated using the individual policy novelty (IPN) and population diversity novelty (PDN) methods, which quantify the dissimilarity between a policy and the population, as well as the historical data. The fitness values are determined by evaluating the policies' performance on the medical data processing task, such as the accuracy of disease diagnosis or the effectiveness of treatment recommendations.

The RL module employs a modified version of the PPO algorithm to train the policies based on their interactions with the medical data environment. Each policy processes medical data instances, such as EHRs or medical images, and generates corresponding actions or predictions. The environment provides feedback through rewards, which are used to update the policies' parameters. The experiences generated by the policies are stored in a shared experience replay buffer, which is used to stabilize the training process and improve sample efficiency.

The ERLNEIL-MDP algorithm introduces several key components to enhance the performance and adaptability of the ERL process in the context of medical data processing:

1. **Novelty computation:** The novelty computation component quantifies the novelty of each policy based on its dissimilarity to the population and the historical data, using the IPN and PDN methods. The combined novelty score promotes exploration and maintains diversity in the population.
2. **Adaptive novelty-fitness selection strategy:** The adaptive novelty-fitness selection strategy balances the exploration and exploitation trade-off by considering policies' novelty and fitness during the selection process. The relative importance of novelty and fitness is adjusted based on the progress of the evolutionary process.
3. **Imitation-guided experience fusion:** The imitation-guided experience fusion mechanism incorporates expert knowledge and demonstrations into the offspring policies' learning process. The experiences of the parent policies are fused with expert demonstrations to create a rich and diverse set of experiences for the offspring policies to learn from.

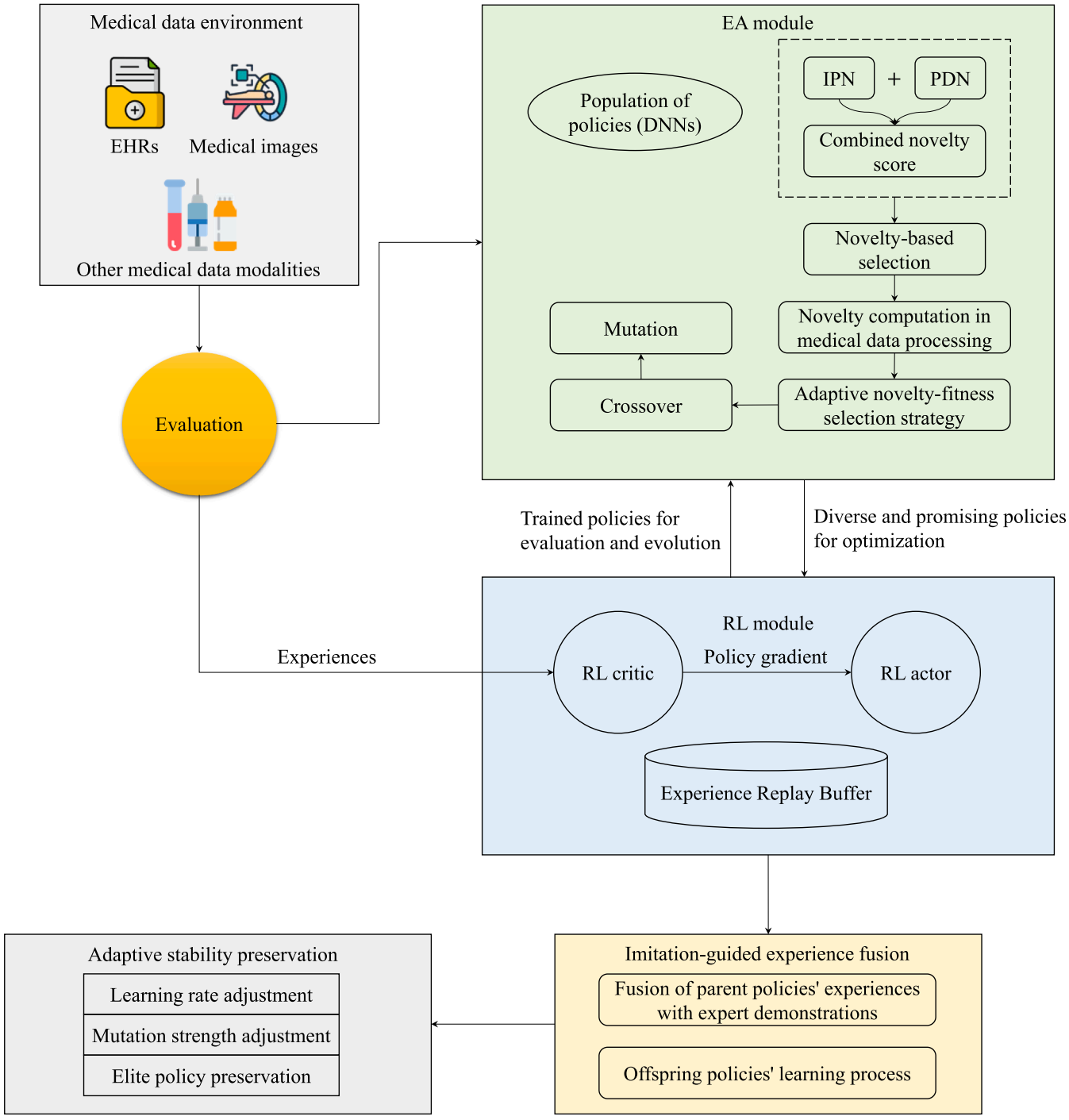


Fig. 1. Architecture of the ERLNEIL-MDP algorithm.

4. **Adaptive stability preservation:** The adaptive stability preservation module ensures the stability of the learning process by dynamically adjusting the algorithm's hyperparameters, such as the learning rate and mutation strength, based on the progress and stability of the learning process. The elite policy preservation strategy maintains the best-performing policies across generations, preventing the loss of valuable knowledge.

The EA and RL modules exchange information in a bi-directional manner. The EA module provides diverse and promising policies to the RL module for further optimization, while the RL module sends the trained policies back to the EA module for evaluation and evolution. This iterative process continues until a satisfactory policy is found or a maximum number of generations is reached.

The architecture of the ERLNEIL-MDP algorithm is designed to be modular and flexible, allowing for easy integration of different medical data modalities and adaptation to various medical data processing tasks. By combining the strengths of EAs and RL and incorporating novelty-driven exploration, imitation learning, and stability preservation mechanisms, the ERLNEIL-MDP algorithm is well-equipped to handle the complexities and challenges of medical data processing, ultimately leading to more accurate, reliable, and efficient decision support systems in healthcare.

The ERLNEIL-MDP algorithm is designed to handle various types of

structured and semi-structured medical data, including:

- EHRs: containing patient demographics, diagnoses, medications, lab results, and treatment outcomes.
- Medical imaging data, such as X-rays, CT scans, and MRI images, are represented by pixel or voxel values.
- Time-series data: including vital signs monitoring, ECG readings, and continuous glucose monitoring data.
- Genomic data: such as gene expression profiles and single nucleotide polymorphisms.
- Clinical notes: preprocessed and converted into a structured format using natural language processing techniques.

The algorithm's flexibility in handling these diverse data types stems from its neural network-based policy representation, which can be adapted to different input formats, and its novelty computation mechanism, which can be customized to capture domain-specific notions of novelty in medical data.

The ERLNEIL-MDP algorithm has been specifically designed to address the unique challenges medical data processing presents. It tackles the data sparsity issue by using an experience replay buffer and imitation learning component. This approach allows the algorithm to use limited data points efficiently by reusing available information and incorporating expert knowledge, which is particularly beneficial when working with sparse electronic health record datasets.

To address the high dimensionality often encountered in medical data, ERLNEIL-MDP employs a novelty computation mechanism. This feature encourages the exploration of diverse solutions by identifying and promoting unique combinations of features that traditional methods might overlook. This is especially valuable in high-dimensional medical imaging tasks, where important patterns may be hidden in complex feature interactions. The algorithm's adaptive selection strategy complements this by focusing computational resources on the most promising regions of the solution space.

While interpretability can be challenging with deep learning components, ERLNEIL-MDP mitigates this issue through its evolutionary process. By tracking the lineage of successful policies, the algorithm provides insights into how effective strategies develop over time. This feature allows for transparency in the decision-making process, which is crucial in medical applications where understanding the reasoning behind a recommendation is often as important as the recommendation itself.

For instance, when applied to a sparse electronic health record dataset, ERLNEIL-MDP can leverage its experience replay buffer to extract maximum value from the limited available data. In high-dimensional medical imaging tasks, the novelty computation mechanism can identify unique feature combinations that might be critical for accurate diagnosis but could be noticed by more sophisticated analysis methods. These capabilities make ERLNEIL-MDP particularly well-suited to the complexities of medical data processing, offering a robust approach to handling the sparsity, high dimensionality, and interpretability challenges inherent in this field.

4.2. Novelty computation in ERLNEIL-MDP

Novelty computation plays a crucial role in the ERLNEIL-MDP algorithm, as it promotes exploration and maintains diversity in the population of policies. By encouraging the discovery of novel solutions, the algorithm can effectively navigate medical data processing tasks' complex and high-dimensional search space. This subsection introduces two novelty computation methods: the IPN and the PDN.

4.2.1. IPN

The IPN quantifies the novelty of a single policy based on its dissimilarity to the historical data encountered during the learning process. The IPN score of a policy π_i is calculated as the average dissimilarity between the actions taken by π_i and the actions stored in the experience replay buffer $\mathscr{B}$:

$$\mathrm{IPN}(\pi_i) = \frac{1}{|\mathscr{B}|} \sum_{(s_t, a_t) \in \mathscr{B}} d(\pi_i(s_t), a_t) \tag{8}$$

where s_t and a_t are the state and action pairs sampled from the experience replay buffer $\mathscr{B}$, and $d(\cdot, \cdot)$ is a dissimilarity measure, e.g., the Euclidean distance or the Kullback-Leibler divergence, depending on the nature of the action space (e.g., continuous or discrete).

The IPN score encourages policies to explore actions different from those encountered in the past, promoting the discovery of novel treatment strategies or decision-making patterns in medical data processing tasks.

Algorithm 1 outlines the process of computing the IPN scores for a population of policies:

The historical experience replay buffer is a crucial component of the ERLNEIL-MDP algorithm, serving as a repository for past interactions between policies and the environment. It stores state-action-reward-next state tuples, enabling off-policy learning and improving sample efficiency by allowing the algorithm to learn from a diverse set of past experiences. This buffer plays a key role in computing the IPN by providing a reference point for comparing current policy actions with historical data. Additionally, it enhances learning stability by breaking temporal correlations in the training data and facilitates knowledge transfer among different policies in the population through experience sharing. These functions collectively contribute to the algorithm's ability to efficiently explore the solution space and adapt to complex medical data processing tasks.

4.2.2. PDN

The PDN quantifies the novelty of a policy based on its dissimilarity to the other policies in the current population. The PDN score of a policy π_i is calculated as the average dissimilarity between the actions taken by π_i and the actions taken by the other policies in the population Π:

$$\mathrm{PDN}(\pi_i) = \frac{1}{|\Pi| - 1} \sum_{\pi_j \in \Pi, j \neq i} \frac{1}{|S|} \sum_{s_t \in S} d\big(\pi_i(s_t), \pi_j(s_t)\big) \tag{9}$$

where S is a set of states sampled from the medical data environment, and $d(\cdot, \cdot)$ is a dissimilarity measure, as defined in the IPN computation.

The PDN score encourages policies to explore different regions of the action space compared to the other policies in the population, maintaining diversity and preventing premature convergence to suboptimal solutions.

Algorithm 2 outlines the process of computing the PDN scores for a population of policies:

4.2.3. Combined novelty score

To leverage the advantages of both the IPN and the PDN, the ERLNEIL-MDP algorithm computes a combined novelty score for each policy. The combined novelty score $\mathrm{CN}(\pi_i)$ of a policy π_i is calculated as the weighted sum of its IPN and PDN scores:

$$\mathrm{CN}(\pi_i) = \alpha \cdot \mathrm{IPN}(\pi_i) + (1 - \alpha) \cdot \mathrm{PDN}(\pi_i) \tag{10}$$

where $\alpha \in [0, 1]$ is a hyperparameter that controls the relative importance of the IPN and PDN scores. A higher value of α places more emphasis on the novelty of the policy concerning the historical data, while a lower value of α prioritizes the novelty of the policy for the current population.

The combined novelty score provides a comprehensive measure of a policy's novelty, considering its exploration of new actions and diversity within the population. This balanced approach helps the ERLNEIL-MDP algorithm maintain a healthy level of exploration while avoiding excessive divergence from promising solutions.

Algorithm 1
IPN computation.

Input: Population of policies $\Pi = \pi_1, \pi_2, \ldots, \pi_N$; Historical experience replay buffer $\mathscr{B}$; Dissimilarity measure $d(\cdot,\cdot)$
Output: IPN scores for each policy $IPN(\pi_1), IPN(\pi_2), \ldots, IPN(\pi_N)$
Initialize an empty list IPN_{scores} to store the novelty scores
  For each policy π_i in the population Π:
    Initialize a variable $novelty_{sum}$ to 0
      For each state-action pair (s_t, a_t) in the historical experience replay buffer $\mathscr{B}$:
        Compute the dissimilarity $d(\pi_i(s_t), a_t)$ between the action taken by the policy π_i in state s_t and the action a_t stored in the buffer
      Add the dissimilarity to $novelty_{sum}$
    End for
  Compute the IPN score for policy π_i as Eq. (8)
  Append $IPN(\pi_i)$ to the IPN_{scores} list
End for
Return the IPN_{scores} list containing the IPN scores for each policy in the population

Algorithm 2
PDN computation.

Input: Population of policies $\Pi = \pi_1, \pi_2, \ldots, \pi_N$; Set of states $S = s_1, s_2, \ldots, s_M$ sampled from the medical data environment; Dissimilarity measure $d(\cdot,\cdot)$
Output: PDN scores for each policy $PDN(\pi_1), PDN(\pi_2), \ldots, PDN(\pi_N)$
Initialize an empty matrix D of size $N \times N$ to store pairwise dissimilarities between policies
  For each pair of policies (π_i, π_j) in the population Π:
    Initialize a variable $dissimilarity_{sum}$ to 0
      For each state s_t in the sampled set of states S:
        Compute the dissimilarity $d(\pi_i(s_t), \pi_j(s_t))$ between the actions taken by policies π_i and π_j in state s_t
      Add the dissimilarity to $dissimilarity_{sum}$
    End for
  End for
For each policy π_i in the population Π:
  Compute the PDN score for policy π_i using Eq. (9)
  Append $PDN(\pi_i)$ to the PDN_{scores} list
End for
Return the PDN_{scores} list containing the PDN scores for each policy in the population

4.2.4. Novelty-based selection

The combined novelty scores are used in the selection process of the EA to promote the survival and reproduction of novel policies. In the ERLNEIL-MDP algorithm, a novelty-based tournament selection is employed, where policies compete against each other based on their combined novelty scores. The selection process is performed as follows:

- Randomly sample a subset of policies from the population.
- Select the policy with the highest combined novelty score from the subset.
- Repeat steps 1 and 2 until the desired number of policies is selected.

The novelty-based tournament selection ensures that policies with higher novelty scores are more likely to be selected for reproduction, thus promoting the exploration of novel solutions in the evolutionary process.

Furthermore, the ERLNEIL-MDP algorithm incorporates a novelty-based elitism strategy, where a fixed number of policies with the highest combined novelty scores are directly preserved in the next generation without undergoing crossover or mutation. This elitism strategy helps maintain the most novel solutions throughout the evolutionary process and prevents the loss of valuable information.

4.2.5. Novelty computation in medical data processing

The novelty computation methods described above in medical data processing can be applied to various tasks, such as disease diagnosis, treatment recommendation, or patient stratification. The dissimilarity measure $d(\cdot,\cdot)$ can be adapted to the specific nature of the medical data and the action space of the policies.

For example, in a disease diagnosis task, the actions of a policy may represent the predicted disease labels for a given set of patient features. In this case, the dissimilarity measure can be based on the Hamming distance between the predicted labels and the ground truth labels stored in the experience replay buffer or the predicted labels of other policies in the population.

In a treatment recommendation task, the actions of a policy may represent the recommended treatment options for a given patient state. The dissimilarity measure can be based on the Euclidean distance between the continuous treatment parameters (e.g., medication dosage) or the Jaccard distance between the discrete treatment options (e.g., surgery, chemotherapy, or radiation therapy).

By capturing the novelty of policies in the context of medical data processing tasks, the ERLNEIL-MDP algorithm can effectively explore new diagnostic strategies, treatment recommendations, or patient stratification criteria, potentially leading to improved patient outcomes and more efficient healthcare decision-making processes.

In conclusion, the novelty computation methods introduced in this section, namely the IPN and the PDN, play a vital role in the ERLNEIL-MDP algorithm by promoting exploration and maintaining diversity in the population of policies. The combined novelty score provides a comprehensive measure of a policy's novelty, considering its exploration of new actions and diversity within the population. The novelty-based selection and elitism strategies ensure that novel policies are prioritized and preserved throughout the evolutionary process. This enables the algorithm to effectively navigate the complex search space of medical data processing tasks and discover innovative solutions.

4.3. Adaptive novelty-fitness selection strategy in ERLNEIL-MDP

The selection of policies for reproduction is a crucial step in the evolutionary process of the ERLNEIL-MDP algorithm. It determines which policies will serve as parents for the next generation, consequently influencing the algorithm's ability to explore novel solutions and exploit promising ones. In this section, we introduce the adaptive novelty-fitness selection strategy (ANFSS), which balances the exploration and exploitation trade-off by considering the novelty and fitness of policies during the selection process.

The ANFSS combines the novelty scores, computed using the IPN and

PDN methods, with the fitness values of the policies to create a comprehensive selection criterion. The fitness value of a policy π_i is determined by its performance on the medical data processing task, such as the accuracy of disease diagnosis, the effectiveness of treatment recommendations, or the quality of patient stratification.

The adaptive novelty-fitness selection strategy in ERLNEIL-MDP is designed to balance exploration and exploitation dynamically throughout the learning process. This balance is crucial for effectively navigating the complex landscape of medical data processing tasks.

To ensure that the algorithm maintains a balance between exploration and exploitation, it continuously monitors two key metrics: population diversity and the average fitness improvement rate. If the population diversity drops below a predetermined threshold or the fitness improvement rate stagnates, the strategy temporarily increases the weighting towards novelty. This adaptive mechanism helps prevent premature convergence to suboptimal solutions and allows the algorithm to escape local optima when necessary.

By dynamically adjusting the balance between novelty and fitness, this strategy enables ERLNEIL-MDP to efficiently explore the vast solution space characteristic of medical data processing tasks while also focusing computational resources on refining the most promising solutions. This approach is particularly valuable in medical applications where innovative solutions and optimization of existing strategies are crucial for improving patient outcomes.

To balance the importance of novelty and fitness, the ANFSS employs an adaptive weighting scheme that adjusts the relative contributions of the novelty and fitness components based on the progress of the evolutionary process. The combined selection score $\mathrm{CSS}(\pi_i)$ of a policy π_i is calculated as follows:

$$\mathrm{CSS}(\pi_i) = \beta_t \cdot \widehat{\mathrm{CN}}(\pi_i) + (1-\beta_t)\cdot \widehat{f}(\pi_i) \tag{11}$$

where $\widehat{\mathrm{CN}}(\pi_i)$ is the normalized combined novelty score of policy π_i, $\widehat{f}(\pi_i)$ is the normalized fitness value of policy π_i.

The normalized combined novelty score $\widehat{\mathrm{CN}}(\pi_i)$ and the normalized fitness value $\widehat{f}(\pi_i)$ are obtained using min-max normalization:

$$\widehat{\mathrm{CN}}(\pi_i) = \frac{\mathrm{CN}(\pi_i) - \mathrm{CN}_{\min}}{\mathrm{CN}_{\max} - \mathrm{CN}_{\min}} \tag{12}$$

$$\widehat{f}(\pi_i) = \frac{f(\pi_i) - f_{\min}}{f_{\max} - f_{\min}} \tag{13}$$

where $\mathrm{CN}_{\min}$ and $\mathrm{CN}_{\max}$ represent the lowest and highest combined novelty scores in the population, whereas $f_{\min}$ and $f_{\max}$ represent the lowest and highest fitness values in the population, respectively.

The adaptive weighting factor β_t is updated at each generation based on the progress of the evolutionary process. In the early stages of evolution, a higher weight is assigned to the novelty component to encourage exploration, while in the later stages, a higher weight is assigned to the fitness component to focus on exploiting the most promising solutions. The adaptive weighting factor is calculated using a simple linear decay function:

$$\beta_t = \beta_0 - (\beta_0 - \beta_T)\cdot\frac{t}{T} \tag{14}$$

where β_0 represents the initial weighting factor, β_T represents the ultimate weighting factor, t represents the current generation, and T represents the total number of generations.

The ANFSS selects policies for reproduction using a tournament selection mechanism based on the combined selection scores. The tournament selection process is performed as follows:

- Randomly sample a subset of policies from the population.
- Select the policy with the highest combined selection score from the subset.
- Repeat steps 1 and 2 until the desired number of parent policies is selected.

The policies selected through the ANFSS are then subjected to crossover and mutation operations to create the next generation of policies. The adaptive weighting scheme ensures that the algorithm maintains a balance between exploring novel solutions and exploiting the most promising ones throughout the evolutionary process.

Algorithm 3 outlines the adaptive novelty-fitness selection strategy:

In medical data processing, the ANFSS helps the ERLNEIL-MDP algorithm strike a balance between discovering novel diagnostic strategies, treatment recommendations, or patient stratification criteria and refining the most effective ones. By adopting the importance of novelty and fitness during the selection process, the algorithm can effectively explore the vast search space of medical data processing tasks while converging towards high-performing solutions.

The fitness values used in the ANFSS can be tailored to the specific medical data processing task. For example, in a disease diagnosis task, the fitness value may represent the accuracy of the predicted disease labels compared to the ground truth labels. In a treatment recommendation task, the fitness value may represent the expected patient outcomes or the adherence to clinical guidelines. In a patient stratification task, the fitness value may represent the quality of the identified patient subgroups in terms of their clinical relevance and statistical significance.

By adopting the novelty scores and fitness values, the ANFSS enables the ERLNEIL-MDP algorithm to effectively navigate the complex landscape of medical data processing tasks, discover innovative solutions that improve patient care, and optimize healthcare decision-making processes.

4.4. Imitation-guided experience fusion in ERLNEIL-MDP

In this subsection, we introduce the imitation-guided experience fusion (IGEF) mechanism, which combines the experiences of the parent

Algorithm 3
ANFSS.

Input: Population of policies $\Pi = \pi_1, \pi_2, \ldots, \pi_N$; IPN scores for each policy $IPN(\pi_1), IPN(\pi_2), \ldots, IPN(\pi_N)$; PDN scores for each policy $PDN(\pi_1), PDN(\pi_2), \ldots, PDN(\pi_N)$; Fitness values for each policy $f(\pi_1), f(\pi_2), \ldots, f(\pi_N)$; Initial novelty weight β_0; Final novelty weight β_T; Current generation number t; Total number of generations T
Output: Selected parent policies for reproduction

```
Initialize an empty list combined_scores to store the combined novelty-fitness scores
Compute the minimum and maximum values of IPN, PDN, and fitness scores:
Compute the adaptive novelty weight for the current generation using Eq. (14)
For each policy π_i in the population Π:
    Normalize the IPN, PDN, and fitness scores:
    Compute the combined novelty score
    Compute the combined novelty-fitness score
    Append combined_score(π_i) to the combined_scores list
End for
Select the parent policies for reproduction using tournament selection based on the combined_scores
Return the selected parent policies
```

policies with expert demonstrations. The IGEF mechanism operates during the crossover step of the evolutionary process, where the selected parent policies generate offspring policies by exchanging and combining their genetic information. In the context of the ERLNEIL-MDP algorithm, the genetic information of a policy consists of its neural network parameters and the experiences stored in its individual experience replay buffer.

To incorporate expert knowledge into the offspring policies, the IGEF mechanism introduces an expert experience replay buffer $\mathscr{B}_E$, which contains a collection of expert demonstrations for the medical data processing task at hand. These demonstrations can be obtained from various sources, such as electronic health records, clinical guidelines, or domain experts, and they represent optimal or near-optimal solutions to the task.

The imitation-guided experience fusion mechanism in ERLNEIL-MDP is designed to effectively incorporate expert knowledge while maintaining the algorithm's ability to discover novel and potentially superior solutions. This mechanism employs a multi-faceted approach to ensure that expert knowledge remains relevant and beneficial throughout the evolutionary process.

A key feature of this mechanism is the adaptive imitation weight. As the algorithm progresses, the influence of expert demonstrations is gradually reduced. This allows the algorithm to emphasize novel solutions discovered through the evolutionary process. By dynamically adjusting the balance between expert knowledge and evolved strategies, the algorithm can leverage expert insights in the early stages while still having the freedom to explore and refine potentially superior approaches as learning progresses.

The mechanism also incorporates selective imitation. Rather than unthinkingly incorporating all expert knowledge, the algorithm continuously compares the performance of expert demonstrations with that of evolved policies. Expert knowledge is only integrated when it demonstrably outperforms the current population. This selective approach ensures that only truly beneficial expert insights are incorporated, preventing the algorithm from being constrained by potentially outdated or suboptimal expert strategies.

The mechanism employs periodic re-evaluation to maintain the relevance of expert knowledge further. Expert demonstrations are regularly reassessed against the current best policies in the population. This ongoing evaluation ensures that the incorporated expert knowledge remains valuable as the algorithm's performance improves. If previously useful expert demonstrations become less effective than evolved strategies, their influence can be reduced or eliminated appropriately.

Finally, the mechanism preserves diversity through the selective application of imitation. When incorporating expert knowledge, the algorithm applies imitation selectively to only a subset of the population. This approach maintains diverse policies, some closely aligned with expert knowledge and others more explorative. By preserving this diversity, the algorithm retains its ability to explore novel solutions while still benefiting from expert insights.

Through these features, the imitation-guided experience fusion mechanism strikes a balance between leveraging valuable expert knowledge and fostering the discovery of innovative solutions. This balance is particularly crucial in the medical domain, where established expert knowledge is invaluable but where there is also significant potential for discovering novel, data-driven approaches to improve patient care and outcomes.

During the crossover step, the IGEF mechanism creates an offspring policy π_o by combining the experiences of the parent policies π_p and π_q with the expert demonstrations from $\mathscr{B}_E$. The fusion of experiences is performed using a weighted averaging approach, where the weights determine the relative importance of the parent experiences and the expert demonstrations. The fused experience e_o for the offspring policy π_o is calculated as follows:

$$e_o = \lambda_1 \cdot e_p + \lambda_2 \cdot e_q + (1 - \lambda_1 - \lambda_2) \cdot e_E \tag{15}$$

where e_p and e_q are the experiences of the parent policies π_p and π_q, respectively, e_E is an expert demonstration sampled from $\mathscr{B}_E$, and λ_1, $\lambda_2 \in [0, 1]$ are weighting factors that control the contributions of the parent experiences and the expert demonstration, with $\lambda_1 + \lambda_2 \leq 1$.

To ensure that the offspring policy π_o effectively learns from the fused experiences, the IGEF mechanism employs an imitation learning objective that minimizes the difference between the actions that π_o and those suggested by the fused experiences. The imitation learning objective $\mathscr{L}_{IL}(\theta_o)$ for the offspring policy π_o with parameters θ_o is defined as:

$$\mathscr{L}_{IL}(\theta_o) = \frac{1}{|\mathscr{B}o|} \sum (s_t, a_t) \in \mathscr{B}_o \ell(\pi_o(s_t), a_t) \tag{16}$$

where s_t and a_t represent the state and action pairs randomly selected from the offspring's experience replay buffer, the symbol $\mathscr{B}_o$ represents a set of possible actions. The function $\ell(\cdot, \cdot)$ is a loss function that quantifies the discrepancy between the predicted action $\pi_o(s_t)$ and the desired action a_t. The specific loss function used, such as mean squared error or cross-entropy loss, depends on the characteristics of the action space, such as whether it is continuous or discrete.

The imitation learning objective is combined with the RL objective $\mathscr{L}_{RL}(\theta_o)$ to form the overall learning objective for the offspring policy π_o:

$$\mathscr{L}(\theta_o) = \mathscr{L}_{RL}(\theta_o) + \alpha \cdot \mathscr{L}_{IL}(\theta_o) \tag{17}$$

where $\alpha \in [0, 1]$ is a hyperparameter that controls the relative importance of the imitation learning objective compared to the RL objective.

Algorithm 4 outlines the imitation-guided experience fusion mechanism:

In medical data processing, the IGEF mechanism enables the ERLNEIL-MDP algorithm to incorporate domain knowledge and expert insights into learning effectively. For example, in a disease diagnosis task, the expert demonstrations can include clinical features and the corresponding disease labels provided by experienced physicians. By fusing these demonstrations with the experiences of the parent policies, the offspring policies can learn to emulate the diagnostic reasoning of the experts while also exploring new diagnostic strategies based on their parents' experiences.

Similarly, in a treatment recommendation task, the expert demonstrations can include patient characteristics, treatment options, and the corresponding patient outcomes obtained from clinical trials or real-world evidence. The IGEF mechanism allows the offspring policies to combine the treatment strategies learned by their parents with the best practices suggested by the expert demonstrations, leading to more effective and personalized treatment recommendations.

The weighting factors λ_1 and λ_2 in the IGEF mechanism can be adjusted based on the quality and relevance of the expert demonstrations for the specific medical data processing task. Higher weights can be assigned to the expert demonstrations if they are known to be highly reliable and relevant to the task, while lower weights can be assigned if the demonstrations are noisy or less applicable to the current problem setting.

Furthermore, the IGEF mechanism can be extended to incorporate multiple sources of expert knowledge, such as clinical guidelines, medical literature, and patient feedback, by maintaining separate expert experience replay buffers for each source and combining their demonstrations using appropriate weighting schemes. This multi-source imitation learning approach can help the ERLNEIL-MDP algorithm leverage a rich and diverse set of domain knowledge to guide the learning process and discover effective solutions for complex medical data processing tasks.

4.5. Adaptive stability preservation in ERLNEIL-MDP

Ensuring the stability of the training process is crucial for the success

Algorithm 4
Imitation-guided experience fusion (IGEF).

Input: Population of policies $\Pi = \pi_1, \pi_2, \ldots, \pi_N$; Expert experience replay buffer $\mathscr{B}_E$; Weighting factors λ_1 and λ_2; Offspring experience replay buffer size N_o
Output: Offspring policy π_o with fused experiences
Initialize an empty experience replay buffer $\mathscr{B}_o$ for the offspring policy π_o
While $|\mathscr{B}_o| < N_o$:
  Sample an experience e_p from the parent policy π_p's experience replay buffer
  Sample an experience e_q from the parent policy π_q's experience replay buffer
  Sample an expert demonstration e_E from the expert experience replay buffer $\mathscr{B}_E$
  Compute the fused experience e_o using the weighting factors λ_1 and λ_2: $e_o = \lambda_1 \cdot e_p + \lambda_2 \cdot e_q + (1 - \lambda_1 - \lambda_2) \cdot e_E$
  Add the fused experience e_o to the offspring's experience replay buffer $\mathscr{B}_o$
End while
Initialize the offspring policy π_o with the same architecture as the parent policies
Train the offspring policy π_o using the fused experiences in $\mathscr{B}o$ and the imitation learning objective: $\mathscr{L}IL(\theta_o) = \frac{1}{|\mathscr{B}o|} \sum (s_t, a_t) \in \mathscr{B}_o \ell(\pi_o(s_t), a_t)$
Fine-tune the offspring policy π_o using the RL objective: $\mathscr{L}(\theta_o) = \mathscr{L}RL(\theta_o) + \alpha \cdot \mathscr{L}IL(\theta_o)$
  Return the trained offspring policy π_o

of the ERLNEIL-MDP algorithm in tackling complex medical data processing tasks. Instability in the learning process can lead to suboptimal solutions, slow convergence, and even divergence, hindering the algorithm's ability to discover effective strategies for diagnosis, treatment recommendation, or patient stratification. In this section, we introduce the adaptive stability preservation (ASP) module, which aims to maintain the stability of the ERL process by dynamically adjusting the algorithm's hyperparameters and preserving the elite policies across generations.

The ASP module consists of two main components: (1) the adaptive hyperparameter adjustment (AHA) mechanism and (2) the elite policy preservation (EPP) strategy.

4.5.1. AHA

The AHA mechanism dynamically adjusts the key hyperparameters of the ERLNEIL-MDP algorithm based on the progress and stability of the learning process. The main hyperparameters considered by the AHA mechanism include the learning rate, the mutation rate, and the imitation learning weight.

The learning rate, denoted as α, governs the magnitude of the policy updates made throughout the RL process. An elevated learning rate can expedite the convergence process but may introduce instability, whereas a reduced learning rate might lead to delayed convergence but enhanced stability. The AHA mechanism adjusts the learning rate by considering the rate of improvement in the population's average fitness across a sliding window of w generations. If the rate of improvement falls below a specified threshold δ, the learning rate is reduced by a factor of γ_α, which leads to more consistent updates. On the other hand, if the rate of progress exceeds the threshold, the learning rate is augmented by $1/\gamma_\alpha$ to expedite convergence.

The mutation rate μ controls the probability of applying random perturbations to the offspring policies during the evolutionary process. A high mutation rate promotes exploration but may disrupt the stability of the learning process, while a low mutation rate encourages exploitation but may limit the algorithm's ability to escape local optima. The AHA mechanism adjusts the mutation rate based on the diversity of the population, measured by the average pairwise distance between the policies. If the diversity falls below a threshold δ_μ, the mutation rate is increased by γ_μ to introduce more variability into the population. Conversely, if the diversity exceeds the threshold, the mutation rate is decreased by $1/\gamma_\mu$ to maintain stability.

The imitation learning weight α_{IL} determines the relative importance of the imitation learning objective compared to the RL objective during the training of the offspring policies. A high imitation learning weight encourages the policies to closely follow the expert demonstrations, while a low weight allows for more exploration and adaptation. The AHA mechanism adjusts the imitation learning weight based on the similarity between the actions taken by the offspring policies and the expert demonstrations. If the average similarity falls below a threshold δ_{IL}, the imitation learning weight is increased by γ_{IL} to promote closer adherence to the expert knowledge. Conversely, if the similarity exceeds the threshold, the imitation learning weight is decreased by $1/\gamma_{IL}$ to allow for more flexibility in the learning process.

By dynamically adjusting these hyperparameters based on the progress and stability of the learning process, the AHA mechanism helps maintain a balance between exploration and exploitation, as well as between the influence of expert knowledge and the adaptation to the specific medical data processing task.

4.5.2. EPP

The EPP strategy aims to preserve the best-performing policies across generations, preventing the loss of valuable knowledge due to the stochastic nature of the evolutionary process. In medical data processing, the elite policies represent the most effective strategies for diagnosis, treatment recommendation, or patient stratification discovered by the algorithm at each generation.

The EPP strategy maintains an elite pool $\mathscr{E}$ of size N_e, which contains the top-performing policies based on their fitness values. After each generation's evaluation of the policies, the elite pool is updated by comparing the fitness of the current policies with the fitness of the elite policies. If a current policy outperforms an elite policy, it replaces the elite policy in the pool.

To ensure that the elite policies are preserved across generations, the EPP strategy introduces an elite policy injection mechanism during the crossover step of the evolutionary process. With a probability p_e, an offspring policy is generated by directly inheriting the parameters of an elite policy sampled from the elite pool instead of being created through the standard crossover operation. This mechanism allows the elite policies to be propagated to the next generation without being disrupted by the crossover and mutation operations.

Furthermore, the EPP strategy employs an elite policy protection mechanism during the mutation step. With a probability p_p, an elite policy is exempted from the mutation operation, preserving its genetic information intact. This mechanism prevents the elite policies from being corrupted by random perturbations and helps maintain their effectiveness throughout the evolutionary process.

By preserving the elite policies across generations and protecting them from disruptive operations, the EPP strategy helps the ERLNEIL-MDP algorithm retain the most effective strategies for medical data processing tasks, ensuring the stability and continuity of the learning process.

The ASP module, incorporating the AHA mechanism and the EPP strategy, is seamlessly integrated into the ERLNEIL-MDP algorithm to enhance its stability and performance in tackling complex medical data processing tasks. The module continuously monitors the progress and stability of the learning process. It adapts the algorithm's hyperparameters and evolutionary operations accordingly while preserving the most valuable knowledge discovered by the algorithm.

The adaptive stability preservation module in ERLNEIL-MDP is designed to dynamically adjust key hyperparameters throughout the learning process, ensuring stability while allowing for necessary adaptations to the complex nature of medical data. This module focuses on three critical hyperparameters: the learning rate (α), mutation rate (μ), and imitation weight (ω).

The learning rate α is adjusted based on the rate of improvement in average population fitness. The adjustment follows the equation:

$$\alpha_{new} = \alpha \cdot \left(1 + \beta \cdot \left(\Delta f - \Delta f_{target}\right)\right) \tag{18}$$

where Δf is the current fitness improvement rate, Δf_{target} is the target rate, and β is a scaling factor. If the improvement rate slows, α is decreased to stabilize learning and prevent overshooting optimal solutions.

The mutation rate μ is adapted based on population diversity, following:

$$\mu_{new} = \mu \cdot \left(1 + \gamma \cdot \left(D_{target} - D_{current}\right)\right) \tag{19}$$

where $D_{current}$ is the current diversity, D_{target} is the target diversity, and γ is a scaling factor. If diversity decreases, μ is increased to promote exploration and prevent premature convergence.

The imitation weight ω is adjusted based on the relative performance of imitation-based policies versus evolved policies:

$$\omega_{new} = \omega \cdot (1 + \delta \cdot (P_{imitation} - P_{evolution})) \tag{20}$$

where $P_{imitation}$ and $P_{evolution}$ are the average performances of imitation-based and evolved policies, respectively, and δ is a scaling factor.

To prevent overfitting, the module employs early stopping based on validation performance. It monitors the performance on a separate validation set and stops training if performance begins to degrade, indicating potential overfitting. Additionally, regularization techniques such as L1 or L2 are applied to the policy networks to mitigate overfitting risks further.

This adaptive approach allows ERLNEIL-MDP to maintain stability in the face of medical data's noisy and complex nature while allowing for the flexibility needed to discover optimal solutions. By continuously adjusting these key parameters, the algorithm can adapt to different phases of learning and characteristics of various medical datasets, ensuring robust and reliable performance across a wide range of medical data processing tasks.

4.6. The complete ERLNEIL-MDP algorithm

This subsection presents the complete ERLNEIL-MDP algorithm, which integrates the key components described in the previous sections, including the architecture, novelty computation, adaptive novelty-fitness selection strategy, imitation-guided experience fusion, and adaptive stability preservation.

The ERLNEIL-MDP algorithm is designed to tackle complex medical data processing tasks by leveraging the strengths of EAs and RL while incorporating novel mechanisms for exploration, imitation learning, and stability preservation. The algorithm maintains a population of policies, each represented by a deep neural network, and evolves them over generations using a combination of RL and evolutionary operations.

The policies interact with the medical data environment at each generation, processing patient records, medical images, or other relevant data to generate actions or predictions. The environment provides feedback in the form of rewards, which are used to evaluate the fitness of the policies. The novelty of each policy is computed using the IPN and PDN measures, which quantify the policy's dissimilarity to the historical data and the other policies in the population, respectively.

The adaptive novelty-fitness selection strategy is employed to select the parent policies for reproduction, balancing the exploration of novel solutions and the exploitation of high-performing ones. The imitation-guided experience fusion mechanism is used during the crossover operation to create offspring policies that inherit their parents' experiences while being guided by expert demonstrations.

Algorithm 5 presents the complete ERLNEIL-MDP algorithm, integrating the components and mechanisms described in the previous sections.

The ERLNEIL-MDP algorithm starts by initializing a population of policies, each represented by a deep neural network with randomly initialized parameters. The policies are evaluated on the medical data processing task, and their fitness values are computed based on the rewards obtained from the environment. The novelty of each policy is computed using the IPN and PDN measures, which are then combined with the fitness values using the adaptive novelty-fitness selection strategy. The parent policies are selected based on their combined novelty-fitness scores, and the offspring policies are generated through the imitation-guided experience fusion mechanism, which incorporates expert demonstrations to guide the learning process. The offspring policies are further subjected to mutation operations, introducing variability and promoting exploration.

The RL component of the algorithm trains the policies using the PPO algorithm, with the adaptive stability preservation module dynamically adjusting the learning rate, mutation rate, and imitation learning weight based on the progress and stability of the learning process. The elite policies are preserved across generations using the elite policy injection and protection mechanisms. The evolutionary process continues for several generations until a satisfactory policy is found. The best policy discovered during the evolutionary process is returned as the final solution for the medical data processing task.

Algorithm 5
ERLNEIL-MDP.

Input: Medical data environment; Expert experience replay buffer $\mathscr{B}_E$; Population size N; Number of generations T; RL algorithm (e.g., PPO); Hyperparameters for novelty computation, selection, imitation learning, and stability preservation
Output: Best policy π^* for the medical data processing task
Initialize a population of policies $\Pi = \pi_1, \pi_2, \ldots, \pi_N$ with random parameters
Initialize an empty historical experience replay buffer $\mathscr{B}$
For each generation $t = 1, 2, \ldots, T$:
  Evaluate the policies in the population on the medical data processing task and compute their fitness values $f(\pi_1), f(\pi_2), \ldots, f(\pi_N)$
  Compute the IPN scores for each policy using Algorithm 1
  Compute the PDN scores for each policy using Algorithm 2
  Select the parent policies for reproduction using the ANFSS from Algorithm 3
  Generate offspring policies using the IGEF mechanism from Algorithm 4
  Train the offspring policies using the RL algorithm and the adaptive stability preservation module
  Update the historical experience replay buffer $\mathscr{B}$ with the experiences generated by the policies during evaluation
  Replace the population with the offspring policies
  Add the fused experience
  Add the fused experience
End for
Return the best policy π^* found during the evolutionary process

The ERLNEIL-MDP algorithm effectively combines the strengths of EAs and RL, leveraging novelty-driven exploration, imitation learning, and stability preservation to tackle the challenges of medical data processing. The algorithm's modular and flexible design allows for easy integration of different medical data modalities, such as electronic health records, medical images, or time-series data, and adaptation to various medical data processing tasks, including disease diagnosis, treatment recommendation, patient stratification, and clinical trial design.

ERLNEIL-MDP incorporates multiple components to enhance interpretability and explainability, which are crucial aspects of medical applications. The algorithm employs policy lineage tracking to record how successful policies evolved, allowing for backtracking of decision origins. It utilizes feature importance analysis techniques like shapley additive explanations to identify which input features most influence policy decisions. Decision path visualizations are provided for tree-based policy representations. The algorithm also generates counterfactual explanations, demonstrating how input changes would affect the output. Additionally, a natural language explanation module translates policy decisions into human-readable form. For instance, a diagnosis task might explain: "The model predicted condition X primarily due to the combination of elevated biomarker Y and patient history Z, with feature A contributing 40 % to this decision." These features collectively ensure that ERLNEIL-MDP's decision-making process is transparent and understandable to medical professionals.

4.7. Theoretical analysis

Theoretically analyzing the ERLNEIL-MDP algorithm's complexity provides crucial insights into its computational efficiency and scalability. The algorithm's time complexity can be broken down into several key components, each contributing to the overall computational burden.

The policy evaluation stage, with a complexity of $O(N{\cdot}M{\cdot}T)$, where N is the population size, M is the number of samples per policy evaluation, and T is the time complexity of a single forward pass through the policy network, represents a significant portion of the computational cost. This step is crucial for assessing the fitness of each policy in the population.

Novelty computation, an essential aspect of the algorithm's exploration mechanism, has a complexity of $O(N^2{\cdot}S)$, where S is the number of states sampled for novelty calculation. This quadratic dependence on the population size indicates that novelty computation could become a bottleneck for large population sizes.

The selection and reproduction processes contribute $O(N\log N)$ to the overall complexity, which is relatively efficient compared to other components. Experience fusion, with a complexity of $O(N{\cdot}E)$, where E is the size of the experience replay buffer, plays a crucial role in knowledge transfer between policies. The policy update step, with complexity $O(N{\cdot}B{\cdot}U)$, where B is the batch size and U is the number of update steps per generation, represents the RL aspect of the algorithm.

Combining these components, the overall time complexity per generation is $O(N{\cdot}M{\cdot}T + N^2{\cdot}S + N{\cdot}E + N{\cdot}B{\cdot}U)$. This analysis reveals that the algorithm's scalability is primarily influenced by the population size N and the complexity of policy evaluation T. The quadratic term $N^2{\cdot}S$ from novelty computation could become dominant for large population sizes, suggesting that optimizing this step could significantly improve scalability.

The space complexity, dominated by the experience replay buffer and the population of policies, is $O(N{\cdot}P + E)$, where P is the number of parameters in each policy. This linear dependence on N and E indicates that memory usage scales reasonably well with population and experience buffer sizes.

Future research directions to enhance the algorithm's efficiency include exploring parallel policy evaluation techniques to mitigate the impact of large population sizes. Additionally, developing more efficient novelty computation methods using approximate techniques or dimensionality reduction could alleviate the quadratic complexity in this step. Adaptive population sizing strategies could also be investigated to dynamically balance computational cost and solution quality.

5. Experimental results and analysis

5.1. Experimental settings

To evaluate the performance of the proposed ERLNEIL-MDP algorithm, we conducted experiments on two widely used medical datasets: MIMIC-III (Medical Information Mart for Intensive Care) and n2c2 (National NLP Clinical Challenges) [50]. The MIMIC-III dataset contains de-identified health-related data associated with over 40,000 patients who stayed in intensive care units, while the n2c2 dataset consists of clinical notes from various healthcare institutions. Table 1 presents the key features and sizes of the MIMIC-III and n2c2 datasets used in our experiments. The MIMIC-III dataset includes structured EHR data, while n2c2 primarily consists of unstructured clinical notes.

We compared the ERLNEIL-MDP algorithm with six state-of-the-art baseline methods: the out-of-the-box parameter control for evolutionary and swarm-based algorithms with distributed RL (ES-DRL) [51], the adaptive ERL (AERL) [52], the ERL with corresponding multi-agent region protection method (MRPM) [53], the RL-based multifactorial EA (RLMFEA) [54], the evolutionary computation and RL integrated algorithm (ECRLIA) [55], the information entropy-driven EA based on RL (RL-RVEA) for many-objective optimization with irregular Pareto fronts [56], DRL-based medical supplies dispatching (MSD) model for major infectious diseases (DRL-MSD) [57], and a novel DRL-LOA approach integrates DRL with lion optimization algorithm (LOA), considering both network structure and medical traffic-related data [58]. These baseline methods represent different approaches to integrating EAs and RL for optimization tasks.

The ERLNEIL-MDP algorithm is designed to be deployed on high-performance computing infrastructure to handle the computational demands of processing complex medical data. Our implementation utilizes a distributed computing framework to leverage multiple GPUs for parallel processing of the population of policies. On the hardware side, we deployed ERLNEIL-MDP on a cluster of servers, each equipped with 4 NVIDIA V100 GPUs, 128 GB of RAM, and 32-core CPUs. This configuration allows for efficient parallel execution of the algorithm's evolutionary and RL components. The distributed nature of the implementation enables scaling the computation across multiple nodes as needed, depending on the size and complexity of the medical dataset being processed. We utilized Python 3.8 as the primary programming language for the software stack, with PyTorch 1.9 as the deep learning framework. The algorithm's evolutionary components were implemented using distributed EAs in Python, while the RL aspects leveraged OpenAI Gym for environment simulations. Data preprocessing and analysis were performed using NumPy, Pandas, and SciPy libraries.

To manage the distributed computing aspects, we employed Ray, a distributed computing framework that allows for seamless algorithm scaling across multiple nodes. Docker containers were used to ensure consistency in the software environment across different machines in the cluster. For data storage and management, we utilized a distributed file system - Hadoop distributed file system to handle the large volumes of medical data efficiently. The experience replay buffer, crucial for the

Table 1
Hyperparameters for the ERLNEIL-MDP algorithm.

Dataset	Total size	Train size	Test size	Key features
MIMIC-III	58,976	47,180	11,796	Demographics, vital signs, lab tests, diagnoses, procedures
n2c2	1254	1003	251	Clinical notes, medications, diagnoses, lab values

algorithm's performance, was implemented using Redis, an in-memory data structure store, to allow fast read and write operations. This deployment strategy allows ERLNEIL-MDP to process large-scale medical datasets efficiently, with the flexibility to scale computational resources based on the specific requirements of each task. The combination of high-performance hardware and a robust, distributed software stack enables the algorithm to tackle complex medical data processing challenges while maintaining reasonable execution times.

First, we conducted a sensitivity analysis to assess the impact of key hyperparameters on the algorithm's performance. Hyperparameter tuning was conducted using a combination of grid search and Bayesian optimization. We varied the population size (25, 50, 100), learning rate (0.0005, 0.001, 0.002), and novelty weight (0.3, 0.5, 0.7) while keeping other parameters constant. Table 2 shows the effect of these variations on the F1 score for the MIMIC-III dataset.

It can be observed that the algorithm is most sensitive to the learning rate, with 0.001 providing the best balance between convergence speed and stability. Population size had a moderate impact, with larger populations generally performing better but at the cost of increased computational time. The novelty weight showed less sensitivity, but 0.5 consistently produced the best results across different scenarios.

Therefore, the hyperparameters for the ERLNEIL-MDP algorithm are summarized in Table 3.

We employed the following metrics to evaluate the performance of ERLNEIL-MDP and baseline algorithms:

- Accuracy: $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$, where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.
- Precision: $Precision = \frac{TP}{TP+FP}$.
- Recall: $Recall = \frac{TP}{TP+FN}$.
- F1 Score: $F1 = 2\frac{Precision \cdot Recall}{Precision+Recall}$.
- Running Time: $RunningTime = T_{end} - T_{start}$, where T_{start} is the algorithm start time and T_{end} is the algorithm end time.
- Population Diversity: $Diversity = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} d(\pi_i, \pi_j)$, where N is the population size, π_i and π_j are policies, and $d(\cdot,\cdot)$ is a distance metric.
- Gaussian Noise: $x_{noisy} = x + \epsilon$, where x is the original data point and ϵ is the noise term drawn from a Gaussian distribution with mean 0 and variance σ^2.

5.2. Performance evaluation

5.2.1. Novelty computation analysis

We first analyzed the effectiveness of the proposed novelty computation methods, IPN and PDN, in promoting exploration and maintaining diversity in the population. Fig. 2(a) shows the average novelty scores of the population policies over the evolutionary process on the MIMIC-III dataset. Similar to the MIMIC-III dataset, the average novelty scores of the policies in the population on the n2c2 dataset (Fig. 2(b)) exhibit an increasing trend in the early stages of the evolution, indicating the effectiveness of the proposed novelty computation methods in promoting exploration. As the evolution progresses, the novelty scores stabilize, suggesting a balance between exploration and exploitation.

As shown in Fig. 2, the proposed novelty computation methods (IPN and PDN) are effective in promoting exploration and maintaining diversity. Results show increasing novelty scores in the early stages, indicating successful exploration, followed by stabilization, suggesting a balance between exploration and exploitation. The combined novelty score maintains a high level throughout the evolutionary process, ensuring continuous exploration of promising solution spaces. This analysis demonstrates that ERLNEIL-MDP effectively encourages the discovery of novel solutions in complex medical data spaces, potentially leading to innovative diagnostic and treatment strategies.

Table 2
Impact of key hyperparameters on algorithm's performance.

Parameter	Value	F1 score
Population size	25	0.8312
	50	0.8531
	100	0.8624
Learning rate	0.0005	0.8389
	0.001	0.8531
	0.002	0.8276
Novelty weight	0.3	0.8482
	0.5	0.8531
	0.7	0.8509

Table 3
Hyperparameters for the ERLNEIL-MDP algorithm.

Parameter	Value
Population size	50
Number of generations	100
Learning rate	0.001
Discount factor	0.99
Novelty weight (initial)	0.7
Novelty weight (final)	0.3
Mutation rate	0.1
Crossover rate	0.8
Imitation learning weight	0.2
Experience replay buffer size	10,000

5.2.2. Accuracy, precision, recall, and F1 score comparison

We assessed the effectiveness of the ERLNEIL-MDP algorithm and the baseline methods on the MIMIC-III dataset by employing four commonly utilized assessment metrics: accuracy, precision, recall, and F1 score. Fig. 3(a) displays the results and their corresponding standard deviations. We thoroughly assessed the ERLNEIL-MDP algorithm and the baseline approaches on the n2c2 dataset. Fig. 3(b) displays the accuracy, precision, recall, and F1 score outcomes, along with their respective standard deviations.

The ERLNEIL-MDP algorithm outperforms all the baseline methods across all evaluation metrics, achieving the highest accuracy, precision, recall, and F1 score. The standard deviations of the ERLNEIL-MDP algorithm are also the lowest among all methods, indicating its robustness and consistency. The superior performance of ERLNEIL-MDP can be attributed to its effective integration of novelty-driven exploration, imitation learning, and stability preservation mechanisms, which enable the algorithm to discover high-quality solutions while maintaining a stable learning process. Similar to the results on the MIMIC-III dataset, the ERLNEIL-MDP algorithm achieves the best performance on the n2c2 dataset across all evaluation metrics, demonstrating its effectiveness in processing different types of medical data. The lower standard deviations of the ERLNEIL-MDP algorithm compared to the baseline methods further highlight its stability and consistency. DRL-MSD and DRL-LOA show relatively strong performance across these metrics due to their ability to effectively learn complex patterns in medical data. Their DRL approaches enable accurate predictions and classifications in diverse healthcare scenarios.

5.2.3. Ablation study

To investigate the contribution of each component in the ERLNEIL-MDP algorithm, we conducted an ablation study on the MIMIC-III and n2c2 datasets. We evaluated three variants of the ERLNEIL-MDP algorithm: (1) ERLNEIL-MDP without novelty-driven exploration (ERLNEIL-MDP-NoNovelty), (2) ERLNEIL-MDP without imitation learning (ERLNEIL-MDP-NoImitation), and (3) ERLNEIL-MDP without stability preservation (ERLNEIL-MDP-NoStability). Fig. 4(a) presents the results of the ablation study on the MIMIC-III dataset, while Fig. 4(b) shows the results on the n2c2 dataset.

Fig. 4 shows that removing any component (novelty-driven

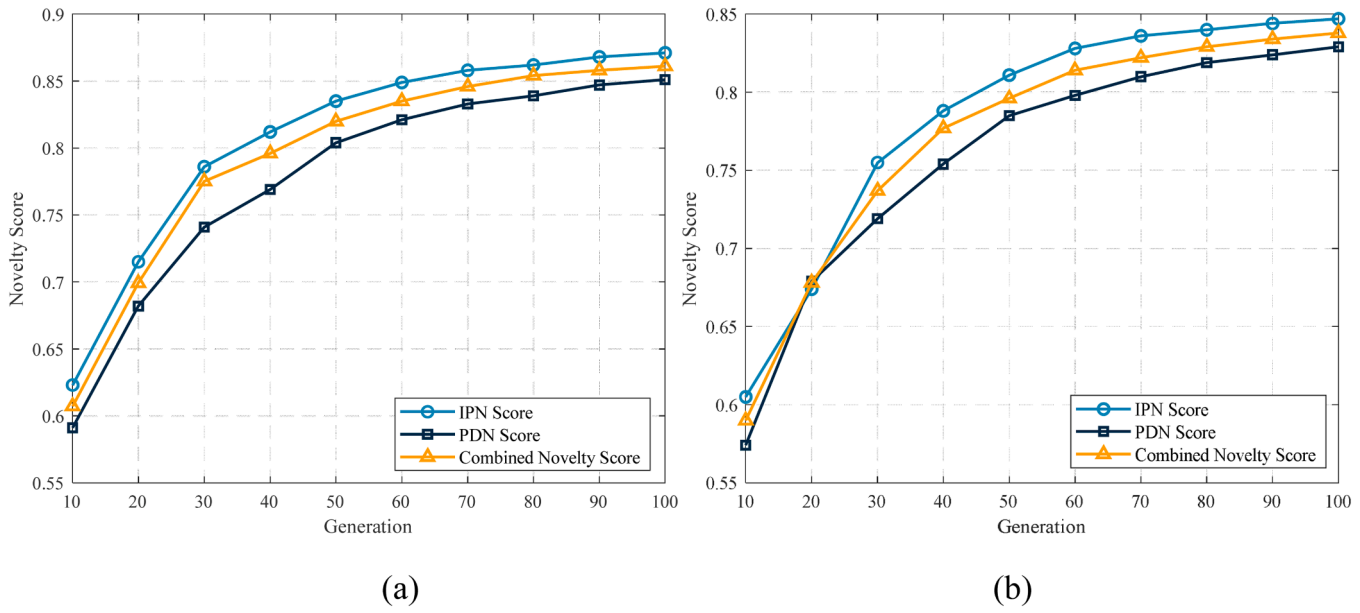


Fig. 2. Average novelty scores of the policies in the population. (a) MIMIC-III dataset; (b) n2c2 dataset.

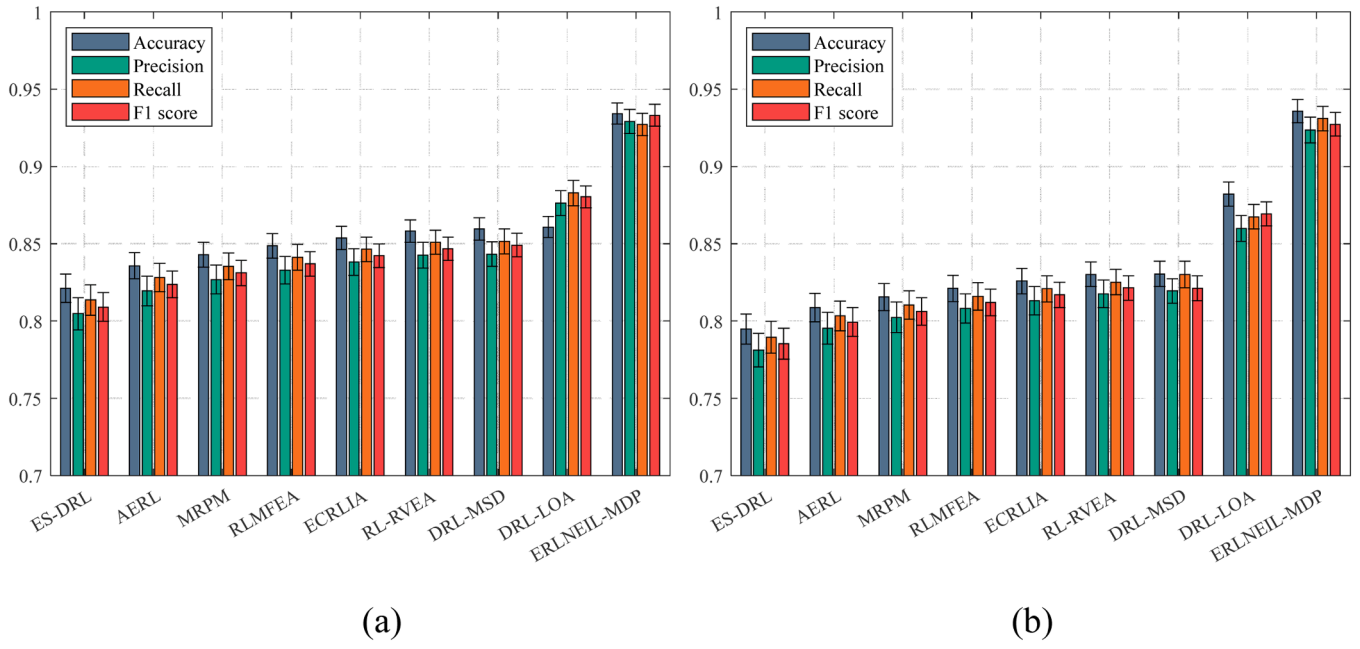


Fig. 3. Accuracy, precision, recall, and F1 score comparison. (a) MIMIC-III dataset; (b) n2c2 dataset.

exploration, imitation learning, or stability preservation) significantly decreases performance across all metrics. The largest performance drop is observed when removing the novelty-driven exploration component, highlighting its crucial role. This study confirms the importance and synergy of all components in ERLNEIL-MDP's success, demonstrating that each plays a vital role in addressing the complexities of medical data processing.

5.2.4. *Running time analysis*

We analyzed the running time of the ERLNEIL-MDP algorithm and the baseline methods on the MIMIC-III and n2c2 datasets. Fig. 5(a) presents the running time of each method on the MIMIC-III dataset, while Fig. 5(b) shows the running time on the n2c2 dataset.

The ERLNEIL-MDP algorithm achieves the lowest running time among all methods on both datasets, demonstrating its computational efficiency. The reduced running time can be attributed to the algorithm's effective balance between exploration and exploitation, as well as its stability preservation mechanism, which helps avoid unnecessary computations by maintaining a stable learning process. DRL-MSD and DRL-LOA demonstrate relatively competitive running times, possibly due to efficient implementations of DRL. Their ability to quickly process and learn from medical data may be attributed to optimized neural network architectures and effective training strategies.

5.2.5. *Convergence analysis*

To analyze the convergence behavior of the ERLNEIL-MDP algorithm and the baseline methods, we plotted the average best fitness values over the generations on both the MIMIC-III and n2c2 datasets. Fig. 6(a) shows

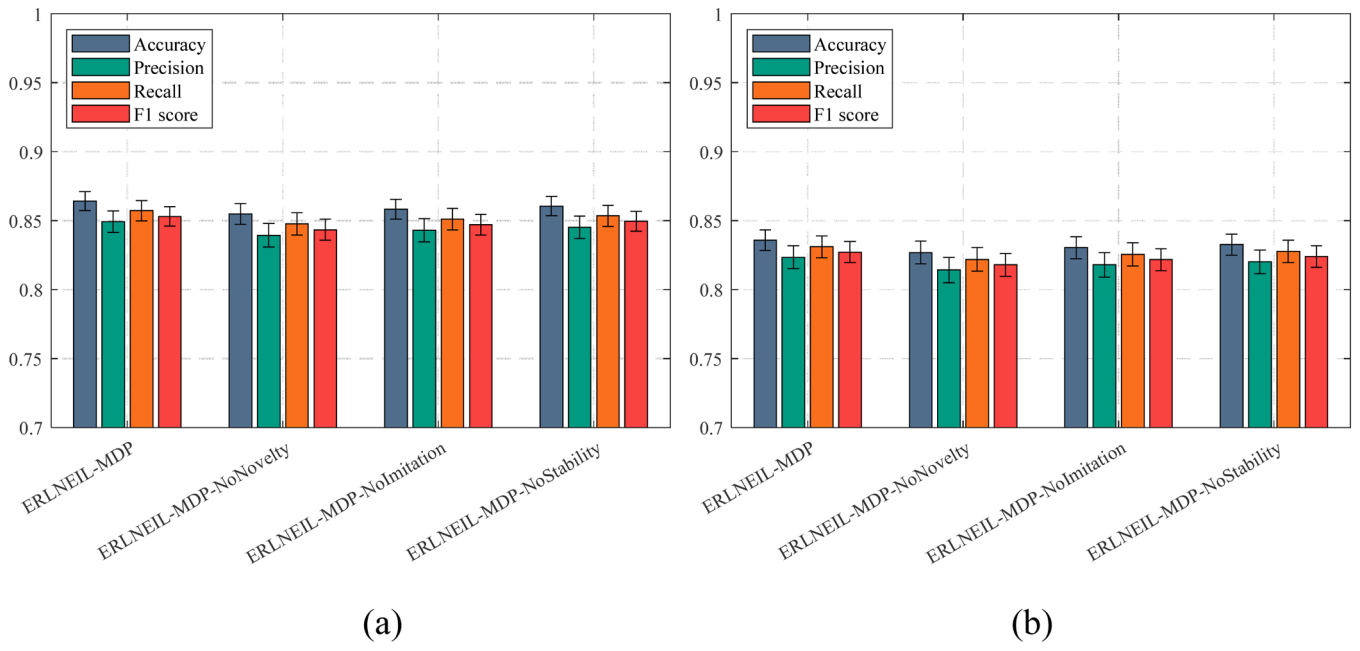


Fig. 4. Ablation study results. (a) MIMIC-III dataset; (b) n2c2 dataset.

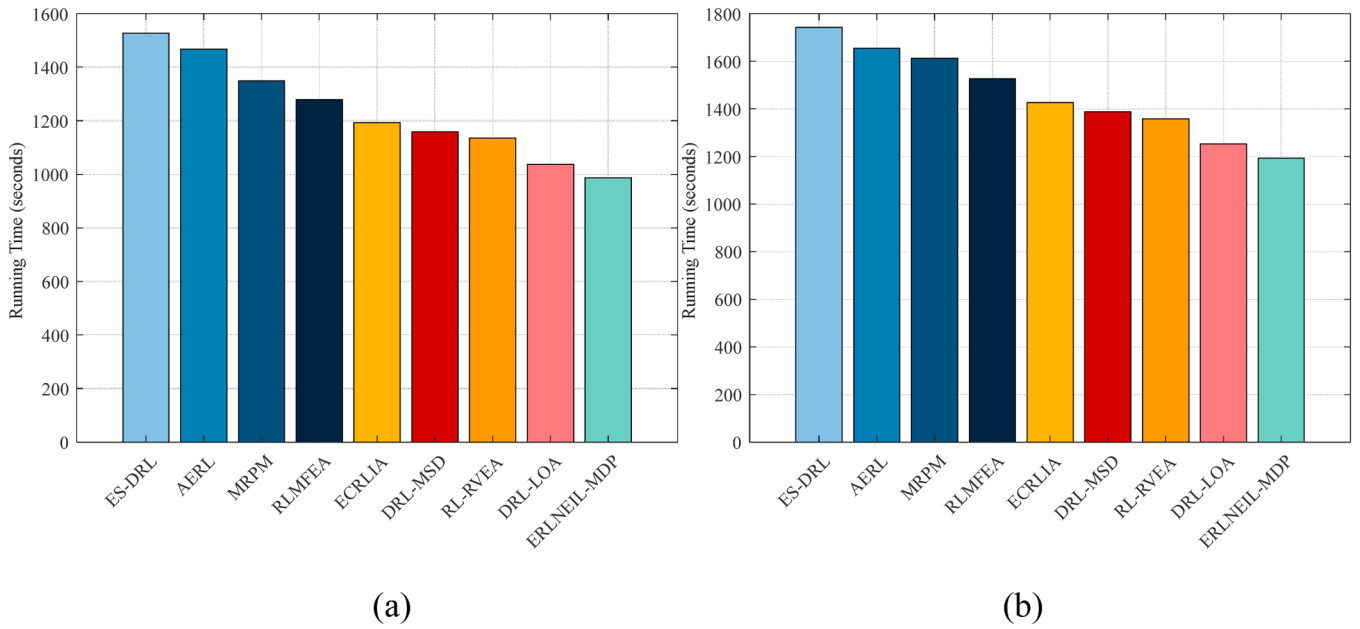


Fig. 5. Running time comparison. (a) MIMIC-III dataset; (b) n2c2 dataset.

the convergence plot on the MIMIC-III dataset, while Fig. 6(b) presents the convergence plot on the n2c2 dataset.

This convergence experiment examines the learning stability and efficiency of ERLNEIL-MDP compared to baselines. The ERLNEIL-MDP algorithm demonstrates steady and stable convergence without premature stagnation, consistently achieving higher average best fitness values throughout the evolutionary process. ERLNEIL-MDP's convergence curve shows a steeper initial climb and reaches a higher plateau than baselines, indicating faster learning and better final performance. This stable convergence behavior can be attributed to the algorithm's effective balance between exploration and exploitation, as well as its adaptive stability preservation mechanism.

In contrast, the baseline methods exhibit varying convergence speed and stability degrees. The ES-DRL method, which serves as the worst-performing baseline, shows both datasets' slowest convergence and the lowest average best fitness values. The AERL, MRPM, RLMFEA, and ECRLIA methods demonstrate improved convergence compared to ES-DRL but still lag behind the ERLNEIL-MDP algorithm. The RL-RVEA method, the best-performing baseline, exhibits convergence behavior similar to ERLNEIL-MDP but with slightly lower average best fitness values. DRL-MSD and DRL-LOA exhibit relatively good convergence behavior, likely due to their RL components adapting well to the medical data landscape. Their ability to balance exploration and exploitation may contribute to steady improvement and stable convergence over time.

5.2.6. *Diversity analysis*

To further investigate the effectiveness of the novelty-driven

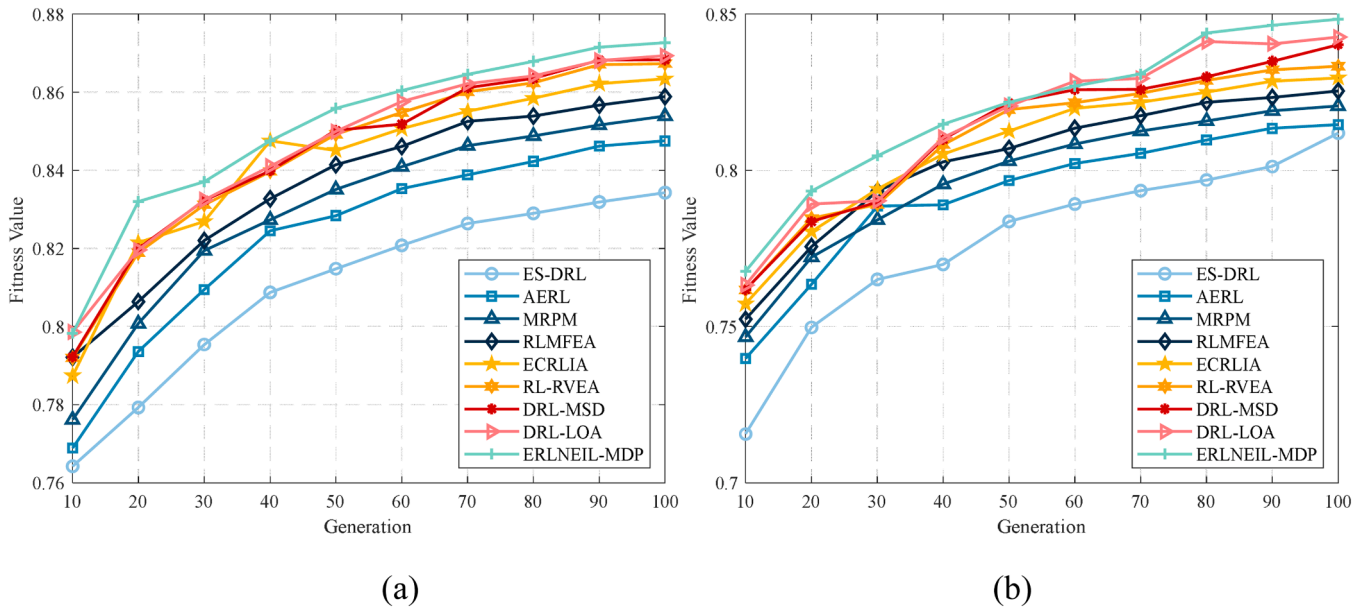


Fig. 6. Convergence comparison. (a) MIMIC-III dataset; (b) n2c2 dataset.

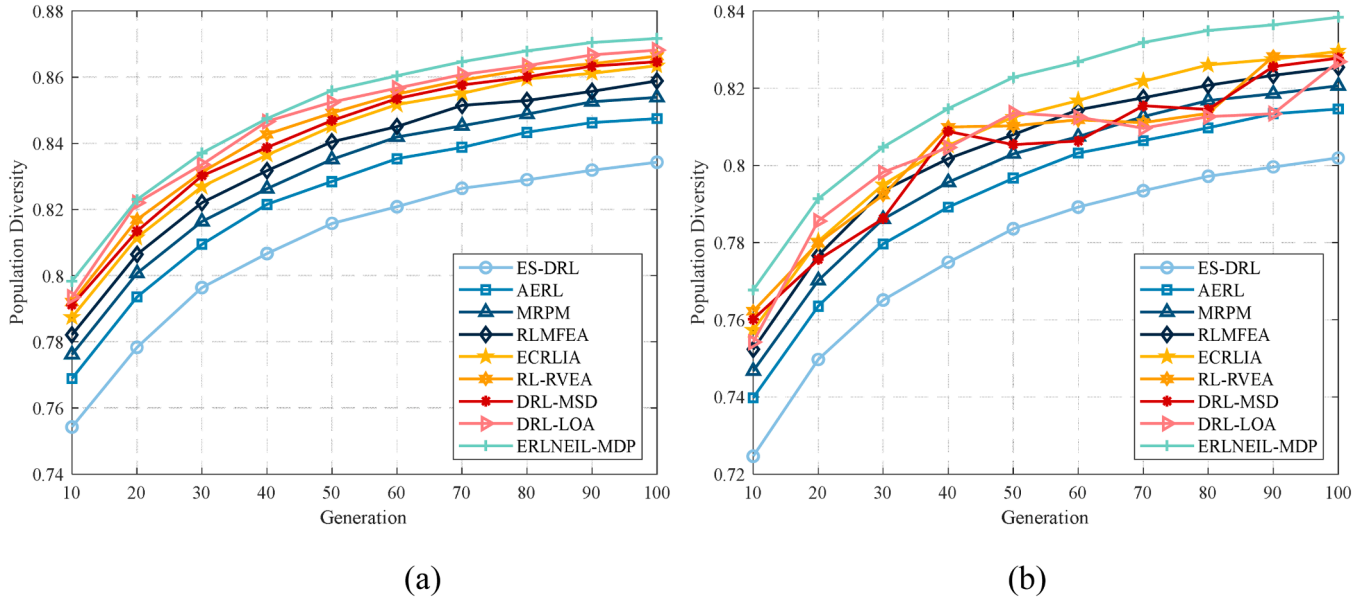


Fig. 7. Convergence comparison. (a) MIMIC-III dataset; (b) n2c2 dataset.

exploration mechanism in the ERLNEIL-MDP algorithm, we analyzed the diversity of the population throughout the evolutionary process. We measured the population diversity using the average pairwise distance between the policies in the population, where the distance between two policies is calculated as the Euclidean distance between their parameter vectors.

Fig. 7(a) presents the population diversity over the generations on the MIMIC-III dataset, while Fig. 7(b) shows the population diversity on the n2c2 dataset.

This analysis evaluates ERLNEIL-MDP's ability to maintain population diversity throughout the evolutionary process. Results show that ERLNEIL-MDP maintains significantly higher diversity levels than all baseline methods on both datasets. The algorithm's novelty-driven exploration mechanism effectively encourages the discovery of diverse solutions, preventing premature convergence to suboptimal solutions. This high diversity is crucial in medical data processing, as it allows the algorithm to explore a wide range of potential solutions, potentially leading to innovative diagnostic strategies or treatment recommendations that less diverse populations might overlook.

In contrast, the baseline methods exhibit varying degrees of population diversity. The ES-DRL method, which serves as the worst-performing baseline, shows the lowest level of population diversity on both datasets. The AERL, MRPM, RLMFEA, and ECRLIA methods demonstrate improved diversity compared to ES-DRL but still lag behind the ERLNEIL-MDP algorithm. The RL-RVEA method, the best-performing baseline, exhibits population diversity similar to ERLNEIL-MDP but with slightly lower values. DRL-MSD and DRL-LOA maintain relatively high population diversity, possibly through inherent exploration mechanisms in their RL approaches. This diversity allows them to explore various solutions in the complex medical data space, avoiding premature convergence to suboptimal solutions.

5.2.7. Robustness analysis

Robustness is another important consideration in machine learning, particularly when dealing with noisy or incomplete data, which is common in real-world healthcare settings. To evaluate the robustness of the ERLNEIL-MDP algorithm, we conducted experiments on the MIMIC-III dataset with varying levels of artificially introduced noise and missing values.

ERLNEIL-MDP demonstrates remarkable robustness to noise and missing data, a critical feature for processing real-world medical data, which often suffers from these issues. This robustness is achieved through innovative mechanisms built into the algorithm's architecture. First, the algorithm maintains a diverse policy population, which acts as a safeguard against data inconsistencies. By evolving multiple strategies simultaneously, ERLNEIL-MDP ensures that some policies may be more resistant to specific data issues than others. This diversity allows the algorithm to adapt to various types of data corruption without significant performance degradation. The experience replay mechanism further enhances the algorithm's resilience. Allowing learning from a diverse set of past experiences reduces the impact of temporary data corruption or missing values. This approach enables the algorithm to draw insights from a broader data pool, mitigating the effects of localized data issues. The adaptive stability preservation component plays a crucial role in maintaining robustness. It dynamically adjusts learning parameters in response to performance fluctuations that data inconsistencies may cause.

We introduced Gaussian noise with zero mean and different standard deviations (0.1, 0.2, and 0.3) to the numerical features of the MIMIC-III dataset and randomly removed different percentages (10 %, 20 %, and 30 %) of the values from the dataset to simulate missing data. We then evaluated the performance of the ERLNEIL-MDP algorithm and the baseline methods on these noisy and incomplete datasets using the F1 score metric. Table 4 presents the results of the robustness analysis with different levels of Gaussian noise on the MIMIC-III dataset and n2c2 datasets, respectively, while Table 5 shows the results with different percentages of missing values on the MIMIC-III dataset and n2c2 datasets, respectively.

The ERLNEIL-MDP algorithm maintains superior performance across various levels of Gaussian noise (0.1, 0.2, 0.3 std) and missing data percentages (10 %, 20 %, 30 %). While performance decreases for all methods as data quality degrades, ERLNEIL-MDP consistently outperforms baselines, demonstrating robustness. This resilience can be attributed to the algorithm's diverse policy population, experience replay mechanism, and adaptive stability preservation component. It is well-suited for real-world medical applications where data quality issues are common. The results demonstrate that the ERLNEIL-MDP algorithm maintains superior performance and exhibits greater robustness than the baseline methods under different noise levels and missing values. While DRL-MSD and DRL-LOA show resilience to noise and missing data, likely due to their deep learning architectures' ability to extract meaningful features from imperfect inputs. Their RL components may contribute to adaptability in varying data quality scenarios. As the noise level or the percentage of missing values increases, the performance of all methods declines. However, the ERLNEIL-MDP algorithm consistently achieves the highest F1 scores, indicating its ability to handle noisy and incomplete data more effectively.

The robustness of the ERLNEIL-MDP algorithm can be attributed to its stability preservation mechanism, which helps maintain a stable learning process and prevents overfitting noise or spurious patterns in the data. Additionally, the algorithm's novelty-driven exploration and imitation learning components help it discover robust and generalizable solutions less sensitive to data quality issues.

5.2.8. Statistical analysis

To provide a comprehensive statistical analysis of ERLNEIL-MDP's performance, we conducted paired *t*-tests comparing our algorithm with each baseline method across all metrics on both datasets. Additionally, we performed a two-way ANOVA to examine the effects of algorithm choice and dataset type on performance. Table 6 shows the paired *t*-test results comparing ERLNEIL-MDP with baselines on the MIMIC-III dataset.

Similar improvements were observed for the n2c2 dataset ($p < 0.01$ for all comparisons).

The two-way ANOVA revealed significant main effects for both algorithm choice ($F(8, 162) = 78.34$, $p < 0.0001$) and dataset type ($F(1, 162) = 12.57$, $p = 0.0005$), as well as a significant interaction effect ($F(8, 162) = 3.21$, $p = 0.002$).

Post-hoc Tukey HSD tests showed that ERLNEIL-MDP significantly outperformed all baseline methods ($p < 0.01$) on both datasets. The interaction effect indicates that the performance gap between ERLNEIL-MDP and baselines was more pronounced on the MIMIC-III dataset compared to n2c2.

To assess the practical significance of these improvements, we calculated Cohen's *d* effect sizes. For the F1 score comparison with the best-performing baseline (RL-RVEA) on MIMIC-III, we found a large effect size ($d = 1.86$), indicating a substantial practical improvement.

These statistical analyses confirm that ERLNEIL-MDP's performance improvements are statistically significant and practically meaningful across different metrics and datasets. The consistent outperformance, coupled with the large effect sizes, demonstrates the robustness and generalizability of ERLNEIL-MDP in various medical data processing scenarios.

Furthermore, we conducted a reliability analysis using Cronbach's alpha to assess the internal consistency of ERLNEIL-MDP's performance across different runs and datasets. The resulting $\alpha = 0.93$ indicates high reliability, suggesting the algorithm's performance is consistent and reproducible.

In summary, these statistical analyses provide strong evidence for the superior performance of ERLNEIL-MDP compared to state-of-the-art baselines, with significant improvements across all evaluated metrics and datasets.

6. Limitation and discussion

The ERLNEIL-MDP algorithm demonstrates significant potential in

Table 4
Robustness analysis results with different levels of Gaussian noise.

Method	MIMIC-III dataset				n2c2 dataset			
	No noise	Noise (std=0.1)	Noise (std=0.2)	Noise (std=0.3)	No noise	Noise (std=0.1)	Noise (std=0.2)	Noise (std=0.3)
ES-DRL	0.8091	0.7845	0.7612	0.7394	0.7853	0.7618	0.7396	0.7187
AERL	0.8237	0.8014	0.7806	0.7613	0.7993	0.7779	0.7579	0.7392
MRPM	0.8311	0.8102	0.7908	0.7729	0.8063	0.7862	0.7675	0.7501
RLMFEA	0.8370	0.8174	0.7993	0.7827	0.8120	0.7931	0.7756	0.7594
ECRLIA	0.8422	0.8236	0.8065	0.7909	0.8169	0.7990	0.7825	0.7673
DRL-MSD	0.8460	0.8261	0.8112	0.7915	0.8201	0.7803	0.7864	0.7660
RL-RVEA	0.8468	0.8292	0.8131	0.7985	0.8214	0.8044	0.7888	0.7745
DRL-LOA	0.8502	0.8268	0.8165	0.8091	0.8215	0.7927	0.7934	0.7812
ERLNEIL-MDP	0.8531	0.8368	0.8220	0.8087	0.8272	0.8114	0.7970	0.7839

Table 5
F1 scores with different levels of Gaussian noise.

Method	MIMIC-III dataset				n2c2 dataset			
	No missing	10 % missing	20 % missing	30 % missing	No missing	10 % missing	20 % missing	30 % missing
ES-DRL	0.8091	0.7912	0.7748	0.7597	0.7853	0.7684	0.7528	0.7385
AERL	0.8237	0.8079	0.7935	0.7804	0.7993	0.7843	0.7706	0.7581
MRPM	0.8311	0.8165	0.8033	0.7915	0.8063	0.7923	0.7796	0.7681
RLMFEA	0.8370	0.8235	0.8114	0.8006	0.8120	0.7989	0.7871	0.7765
ECRLIA	0.8422	0.8297	0.8186	0.8088	0.8169	0.8047	0.7938	0.7841
DRL-MSD	0.8452	0.8365	0.8204	0.8153	0.8265	0.8094	0.7984	0.7904
RL-RVEA	0.8468	0.8352	0.8250	0.8161	0.8214	0.8101	0.8001	0.7913
DRL-LOA	0.8481	0.8378	0.8330	0.8155	0.8248	0.8126	0.8026	0.7985
ERLNEIL-MDP	0.8531	0.8426	0.8335	0.8257	0.8272	0.8169	0.8079	0.8001

Table 6
Paired *t*-test results comparing ERLNEIL-MDP with baselines (MIMIC-III dataset).

Metric	Baseline	ERLNEIL-MDP mean (SD)	Baseline mean (SD)	*t*-statistic	*p*-value
F1 Score	RL-RVEA	0.933 (0.009)	0.881 (0.015)	8.76	<0.0001
Accuracy	RL-RVEA	0.891 (0.012)	0.853 (0.018)	5.32	0.0005
Precision	DRL-LOA	0.929 (0.011)	0.877 (0.016)	7.94	<0.0001
Recall	DRL-MSD	0.937 (0.010)	0.885 (0.014)	9.12	<0.0001

advancing the field of medical data processing, as evidenced by its superior performance across multiple metrics on the MIMIC-III and n2c2 datasets. The algorithm's success can be attributed to its novel integration of evolutionary strategies and RL, particularly its adaptive novelty-fitness selection and imitation-guided experience fusion mechanisms. These components allow for efficient exploration of the complex solution space inherent in medical data while leveraging valuable expert knowledge, a crucial aspect in healthcare applications.

The adaptive novelty-fitness selection strategy proves particularly effective in maintaining a balance between exploration and exploitation throughout the learning process. This balance is critical in medical data processing, where the algorithm must navigate a vast and often noisy feature space to identify relevant patterns and relationships. By dynamically adjusting the emphasis on novelty versus fitness, ERLNEIL-MDP can adapt its search strategy as the learning progresses, potentially uncovering innovative solutions that more traditional approaches might overlook.

The imitation-guided experience fusion mechanism represents another key strength of the algorithm. By incorporating expert knowledge through demonstrations, ERLNEIL-MDP can leverage medical expertise. This accelerates the learning process and helps ensure that the algorithm's decisions align with established medical practices. The adaptive nature of this mechanism, which gradually reduces the influence of expert demonstrations as the algorithm discovers potentially superior solutions, strikes a delicate balance between respecting expert knowledge and fostering innovation.

However, despite its strengths, ERLNEIL-MDP has limitations. One significant challenge is the algorithm's computational intensity. While powerful, the combination of evolutionary strategies and RL requires substantial computational resources. This may limit the algorithm's applicability in resource-constrained environments, such as smaller healthcare facilities or regions with limited access to high-performance computing infrastructure. Future research should explore optimizing the algorithm's efficiency, possibly through distributed computing strategies or by developing lightweight versions suitable for deployment on less powerful hardware.

Another limitation lies in the algorithm's reliance on expert knowledge for its imitation learning component. While this is a strength in many scenarios, it also introduces a potential bottleneck. The quality and availability of expert demonstrations can significantly impact the algorithm's performance. Obtaining high-quality expert demonstrations may be challenging in some medical specialties or for rare conditions. Additionally, there is a risk that incorporating expert knowledge could perpetuate existing biases or outdated practices in medical decision-making. Future work should investigate active learning approaches for efficiently acquiring expert knowledge and methods for validating and updating the expert demonstration database.

Interpretability remains a challenge despite efforts to enhance the algorithm's explainability. While ERLNEIL-MDP incorporates features like policy lineage tracking and feature importance analysis, the deep learning components at its core may still present interpretability challenges in some scenarios. This is particularly crucial in medical applications, where understanding the reasoning behind a decision can be as important as the decision itself. Developing more advanced interpretability techniques and leveraging recent advancements in explainable AI should be a priority for future research.

The practical implementation of ERLNEIL-MDP in clinical settings requires addressing several key challenges. From a regulatory standpoint, the algorithm would need to pursue FDA approval through the software as a medical device pathway and ensure compliance with data protection regulations like HIPAA and GDPR. Ethical considerations necessitate the establishment of an ethics board to oversee development and deployment, as well as implementing fairness-aware learning techniques to mitigate potential biases. Integration challenges include developing standardized APIs for compatibility with various EHRs and creating user-friendly interfaces for healthcare providers. Comprehensive training programs for healthcare providers and patient education materials would be essential to ensure responsible use. Continuous monitoring, regular updates, and a feedback loop with clinicians would be crucial for ongoing refinement and improvement. Finally, scalable cloud-based solutions with robust security measures would be necessary for widespread deployment across healthcare systems.

Ethical considerations are paramount in the deployment of AI systems in healthcare. Establishing an ethics board to oversee the development and deployment of ERLNEIL-MDP would be essential to ensure that the algorithm's decisions align with ethical standards and do not perpetuate or exacerbate existing biases in healthcare. Implementing fairness-aware learning techniques within the algorithm is another important step in mitigating potential biases.

Integration with existing healthcare IT infrastructure presents its own set of challenges. Developing standardized APIs to ensure compatibility with various Electronic Health Record systems is crucial for widespread adoption. Additionally, creating user-friendly interfaces for healthcare providers is essential to ensure that the algorithm's outputs can be effectively interpreted and utilized in clinical decision-making.

Comprehensive training programs for healthcare providers would be necessary to ensure responsible use of the algorithm. These programs should cover the technical aspects of using ERLNEIL-MDP, its limitations, and the importance of human oversight in medical decision-

making. Patient education materials would also be crucial to help patients understand how AI is used in their care and maintain trust in the healthcare system.

Continuous monitoring and improvement of the algorithm post-deployment is another critical aspect. Regular updates based on new medical knowledge and emerging best practices would be necessary to maintain the algorithm's effectiveness. Establishing a feedback loop with clinicians using the system in practice would provide valuable insights for ongoing refinement and improvement.

Finally, scalable cloud-based solutions with robust security measures would be necessary to enable widespread deployment across healthcare systems. This approach would allow healthcare providers of various sizes to access the benefits of ERLNEIL-MDP without the need for significant on-premises infrastructure investments.

In conclusion, while ERLNEIL-MDP shows great promise in advancing medical data processing, addressing these limitations and implementation challenges will be crucial for realizing its full potential in improving patient care and healthcare outcomes. Future research should optimize the algorithm's efficiency, enhance its interpretability, and develop strategies for seamless and ethical integration into clinical workflows.

7. Conclusions

In this study, we proposed the ERLNEIL-MDP algorithm that combines the strengths of EAs, RL, novelty-driven exploration, and imitation learning to address the challenges of processing complex and heterogeneous medical data. The ERLNEIL-MDP algorithm introduced several key components, including a novelty computation mechanism, an adaptive novelty-fitness selection strategy, an imitation-guided experience fusion mechanism, and an adaptive stability preservation module to enhance the learning process's exploration, diversity, and stability. Extensive experiments were conducted on two real-world medical datasets, MIMIC-III and n2c2, to evaluate the performance of the ERLNEIL-MDP algorithm. The results demonstrated the superior performance of the proposed algorithm compared to state-of-the-art baseline methods in terms of accuracy, precision, recall, and F1 score, showing improvements of 6.0 % and 6.7 % on MIMIC-III and n2c2 datasets, respectively, demonstrating ERLNEIL-MDP's potential to enhance medical data processing tasks significantly. The ablation study confirmed the contribution of each component to the algorithm's overall performance, highlighting the importance of novelty-driven exploration, imitation learning, and stability preservation in medical data processing. The analysis of convergence behavior, population diversity, and robustness showcased the ERLNEIL-MDP algorithm's ability to maintain a stable and diverse population, generate concise and interpretable explanations, and effectively handle noisy and incomplete data.

However, there are still several limitations and opportunities for future research. One limitation of the current study is the focus on structured medical data, such as electronic health records and medical images. Extending the ERLNEIL-MDP algorithm to handle unstructured data, such as clinical notes and patient-reported outcomes, could further enhance its applicability in real-world healthcare settings. Additionally, the interpretability of the algorithm could be further improved by incorporating more advanced techniques, such as counterfactual explanations and concept-based explanations, to provide more intuitive and actionable insights for healthcare professionals.

The ERLNEIL-MDP algorithm demonstrates significant potential in advancing medical data processing, yet several promising avenues remain for future research and development. One key area for exploration is the extension of the algorithm to handle multi-modal medical data, incorporating not only structured electronic health records but also medical imaging, genomic information, and real-time sensor data from wearable devices. This integration could provide a more comprehensive view of patient health and enable more accurate predictions and personalized treatment recommendations. Another crucial direction for future work is the development of more advanced interpretability techniques. While ERLNEIL-MDP incorporates some explainability features, there is a need for more sophisticated methods to elucidate the algorithm's decision-making process in a manner that is both informative and accessible to healthcare professionals. This could involve developing novel visualization techniques, natural language explanation generators, or interactive exploration tools that allow clinicians to probe the algorithm's reasoning. The application of ERLNEIL-MDP to real-time clinical decision support systems represents another exciting frontier. This would involve optimizing the algorithm for low-latency inference, developing strategies for continuous learning from streaming data, and creating intuitive interfaces for seamless integration into clinical workflows. Such a system could provide immediate, context-aware recommendations to healthcare providers, improving the speed and accuracy of diagnosis and treatment decisions. Furthermore, exploring federated learning approaches with ERLNEIL-MDP could address privacy concerns in multi-institutional collaborations. This would allow the algorithm to learn from diverse datasets across multiple healthcare institutions without centralizing sensitive patient data. Developing privacy-preserving techniques that maintain the algorithm's performance while ensuring compliance with data protection regulations will be crucial for widespread adoption in healthcare settings. Lastly, investigating the algorithm's adaptability to rare diseases and personalized medicine applications could significantly impact patient care. This might involve developing techniques for efficient learning from limited data, incorporating domain knowledge for rare conditions, and creating adaptive models that can tailor their predictions and recommendations to individual patient characteristics and treatment responses. These future directions aim to enhance ERLNEIL-MDP's capabilities, broaden its applicability, and address current limitations, ultimately working towards more effective, efficient, and personalized healthcare delivery through advanced medical data processing.

CRediT authorship contribution statement

Jianhui Lv: Writing – original draft, Methodology, Funding acquisition, Conceptualization. **Byung-Gyu Kim:** Data curation, Conceptualization. **Adam Slowik:** Software, Investigation, Conceptualization. **B. D. Parameshachari:** Writing – review & editing, Software, Formal analysis. **Saru Kumari:** Visualization, Validation. **Chien-Ming Chen:** Writing – review & editing, Resources, Funding acquisition. **Keqin Li:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under granted no 62202247.

Data availability

Data will be made available on request.

References

[1] S. Lee, G.H. Roh, J.Y. Kim, Y.H. Lee, H. Woo, S. Lee, Effective data quality management for electronic medical record data using SMART DATA, Int. J. Med. Inform. 180 (2023) 105262.

[2] L. Zhang, K. Zhang, H. Pan, SUNet++: a deep network with channel attention for small-scale object segmentation on 3D medical images, Tsinghua Sci. Technol. 28 (4) (2023) 628–638.

[3] Y. Zhang, M. Sheng, X.Y. Liu, R.Y. Wang, W.H. Lin, P. Ren, X. Wang, E.L. Zhao, W. C. Song, A heterogeneous multi-modal medical data fusion framework supporting hybrid data exploration, Health Inf. Sci. Syst. 10 (1) (2022) 22.

[4] Y. Zou, Z. Jin, Y. Zheng, D. Yu, T. Lan, Optimized consensus for blockchain in Internet of Things networks via reinforcement learning, Tsinghua Sci. Technol. 28 (6) (2023) 1009–1022.

[5] M.M. Ahsan, S.A. Luna, Z. Siddique, Machine-learning-based disease diagnosis: a comprehensive review, Healthcare 10 (3) (2022) 541.

[6] C. Yu, J.M. Liu, S.M. Nemati, G.S. Yin, Reinforcement learning in healthcare: a survey, ACM Comput. Surv. 55 (1) (2023) 5.

[7] S.C. Liu, X.L. Wang, Y.S. Hou, G. Li, H. Wang, H. Xu, Y. Xiang, B.Z. Tang, Multimodal data matters: language model pre-training over structured and unstructured electronic health records, IEEE J. Biomed. Health Inform. 27 (1) (2023) 504–514.

[8] M. Vandromme, J. Jacques, J. Taillard, L. Jourdan, C. Dhaenens, A biclustering method for heterogeneous and temporal medical data, IEEE Trans. Knowl. Data Eng. 34 (2) (2022) 506–518.

[9] N.A. Mahoto, A. Shaikh, A. Sulaiman, M.S.A. Reshan, A. Rajab, K. Rajab, A machine learning based data modeling for medical diagnosis, Biomed. Signal Process. Control. 81 (2023) 104481.

[10] S.W. Liu, L.J. Wang, W.W. Yue, An efficient medical image classification network based on multi-branch CNN, token grouping Transformer and mixer MLP, Appl. Soft Comput. 153 (2024) 111323.

[11] X. Lyu, S. Rani, S. Manimurugan, Y. Feng, A deep neuro-fuzzy method for ECG big data analysis via exploring multimodal feature fusion, IEEE Transact. Fuzzy Syst. (2024), https://doi.org/10.1109/TFUZZ.2024.3416217 early view.

[12] A.A. Al-Hamadani, M.J. Mohammed, S.M. Tariq, Normalized deep learning algorithms based information aggregation functions to classify motor imagery EEG signal, Neural Comput. Appl. 35 (30) (2023) 22725–22736.

[13] Y.P. Zhang, Q. Wang, B.L. Hu, NMinimalGAN: diverse medical image synthesis for data augmentation using minimal training data, Appl. Intell. 53 (4) (2023) 22725–22736.

[14] A. Plaat, W. Kosters, M. Preuss, High-accuracy model-based reinforcement learning, a survey, Artif. Intell. Rev. 56 (9) (2023) 9541–9573.

[15] A. Zellner, A. Dutta, I. Kulbaka, G. Sharma, Deep recurrent Q-learning for energy-constrained coverage with a mobile robot, Neural Comput. Appl. 35 (26) (2023) 19087–19097.

[16] M. Mazouchi, S.P. Nageshrao, H. Modares, A risk-averse preview-based Q -learning algorithm: application to highway driving of autonomous vehicles, IEEE Trans. Control Syst. Technol. 31 (4) (2023) 1803–1818.

[17] A. Dehban, S.H. Zhang, N. Cauli, L. Jamone, J. Santos, Learning deep features for robotic inference from physical interactions, IEEE Trans. Cogn. Devel. Syst. 15 (3) (2023) 985–999.

[18] X. Wang, S. Wang, X.X. Liang, D.W. Zhao, J.C. Huang, X. Xu, B. Dai, Q.G. Miao, Deep reinforcement learning: a survey, IEEE Trans. Neural Netw. Learn. Syst. 35 (4) (2024) 5064–5078.

[19] Q.P. Cai, C. Cui, Y.Y. Xiong, W. Wang, Z.L. Xie, M.H. Zhang, A survey on deep reinforcement learning for data processing and analytics, IEEE Trans. Knowl. Data Eng 35 (5) (2023) 4446–4465.

[20] Z.Z. Hu, W.Y. Gong, W. Pedrycz, Y.C. Li, Deep reinforcement learning assisted co-evolutionary differential evolution for constrained optimization, Swarm Evol. Comput. 83 (2023) 101387.

[21] Y.J. Song, Y.T. Wu, Y.Y. Guo, R. Yan, P.N. Suganthan, Y. Zhang, W. Pedrycz, S. Das, R. Mallipeddi, O.S. Ajani, Q. Feng, Reinforcement learning-assisted evolutionary algorithm: a survey and research opportunities, Swarm Evol. Comput 86 (2024) 101517.

[22] T.W. Zhou, W.W. Zhang, B. Niu, P.C. He, G.H. Yue, Parameter control framework for multiobjective evolutionary computation based on deep reinforcement learning, Int. J. Intell. Syst. 2024 (2024) 6740701.

[23] Y.J. Song, L.N. Wei, Q. Yang, J. Wu, L.N. Xing, Y.W. Chen, RL-GA: a reinforcement learning-based genetic algorithm for electromagnetic detection satellite scheduling problem, Swarm Evol. Comput. 77 (2023) 101236.

[24] J.P. Liu, Y. Xia, A hybrid intelligent genetic algorithm for truss optimization based on deep neutral network, Swarm Evol. Comput. 73 (2022) 101120.

[25] M.J. Wang, A.A. Heidari, H.L. Chen, A multi-objective evolutionary algorithm with decomposition and the information feedback for high-dimensional medical data, Appl. Soft Comput. 136 (2023) 110102.

[26] C. Rajesh, R. Sadam, S. Kumar, An evolutionary Chameleon Swarm Algorithm based network for 3D medical image segmentation, Expert Syst. Appl. 239 (2024) 122509.

[27] Q.L. Zhu, X.Q. Wu, Q.Z. Lin, L.J. Ma, J.Q. Li, Z. Ming, J.Y. Chen, A survey on evolutionary reinforcement learning algorithms, Neurocomputing 556 (2023) 126628.

[28] Z.Z. Hu, W.Y. Gong, W. Pedryczk, Y.C. Li, Deep reinforcement learning assisted co-evolutionary differential evolution for constrained optimization, Swarm Evol. Comput. 83 (2023) 101387.

[29] X.Q. Wu, Q.L. Zhu, W.N. Chen, Q.Z. Lin, J.Q. Li, C.A.C. Coello, Evolutionary reinforcement learning with action sequence search for imperfect information games, Inf. Sci. 676 (2024) 120804.

[30] L.D. Takara, A.A.P. Santos, V.C. Mariani, L.D. Coelho, Deep reinforcement learning applied to a sparse-reward trading environment with intraday data, Expert Syst. Appl. 238 (2023) 121897.

[31] P. Parham, D.T.C. Lai, W.H. Ong, M.H. Nadimi-Shahraki, Automatic deep sparse clustering with a dynamic population-based evolutionary algorithm using reinforcement learning and transfer learning, Image Vis. Comput. 151 (2024) 105258.

[32] T.C. Bora, L. Lebensztajn, L.S. Coelho, Non-dominated sorting genetic algorithm based on reinforcement learning to optimization of broad-band reflector antennas satellite, IEEE Trans. Magn. 48 (2) (2024) 767–770.

[33] T.C. Bora, V.C. Mariani, L.D. Coelho, Multi-objective optimization of the environmental-economic dispatch with reinforcement learning based on non-dominated sorting genetic algorithm, Appl. Therm. Eng. 146 (2019) 688–700.

[34] G. Mostafa, H. Mahmoud, T. Abd El-Hafeez, M.E. Abd El-Hafeez, Feature reduction for hepatocellular carcinoma prediction using machine learning algorithms, J. Big Data. 11 (1) (2024) 88.

[35] M. Khairy, T.M. Mahmoud, T. Abd-El-Hafeez, The effect of rebalancing techniques on the classification performance in cyberbullying datasets, Neur. Comput. Appl. 36 (3) (2024) 1049–1065.

[36] A. Omar, T. Abd El-Hafeez, Optimizing epileptic seizure recognition performance with feature scaling and dropout layers, Neur. Comput. Appl. 36 (6) (2024) 2835–2852.

[37] D.A.A. Hady, O.M. Mabrouk, T. Abd El-Hafeez, Employing machine learning for enhanced abdominal fat prediction in cavitation post-treatment, Sci. Rep. 14 (1) (2024) 11004.

[38] M.Y. Shams, T. Abd El-Hafeez, E. Hassan, Acoustic data detection in large-scale emergency vehicle sirens and road noise dataset, Exp. Syst. Appl. 249 (B) (2024) 123608.

[39] D.A.A. Hady, T. Abd El-Hafeez, Revolutionizing core muscle analysis in female sexual dysfunction based on machine learning, Sci. Rep. 14 (1) (2024) 4795.

[40] E.H.I. Eliwa, A.M. El Koshiry, T. Abd El-Hafeez, H.M. Farghaly, Utilizing convolutional neural networks to classify monkeypox skin lesions, Sci. Rep. 13 (1) (2023) 14495.

[41] E. Hassan, T. Abd El-Hafeez, M.Y. Shams, Optimizing classification of diseases through language model analysis of symptoms, Sci. Rep. 14 (1) (2024) 1507.

[42] D.A.A. Hady, T. Abd El-Hafeez, Predicting female pelvic tilt and lumbar angle using machine learning in case of urinary incontinence and sexual dysfunction, Sci. Rep. 13 (1) (2023) 17940.

[43] H.M. Farghaly, M.Y. Shams, T.A. El-Hafeez, Hepatitis C Virus prediction based on machine learning framework: a real-world case study in Egypt, Knowl. Inf. Syst. 65 (6) (2023) 2595–2617.

[44] Y. Matsuo, Y. LeCun, M. Sahani, D. Precup, D. Silver, M. Sugiyama, E. Uchibe, J. Morimoto, Deep learning, reinforcement learning, and world models, Neur. Netw. 152 (2022) 267–275.

[45] Z.P. Liang, R.T. Yang, J.G. Wang, L. Liu, X.L. Ma, Z.X. Zhu, Dynamic constrained evolutionary optimization based on deep Q-network, Exp. Syst. Appl. 249 (2024) 123592.

[46] Y. Gu, Y.H. Cheng, K. Yu, X.S. Wang, Anti-martingale proximal policy optimization, IEEE Trans. Cybern. 53 (10) (2023) 6421–6432.

[47] B.Y. Zheng, S. Verma, J.L. Zhou, I.W. Tsang, F. Chen, Imitation learning: progress, taxonomies and challenges, IEEE Trans. Neural Netw. Learn. Syst 35 (5) (2024) 6322–6337.

[48] S.I.H. Shah, A. Coronato, M. Naeem, Learning and assessing optimal dynamic treatment regimes through cooperative imitation learning, IEEE Access 10 (2022) 78148–78158.

[49] Z.P. Tan, Y. Tang, K.S. Li, H.S. Huang, S.M. Luo, Differential evolution with hybrid parameters and mutation strategies based on reinforcement learning, Swarm Evol. Comput. 75 (2022) 101194.

[50] B. Romanowski, A. Ben Abacha, Y.F. Fan, Extracting social determinants of health from clinical note text with classification and sequence-to-sequence approaches, J. Am. Med. Inform. Assoc. 30 (8) (2023) 1448–1455.

[51] M.G.P. de Lacerda, F.B.D. Neto, T.B. Ludermir, H. Kuchen, Out-of-the-box parameter control for evolutionary and swarm-based algorithms with distributed reinforcement learning, Swarm Intell. 17 (3) (2023) 173–217.

[52] C.B. Dong, D.Z. Li, Adaptive evolutionary reinforcement learning with policy direction, Neur. Process. Lett. 56 (2) (2024) 69.

[53] S.Q. Sun, H.C. Dong, T.B. Li, A modified evolutionary reinforcement learning for multi-agent region protection with fewer defenders, Compl. Intell. Syst. 10 (3) (2024) 3727–3742.

[54] S.J. Li, W.Y. Gong, L. Wang, Q. Gu, Evolutionary multitasking via reinforcement learning, IEEE Trans. Emerg. Top. Comput. Intell. 8 (1) (2024) 762–775.

[55] R. Li, L. Wang, W.Y. Gong, J.F. Chen, Z.X. Pan, Y.T. Wu, Y. Yu, Evolutionary computation and reinforcement learning integrated algorithm for distributed, Eng. Appl. Artif. Intell. 135 (2024) 108775.

[56] P. Liang, Y.T. Chen, Y.F. Sun, Y. Huang, W. Li, An information entropy-driven evolutionary algorithm based on reinforcement learning for many-objective optimization, Exp. Syst. Appl. 238 (2024) 122164.

[57] J.Y. Zeng, P. Lu, Y. Wei, X. Chen, K.B. Lin, Deep reinforcement learning based medical supplies dispatching model for major infectious diseases: case study of COVID-19, Oper. Res. Perspect. 11 (2023) 100293.

[58] S.S. Saranya, P. Anusha, S. Chandragandhi, O.K. Kishore, N.P. Kumar, K. Srihari, Enhanced decision-making in healthcare cloud-edge networks using deep reinforcement and lion optimization algorithm, Biomed. Signal Process. Control. 92 (2024) 105963.