Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A New Multi-Objective Optimization of Regenerative Clausius Rankine Cycle with Two Feedwater Heaters

Amir Basiriparsa

amirbparsa@hut.ac.ir

Hamedan University of Technology

MohammadJavad Faraji

Hamedan University of Technology

Alireza Kokabi

Hamedan University of Technology

Research Article

Keywords: Deep reinforcement learning, Soft actor-critic, Exergy efficiency, Thermal efficiency, Feed water heater, Optimization

Posted Date: September 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10784481/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A New Multi-Objective Optimization of Regenerative Clausius Rankine Cycle with Two Feedwater Heaters

Amir Basiriparsa*

Department of Mechanical Engineering, Hamedan University of Technology, Hamedan, Iran.

* [Corresponding author] Email: amirbparsa@hut.ac.ir

MohammadJavad Faraji

Department of Electrical Engineering, Hamedan University of Technology, Hamedan, Iran.

Email: javadfaraji9877@gmail.com

Alireza Kokabi

Department of Electrical Engineering, Hamedan University of Technology, Hamedan, Iran.

Email: alireza.kokabi@hut.ac.ir

Abstract

This paper presents a novel intelligent framework for autonomous operational adaptation in power plants, leveraging advanced deep reinforcement learning (DRL) techniques. Specifically, a multi-objective optimization approach is developed to analyze and enhance the performance of a regenerative Rankine cycle equipped with two feedwater heaters. The framework employs a soft actor-critic (SAC) algorithm that simultaneously targets the maximization of specific work output, thermal efficiency, and exergy efficiency. These objectives are dynamically optimized with minimal sensitivity to environmental variations or plant-specific constraints, achieving fluctuation margins below 0.58%. This robustness enables consistent high-performance operation across diverse geographical locations, equipment configurations, and seasonal conditions, making the solution broadly applicable to modern power generation systems. The proposed SAC-based controller demonstrates superior adaptability and precision in real-time decision-making, ensuring continuous optimization without manual intervention. As a result, the system achieves optimized values of 38.99% for thermal efficiency, 82.35% for exergy efficiency, and 888.2 kJ/kg for specific work. These findings highlight the potential of intelligent control strategies in improving energy conversion processes and operational resilience in thermal power plants. The framework sets a foundation for future integration of AI-driven optimization in sustainable energy systems.

Keywords: Deep reinforcement learning; Soft actor-critic**;** Exergy efficiency; Thermal efficiency; Feed water heater; Optimization.

Nomenclature

a_t	mathematical expectation at step t
c	the constants utilized in the reward function
e	specific flow exergy, (kJ/kg)
f	Ratio of mass flow rate
h	specific enthalpy, (kJ/kg)
i	specific exergy loss, (kJ/kg)
k	the constants utilized in the reward function
$\dot{m}$	mass flow rate, (kg/s)
P	pressure, (kPa)
q	specific heat transfer, (kJ/kg)
s	specific entropy, $(kJ/kg\,K)$
s_t	mathematical expectation at step t
T	temperature, $(^oC$ or $K)$
v	specific volume, (m^3/kg)
w	specific work, (kJ/kg)

Greek letters

γ	discount factor
η	efficiency

Subscript, Superscript

g	gas (heat source in steam generator)
in	inlet
out	outlet
s	isentropic
t	time
w	water (cooling fluid in condenser)
0	dead state

Abbreviations

ANN	Artificial Neural Network
CHP	Combined Heat and Power
DRL	Deep Reinforcement Learning
EES	Engineering Equation Solver
MDP	Markov Decision Process
ORC	Organic Rankine Cycles
PID	Proportional-Integral-Derivative
SAC	Soft Actor-Critic
WHR	Waste Heat Recovery

1. Introduction

One of the biggest crises currently facing the world is the energy crisis and environmental issues. Therefore, increasing the efficiency of energy production and reducing environmental pollutants in today's world is one of the most important concerns of engineers around the world. One of the reasons for environmental pollution is using of fossil fuels. In addition to causing pollution due to the fuel process, the use of these fuels also leads to a crisis in energy consumption due to severe heat loss [1]. The Rankine cycle is the simplest thermodynamic power cycle that produces mechanical energy from thermal energy as an input. This cycle consists of four processes: isentropic compression by an ideal pump, isobar heat transfer in the evaporator, isentropic expansion in the turbine and finally isobar heat transfer in the evaporator [2].

Therefore, one of the useful ways to reduce energy consumption is to use waste heat to produce energy. One of the uses of this waste heat is in the Clausius Rankine cycle. In order to evaluate regenerative Clausius Rankine cycles, it is necessary to investigate the effect of various parameters on the overall thermal efficiency of the cycle, the cycle total irreversibility, the amount of waste heat required to operate the cycle and the second law efficiency of the system. Various works have been done in this area, including Bejan et al. [3], Moran and Shapiro [4], and others. One of the main challenges in regenerative Clausius Rankine cycles involves obtaining a suitable heat source that can provide the required heat in the steam generator. The suitability of the low-temperature cycle of a vapor power-plant based on an organic fluid for cooling technical equipment that requires cooling during operation, e.g. power transformers, has been presented by Nowak et al. [5]. Also Fischer [6] has done a comparison of optimized trilateral cycle systems with water as working fluid and optimized Organic Rankine Cycle systems (ORC) with pure organic working fluids. In addition, Yamamoto et al. [7] presented a study to find the optimum conditions needed

for the working fluid to yield higher turbine output of ORC from a theoretical stand point. The parametric analysis of regenerative Clausius and ORCs with two feedwater heaters for water and R717 as well as their optimization with a new procedure based on ANN and ABC has been done in another study by Rashidi et al. [8].

The objective of this paper is the evaluation of the performance of a regenerative Clausius Rankine cycle with two feed water heaters, using a combination of first and second law analysis and optimizing them.

There are some studies about application of Artificial Neural Network (ANN) in thermodynamics. Deep reinforcement learning (DRL) has emerged as a promising and powerful approach for addressing complex decision-making and control problems across various systems by combining deep learning and reinforcement learning to tackle challenges associated with increasing complexity, uncertainty, and high-dimensional data [9, 10]. Its effectiveness has been particularly notable in optimizing energy systems, including combined cooling, heating, and power systems with renewable energy and storage [11], and integrated energy systems incorporating wind power and power-to-gas technology [12]. DRL algorithms, such as the Deep Q-Network (DQN), actor-critic (A2C), Proximal Policy Optimization (PPO), and soft actor-critic (SAC) has demonstrated superior performance compared to traditional control methods such as proportional-integral-derivative (PID) controllers or optimization techniques such as Particle Swarm Optimization. Studies underscore DRL's advantages, including its ability to handle complex nonlinear systems with uncertainties, its model-free framework, and its capacity for real-time decision-making [13]. Recent applications, such as optimizing nuclear-renewable integrated energy systems, highlight DRL's potential to enhance efficiency, reduce costs, and improve flexibility. Algorithms such as PPO outperform conventional methods in maximizing the revenue

from electricity and hydrogen production [14]. Its versatility has been demonstrated in applications such as a DRL optimization framework using an A2C agent to balance power generation and NOx emissions in fluidized bed power plants, achieving superior performance with reduced computational costs [15].

DRL has shown remarkable potential in optimizing and controlling complex thermodynamic systems such as Organic Rankine Cycles (ORC) and regenerative Clausius cycles, addressing the challenges posed by system complexity, nonlinearity, and dynamic operating conditions. Applications range from enhancing heat engine efficiency and minimizing entropy production to optimize energy management in power stations [16]. DRL has effectively identified thermodynamic trajectories for maximal efficiency, executing established cycles, such as Carnot and Otto, while discovering novel cycles under irreversible processes [17]. Notable applications include optimizing single mixed refrigerant processes for natural gas liquefaction using the Deep Q-Network (DQN) algorithm, which significantly reduces the specific power requirements while maintaining heat transfer constraints [18]. In combined heat and power (CHP) systems, DRL has facilitated economic dispatch with adaptability to vary scenarios, thereby reducing computational complexity without sacrificing accuracy [19]. For ORC waste heat recovery (WHR) systems in heavy-duty vehicles, DRL-trained control strategies excel in handling transient dynamic regimes, thereby enhancing performance across drive cycles [20]. Safe reinforcement learning techniques have further ensured stability and effective set-point tracking in ORC systems by addressing the inherent nonlinearity and control constraints. Zhang [21] develops a method combining surrogate models with Sim2Real transfer learning to optimize superheat control in ORC using DRL, improving real-world deployment by correcting model inaccuracies in dynamic thermal systems. Moreover, DRL applications in thermal power system scheduling have reduced fuel

consumption and improved efficiency by employing methods that either directly learn control strategies or optimize traditional PID controllers to achieve superior tracking performance under transient heat sources [22].

In this study, the specific network, thermal efficiency and exergy efficiency of regenerative Clausius Rankine cycle with two feed water heaters are selected as objective functions for parametric optimization. In order to optimize these functions, a new procedure based on the DRL is proposed. A soft actor-critic (SAC) dependent of the specific work, thermal efficiency and exergy efficiency is implemented to dynamically maximize the mentioned parameters. The SAC can handle these parameters almost independent of prevailing environmental conditions or plant characteristics with negligible fluctuations.

2. Thermodynamic model

Fig. 1 shows a simple schematic of a regenerative Rankine cycle with two feedwater heaters. The T-s and P-h diagrams are also shown in Figs. 2a and 2b, respectively. The main components of this cycle include three pumps, a steam generator, three turbines, a condenser and two feedwater heaters. For the present study, the heat source is an industrial gas idealized as air.

The following assumptions are made in the study of this cycle: All components are considered as a steady-state steady-flow system, heat losses, friction, and changes in kinetic and potential energies are neglected. The condenser and feedwater heater outlets are assumed to be saturated liquid. The maximum cycle pressure (P_6) is related to the outlet pressure of the first pump.

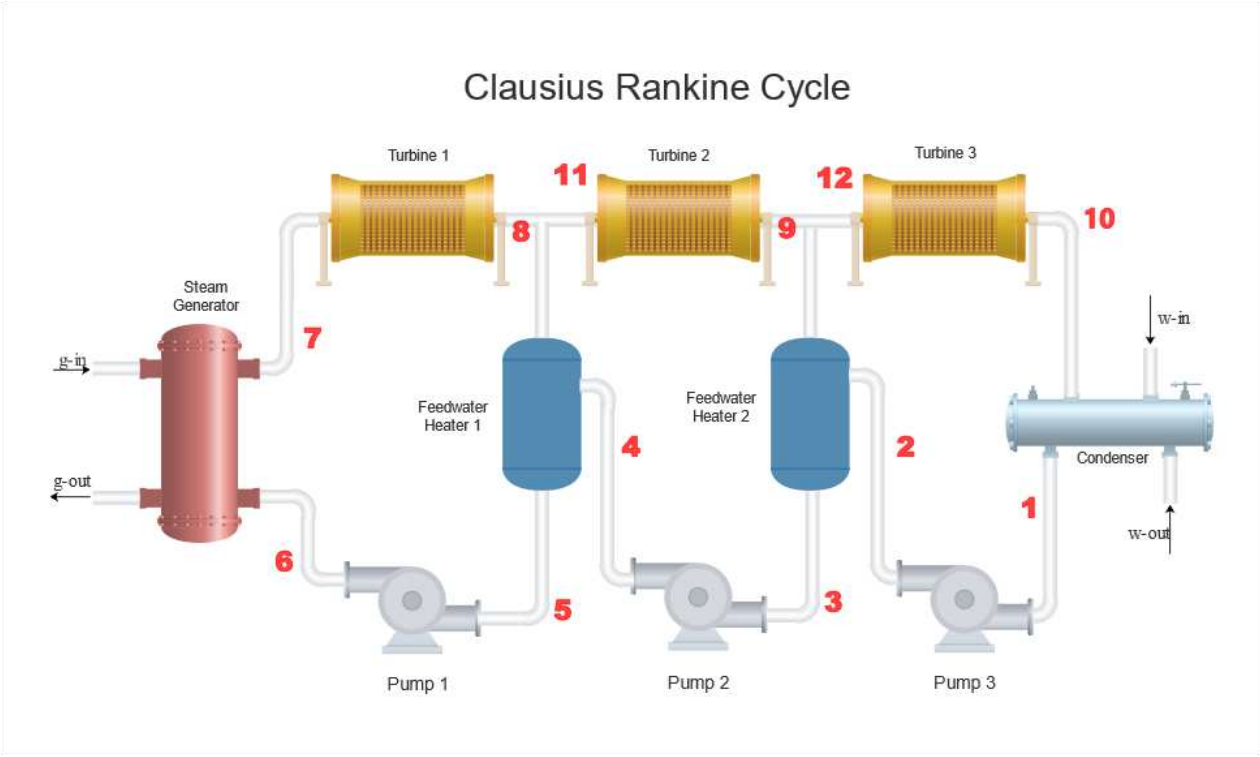

Fig. 1: Simple schematic of the regenerative Rankine cycle with two feed water heaters.

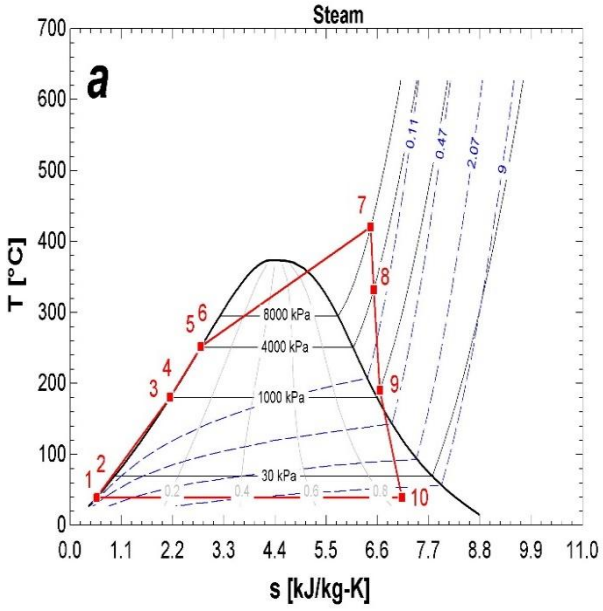

Fig. 2a: Schematic cycle T–S chart.

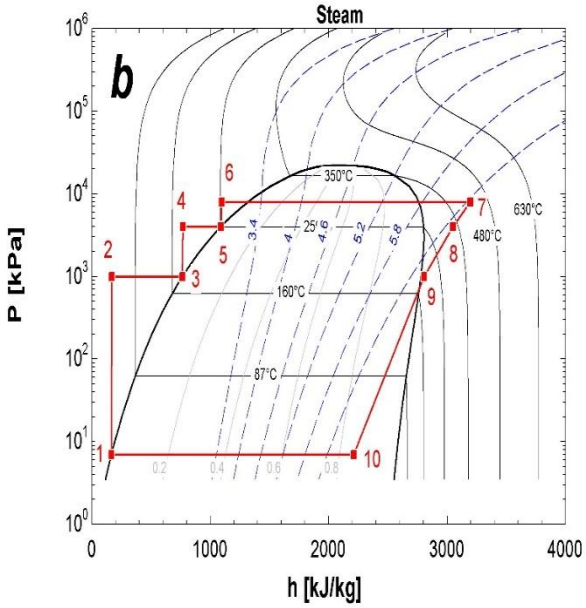

Fig. 2b: Schematic cycle P-h chart.

Since the boiler is a device in which high-pressure and high-temperature steam is produced by extracting heat from an industrial gas source, the maximum cycle temperature (T_7) is correspond to the outlet temperature of the steam generator.

Power is generated by passing high-pressure, high-temperature steam through a turbine and expanding it. Since the main goal of this study is to analyze and optimize the Clausius-Rankin cycle, water has been chosen as the main working fluid. One of the components of power cycles that is used to increase the efficiency of the cycle is the feedwater heater. This component is used to preheat the water delivered to the boiler. Preheating the feedwater reduces the irreversibility in steam production and therefore improves the thermodynamic efficiency of the system. It also prevents thermal shock to the boiler metal when the feedwater returns to the steam cycle. In a modified Rankine cycle steam plant, feedwater heaters allow the feedwater to gradually reach saturation temperature. This minimizes the associated irreversibility with heat transfer. The energy used to heat the feedwater is derived from the steam extracted between the stages of the steam turbine. Therefore, the steam used to do the expansion work in the turbine is not used for this purpose. The percentage of the total mass flow rate used for the feedwater heater is called the extraction fraction and must be carefully optimized for maximum thermal efficiency of the plant, as increasing this fraction reduces the turbine output power.

To optimize this cycle, changes in three parameters are considered: the outlet pressure of the third pump (P_2), the outlet pressure of the second pump (P_4), the outlet pressure of the first pump (P_6). The thermodynamic analysis is performed with EES. This software [23] has considerable power in thermodynamic analyses. In the next section, the thermodynamic equations of the cycle analysis are introduced.

The analysis begins by applying the first law of thermodynamics to the system, which results in the extraction of the specific net output power of the cycle and the thermal efficiency. With the assumptions made earlier, their values depend on three independent parameters: the outlet pressures of the third, second, and first pumps.

In this research, almost all the physical parameters affecting the cycle will be evaluated. the specific net work, thermal efficiency and exergy efficiency are considered as the final objective functions.

Of course, it is clear that the changes of some parameters are beyond our control, so from all the parameters the outlet pressure of the third pump (P_2), the outlet pressure of the second pump (P_4) and the outlet pressure of the first pump (P_6) are selected as the main variables for analysis and optimization. For the different components of the cycle and based on the first law of thermodynamics, it can be written:

Pumps:

$$w_{Pump} = v\left[\frac{P_{in} - P_{out}}{\eta_{Pump}}\right] \Rightarrow h_{out} = h_{in} - w_{Pump}. \tag{1}$$

Turbines:

$$w_{Turbine} = \left[h_{in} - h_{out,s}\right]\eta_{Turbine} \Rightarrow h_{out} = h_{in} - w_{Turbine}. \tag{2}$$

Steam Generator:

$$\begin{cases} q_{Steam\ Generator} = \dot{m}_g(h_{g-in} - h_{g-out}) = \dot{m}_5(h_7 - h_6), \\ \text{Let define: } \dfrac{\dot{m}_g}{\dot{m}_5} = f_g, \end{cases} \Rightarrow f_g(h_{g-in} - h_{g-out}) = (h_7 - h_6).$$

$$q_{Steam\ Generator} = h_7 - h_6. \tag{3}$$

Firs feedwater heater:

$$\begin{cases} \dot{m}_8 h_8 + \dot{m}_4 h_4 = \dot{m}_5 h_5, \\ \text{Let define: } \dfrac{\dot{m}_8}{\dot{m}_5} = f_1, \end{cases} \Rightarrow f_1 h_8 + \left[1 - f_1\right] h_4 = h_5. \tag{4}$$

Second feedwater heater:

$$\begin{cases} \dot{m}_9 h_9 + \dot{m}_2 h_2 = \dot{m}_3 h_3, \\ \text{Let define: } \dfrac{\dot{m}_9}{\dot{m}_3} = f_2, \end{cases} \Rightarrow f_2 h_9 + \left[1 - f_2\right] h_2 = h_3. \tag{5}$$

Condenser:

$$\begin{cases} q_{Condenser} = \dot{m}_w (h_{w-out} - h_{w-in}) = \dot{m}_1 (h_{10} - h_1), \\ \text{Let define: } \dfrac{\dot{m}_w}{\dot{m}_5} = f_w, \end{cases} \Rightarrow f_w (h_{w-out} - h_{w-in}) = (1 - f_1 - f_2)(h_{10} - h_1). \tag{6}$$

Specific net work:

$$w_{Total-T} = w_{Turbine1} + \left[1 - f_1\right] w_{Turbine2} + \left[(1 - f_1)(1 - f_2)\right] w_{Turbine3}, \tag{7}$$

$$w_{Total-P} = w_{Pump1} + \left[1 - f_1\right] w_{Pump2} + \left[(1 - f_1)(1 - f_2)\right] w_{Pump3}, \tag{8}$$

$$w_{net} = w_{Total-T} - w_{Total-P}. \tag{9}$$

Thermal efficiency of the cycle:

$$\eta_{thermal} = \frac{w_{net}}{q_{Steam\ Generator}}. \tag{10}$$

After analyzing the first law of thermodynamics for the components and the entire cycle in question, in order to calculate the exergy efficiency of this cycle, the specific flow exergy in the various components of the cycle is first calculated. This value is obtained by the following expression:

$$e = \left[\left(h - h_0\right) - T_0\left(s - s_0\right)\right], \tag{11}$$

The subscript 0 represents the reference state. In this study, this state is assumed to be at $20^o C$ and $101\,Kpa$. For each component of this cycle, the specific exergy loss is calculated:

Steam Generator:

$$i_{Steam\ Generator} = \left[e_6 - e_7\right] + f_g\left[e_{g-in} - e_{g-out}\right], \tag{12}$$

Turbines:

$$i_{Turbine-1} = \left[e_7 - e_8 - w_{Turbine-1}\right], \tag{13}$$

$$i_{Turbine-2} = \left(1-f_1\right)\left[e_8 - e_9 - w_{Turbine-2}\right], \tag{14}$$

$$i_{Turbine-3} = \left(1-f_1\right)\left(1-f_2\right)\left[e_9 - e_{10} - w_{Turbine-3}\right]. \tag{15}$$

Condenser:

$$i_{Condenser} = \left(1-f_1\right)\left(1-f_2\right)\left[e_{10} - e_1\right] + f_w[e_{w-in} - e_{w-out}]. \tag{16}$$

Pumps:

$$i_{Pump-1} = \left[e_5 - e_6 - w_{Pump-1}\right], \tag{17}$$

$$i_{Pump-2} = \left(1-f_1\right)\left[e_3 - e_4 - w_{Pump-2}\right], \tag{18}$$

$$i_{Pump-3} = \left(1-f_1\right)\left(1-f_2\right)\left[e_1 - e_2 - w_{Pump-3}\right]. \tag{19}$$

Feedwater heaters:

$$i_{Feedwater-1} = f_1 e_8 + \left(1-f_1\right)e_4 - e_5, \tag{20}$$

$$i_{Feedwater-2} = f_2\left(1-f_1\right)e_9 + \left(1-f_1\right)\left(1-f_2\right)e_2 - \left(1-f_1\right)e_3, \tag{21}$$

Therefore the total specific exergy loss is given as

$$\begin{aligned}\sum i = {} & i_{Turbine-1} + i_{Turbine-2} + i_{Turbine-3} + i_{Condenser} + i_{Pump-1} + \\ & i_{Pump-2} + i_{Pump-3} + i_{Feedwater-1} + i_{Feedwater-2} + i_{Steam\ Generator}.\end{aligned} \tag{22}$$

Finally, the overall exergetic efficiency is expressed by

$$\eta_{exergy} = \frac{f_{g,in}e_{g,in} - \sum i}{f_{g,in}e_{g,in}}. \tag{23}$$

3. Parametric Optimization Methodology

In this study, three main parameters including specific network, thermal efficiency and exergy efficiency are used as objective functions for parametric optimization. In order to optimize these functions, a new procedure based on deep reinforcement learning algorithms is proposed.

Reinforcement learning involves the use of an agent that acquires knowledge through environmental interactions. The agent receives the environment state, performs an action, and obtains a numerical reward at each time interval. Through this process, the agent develops a policy π that maps the states to action probability distributions. Deep neural networks serve as function approximators, enabling the learning of intricate behaviors and adaptation to extensive state-action spaces. Training of the policy and value functions occurs by minimizing a loss function using stochastic gradient descent. Model-free DRL algorithms have been demonstrated for various challenging decision-making and control tasks [24]. SAC, an off-policy model-free actor-critic DRL algorithm, is a well-known method based on the maximum entropy RL framework. SAC is achieved by combining off-policy updates with a stable stochastic actor-critic formulation. It provides sample efficiency and stability and is well-suited for continuous control problems. It trains a stochastic neural network policy by updating a state-action Q-value function and the policy π itself. SAC samples experience tuples of (state, action, reward, and next state) from a replay buffer to update the Q-value by minimizing the Bellman equation. The value function of SAC can be described as follows:

$$Q^*(s_t, a_t) = r(s_t, a_t) + \gamma \mathrm{E}_{(s_{t+1}, a_{t+1})}\{Q(s_{t+1}, a_{t+1})\} \tag{24}$$

where γ is the discount factor and s_t, a_t represent the mathematical expectation at step t. Policy π is updated by maximizing the expected Q-values, while maximizing the entropy of the policy to encourage exploration. This entropy maximization ensures sufficient randomness in the action

selection. It deploys the target networks when updating the Q-values to improve stability. The policy converges to the optimal stochastic policy that maximizes the expected return while balancing exploration via entropy regularization. The structure of the neural networks learning the model is shown in Fig. 3.

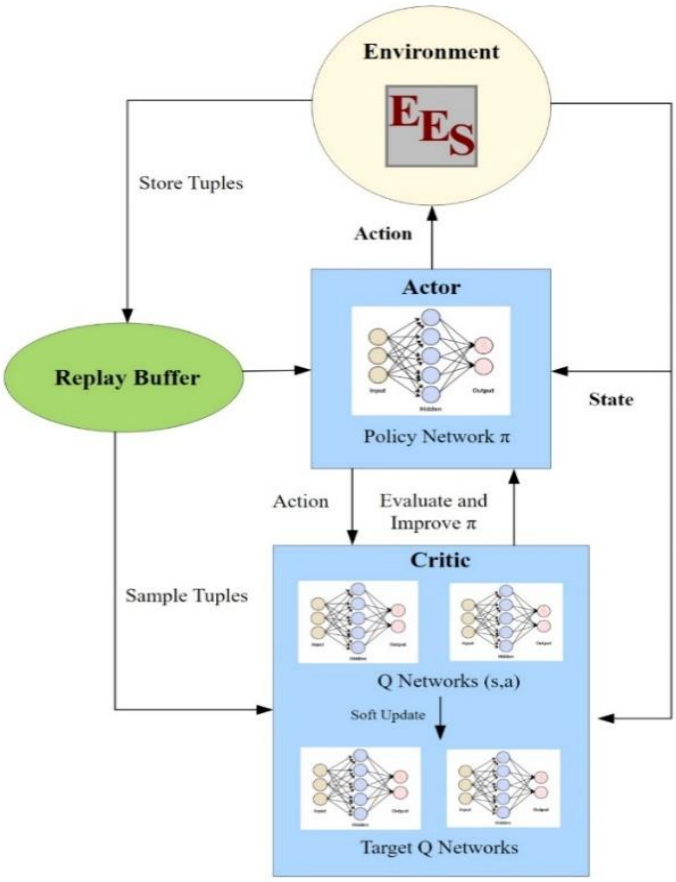

Fig. 3: The structure of the model of neural networks learning.

The actor aims to maximize the expected reward while maximizing entropy, and optimizes the following objective function:

$$J(\pi)=\sum_{t=0}^{T}\mathrm{E}_{(s_t,a_t)\sim\rho_\pi}\left[r\left(s_t,a_t\right)+\alpha\mathrm{H}\left(\pi\left(\cdot\middle|s_t\right)\right)\right] \tag{25}$$

where r is the reward obtained when the agent executes the action at in state s_t, π is the policy, ρ_π is the joint distribution over states and actions induced by policy π, α is the weight of the entropy term and although known as the trade-off coefficient. H is the entropy which is obtained according to Eq (26). The main objective of SAC is to determine the optimal policy that maximizes the entropy-augmented objective function $J(\pi)$ [25].

$$\mathrm{H}\left(\pi\left(\cdot|s_t\right)\right) = -\mathrm{E}_\pi\left[\log\pi\left(\cdot|s_t\right)\right] \tag{26}$$

Also the hyperparameters of SAC are shown in Table 1.

Table 1: The hyperparameters of SAC.

Hyperparameters	SAC
Learning Rate	0.0003
Discount Factor	0.99
Replay Buffer Size	1,000,000
Batch Size	256
Polyak Coefficient	0.005
Actor-Critic ANN Architecture	[256,256] [256,256], [256,256]
Activation Function of Each Layer	ReLU()
Steps Per Episode	10

Formulation of Markov Decision Process (MDP)

The Clausius Rankine cycle optimization problem should be formulated as an MDP when implementing the DRL algorithm to solve this problem. The key concepts of environment, episode, state, action, and reward are defined below:

1) Environment: In the proposed MDP framework, the environment is modeled as the ORC with two feedwater heaters. This cycle is implemented within the EES software, while the DRL model, developed in Python, facilitates a bidirectional interaction with the EES platform. The model's actions are provided as inputs to the cycle equations in EES, which are then solved to produce the state vector, encapsulating critical output data. This state vector is subsequently fed back into the DRL model, enabling the computation and application of the reward function. Upon completion of the training phase, the model is capable of autonomously recommending and applying optimal operating points under varying real-world conditions.

2) Episode: The episode is a sequence of steps in which the agent interacts with the environment in response to a specific power station condition. Each exploration step is also referred to as iteration. During an episode, the agent explores by suggesting actions and receiving states. Afterward, at each step, the reward is calculated until reaching the maximum number of steps. During the training phase, each episode begins with the initialization of stochastic variable values, randomly assigned within the specified ranges outlined in Table 2. These values remain constant throughout the episode and are selected based on their higher probability of occurrence relative to other values, ensuring generalizability to any range. Throughout the training process, the model is designed to learn and identify the optimal operating points of the power plant under varying operational conditions.

Table 2: The initialization of stochastic variable values.

Stochastic Variables	Ranges
Environment Temperature (°C)	(15,35)
Inlet Temperature (°C)	(400,450)

The ratio of heat source mass flow rate to basic mass flow rate	(6,10)
Pump Efficiency (%)	(60,90)
Turbine Efficiency (%)	(60,90)

3) State: In DRL, the state vector represents a comprehensive set of features describing the environment at a given time, capturing all relevant information necessary for decision-making. This vector serves as the input to the agent's neural network, enabling it to predict optimal actions based on the current state and learned policies. In the proposed framework the state S is defined as a vector containing cycle information, including w_{net}, $\eta_{Thermal}$, η_{Exergy}, f_1, f_2, f_g, f_w, T_0, T_{in}, η_{Pumps}, $\eta_{Turbines}$.

4) Action: The pressure operating points of the pumps, denoted as P_2, P_4 and P_6 are defined as the action variables within the proposed DRL framework. The model is tasked with recommending operating points across varying conditions to optimize the cycle's performance and maximize overall efficiency. The permissible ranges for these pressure points are as follows: P_2 ranges from 0 to $2\,Mpa$. P_4 from 0 to $8\,Mpa$ and P_6 from 0 to $10\,Mpa$. Additionally, the pressure P_2 must not exceed two other points, and P_4 must remain less than P_6. These constraints are enforced within the model via the reward function to ensure compliance with operational requirements.

5) Reward: The reward scheme needs to be carefully designed to achieve proper system performance. Given that the objective of the proposed framework is to maximize thermal efficiency, exergy efficiency, and w_{net}, the reward function is defined in the following penalized

form. This ensures that minimizing the reward function leads to the maximization of the aforementioned parameters.

$$r_t = \begin{cases} if \quad \eta_{thermal} > 1 \| \eta_{exergy} > 1 \Rightarrow -\sum k_1 \\ if \quad P_2 \geq P_4 \| P_2 \geq P_6 \| P_4 \geq P_6 \Rightarrow -\sum k_2 \\ Else \Rightarrow c_1 \times (\eta_{thermal} - 1) + c_2 \times (\eta_{exergy} - 1) + c_3 \times (W_{net}^{\frac{MJ}{kg}} - 1) \end{cases} \tag{27}$$

The reward function is meticulously designed to prevent the model from generating unrealistic or invalid outputs under any circumstances. It ensures that the calculated efficiencies remain below 100% and that upstream pressures within the cycle do not exceed downstream pressures. When these conditions are satisfied, the model receives its primary reward. This reward is defined as a penalty-based function that inherently avoids a zero value due to the inclusion of efficiencies in its formulation. By minimizing this penalty, the model seeks to maximize both the efficiencies and w_{net}.

The constants utilized in the reward function, detailed in Table 3, are weighted according to the relative significance of efficiencies and w_{net} to facilitates optimal action selection. A notable advantage of this model lies in its ability to identify optimal operating points across varying conditions with exceptional speed. After training, the model can adapt to changes in stochastic parameters and recommend the best operating points in under a second. This approach eliminates the iterative processes typical of traditional optimization algorithms, enabling rapid and efficient decision-making during execution.

Table 3: The constants utilized in the reward function.

Constants	k_1	k_2	c_1	c_2	c_3
Value	100	10	0.5	0.2	0.3

4. Verification

One of the important stages of a research is the validation and comparison of the obtained results with the results of other past studies. In the present study, in order to validate the obtained results for the conventional Rankine cycle, we compare the results of the previous study [7]. Therefore, for this comparison, in the initial cycle (Rankin cycle with two feedwater heaters), the effects of the feedwater heater are ignored and removed from the cycle. With this assumption, the results will be comparable to Ref [7]. This comparison can be seen in Fig. 4.

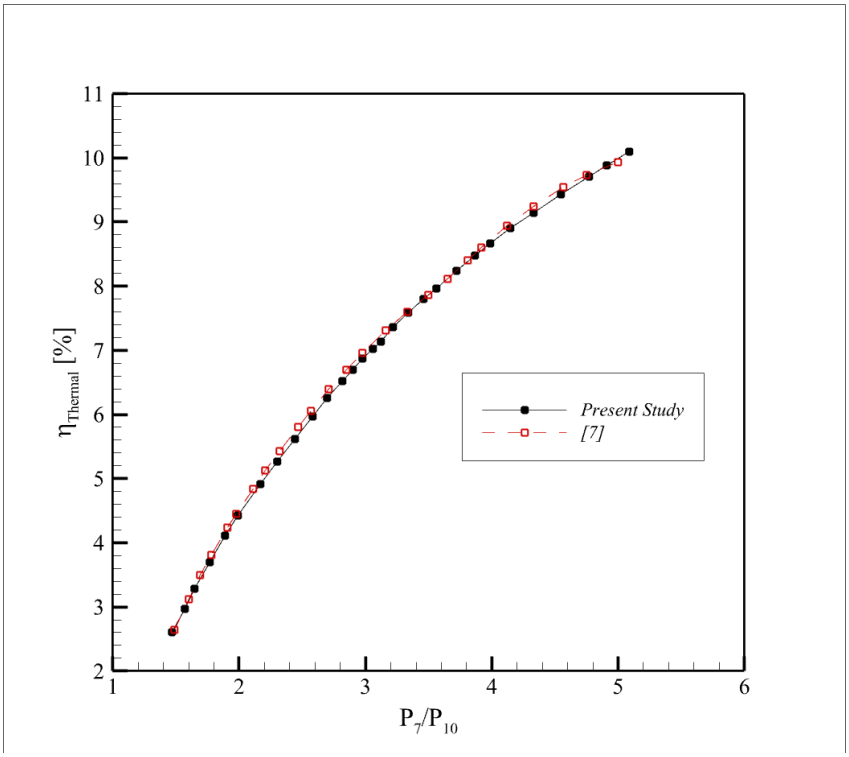

Fig. 4: Validation of the numerical model with the previously published data [7] for the ordinary Rankine cycle assuming working fluid as water.

As can be seen, there is a very good agreement between the results and this agreement indicates the accuracy of the calculations and the numerical code written in the software. It should be noted that in this comparison, the pressure ratio (P_7 / P_{10}) is considered as the independent

variable and the thermal efficiency as the dependent variable, and the other cycle parameters are adjusted according to Ref [7].

5. Simulation Results and discussion

In this section, all the results of the regenerative Rankine cycle with two feedwater heaters simulation, the process of learning the optimization method, and the obtained results from the optimization of this cycle will be presented and analyzed.

A. Learning rate of DRL

In this subsection, the learning rate and convergence rate of the proposed DRL are discussed. To evaluate the learning rate of the algorithm, Fig. 5 shows the cumulative reward trajectory in this method. In the DRL training process, the cumulative reward gradually increases in each iteration. In other words, in the exploration process, the agent gradually learns to take an action that can reduce the reward. The average cumulative reward graph can be seen in Fig. 5. This figure shows that with the increase in training steps, the average reward gradually increases. After training for about 1500 time steps, the reward value tends to stability and the average cumulative reward tends to zero. This indicates the convergence of the desired method and reaching a steady state.

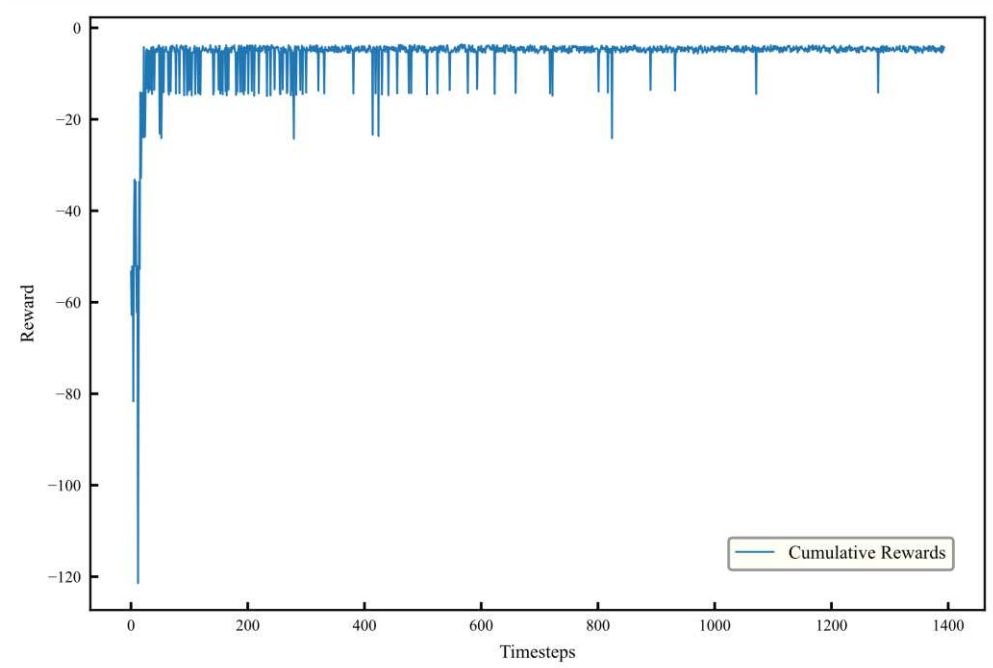

Fig. 5: The cumulative reward trajectory in DRL algorithm.

B. Optimizations results (by DRL)

Figs. 6-10 present the thermal efficiency and the exergy efficiency graphs, changes as stochastic and deterministic variables against changes in various parameters, including flow rate ratio, pump efficiency, turbine efficiency, environment temperature, and inlet temperature.

The variations in the thermal and exergy efficiencies as stochastic and deterministic variables with the ratio of heat source mass flow rate to basic mass flow rate $(\dot{m}_g / \dot{m}_5 = f_g)$ are presented in Fig. 6. According to this figure, it is clear that with an increase in this flow rate ratio, the thermal efficiency increases sharply and gradually initially, then when this ratio reaches a number of about 8, it reaches relative stability in the range of 38.6%. The trend of changes in the stochastic variable also confirms the same behavior. But in the case of exergy efficiency, with increasing flow rate ratio, there is a relatively sharp decrease at first until it reaches a minimum value of about 81.5%

and then increases. This means that changes in flow rate ratio lead to an extremum point in exergy efficiency.

In Fig. 7, changes in the thermal and exergy efficiencies as stochastic and deterministic variables are observed with the isentropic efficiency of the pumps. As the isentropic efficiency of the pumps increases, a completely opposite behavior occurs for the thermal and exergy efficiency of the cycle, in that the thermal efficiency decreases and the exergy efficiency increases, although these changes are very limited and about 0.1% to 0.2%. As a wise conclusion, the intersection of these two graphs can be introduced as an optimum point. The stochastic variables changes are interestingly similar to each other.

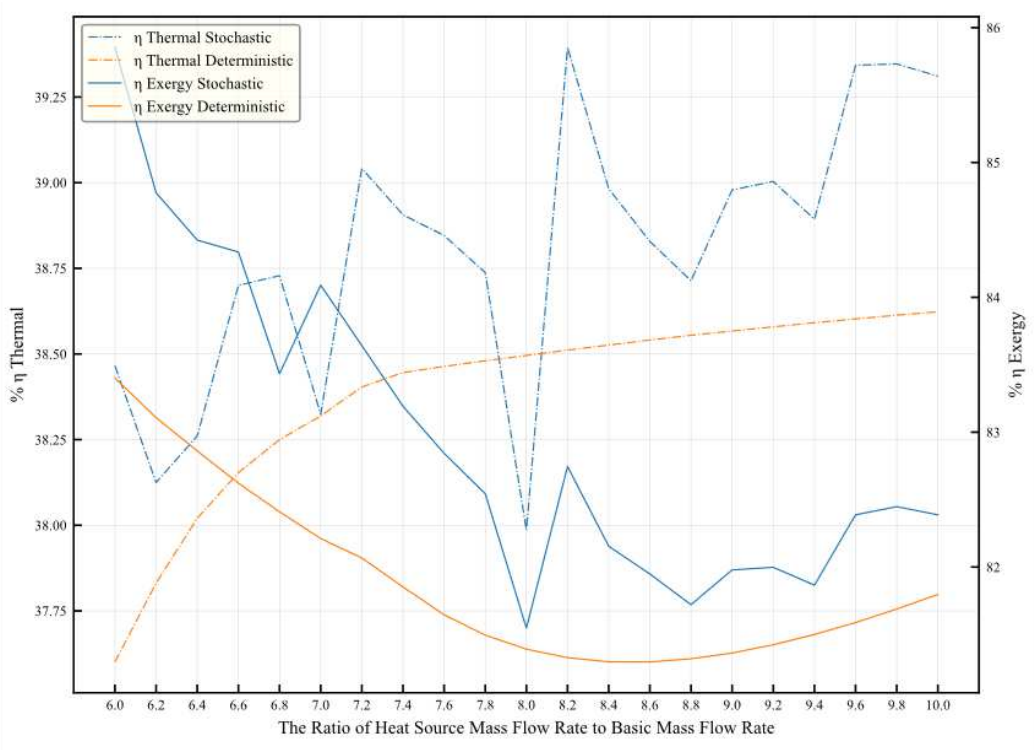

Fig. 6: The variation of the thermal and exergy efficiencies as stochastic and deterministic variables with f_g.

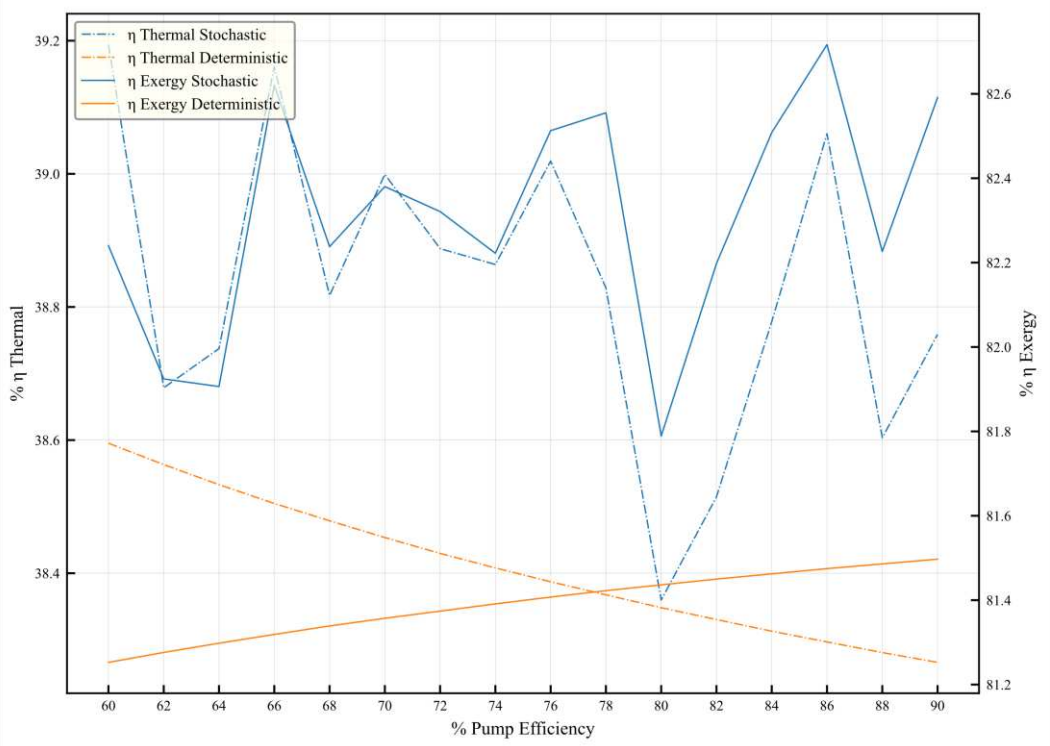

Fig. 7: The variation of the thermal and exergy efficiencies as stochastic and deterministic variables with η_{Pumps}.

Fig. 8 shows the variations in the thermal and exergy efficiencies as random and deterministic variables with the isentropic efficiency of the turbines. In contrast to the previous figure, it is clear that with increasing isentropic efficiency of the turbines, both thermal and exergy efficiency of the cycle increase sharply. With an increase of about 30% in isentropic efficiency of the turbines, there was about a 10% increase in thermal efficiency and a 20% increase in exergy efficiency.

One of the unique independent variables of this study is environment temperature. The effect of environment temperature in the thermal and exergy efficiencies as stochastic and deterministic variables is presented in Fig. 9. As the environment temperature increases from $15^{o}C$ to $25^{o}C$, the thermal efficiency decreases by nearly 0.4%. However, as a result of this change in the environment temperature, the exergy efficiency increases by less than 4%.

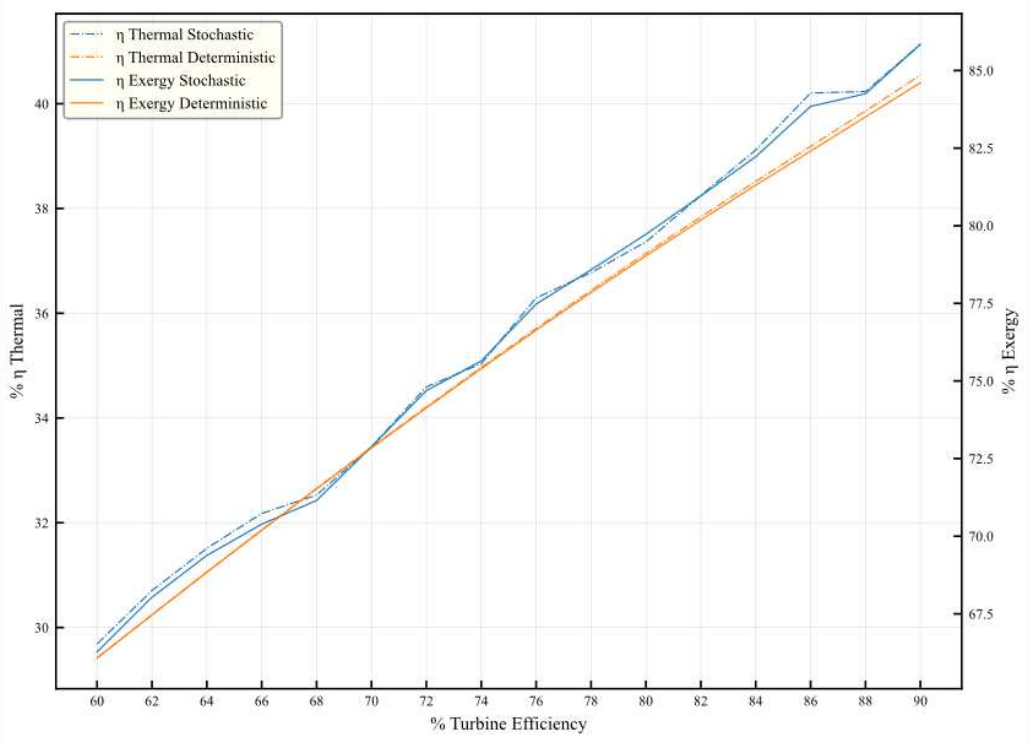

Fig. 8: The variation of the thermal and exergy efficiencies as stochastic and deterministic variables with $\eta_{Turbines}$.

Therefore, the behavior of thermal efficiency and exergy is completely opposite to each other. The interesting thing is that the changes in the stochastic variables are quite irregular and oscillating.

The last figure in this section shows the behavior of the thermal and exergy efficiencies as stochastic and deterministic variables due to changes in the inlet temperature of air heat source. From Fig. 10, it can be seen that with increasing this temperature, the thermal efficiency has a general increasing trend, which becomes more intense up to a temperature of $400^{o}C$ and then becomes much milder. Although this increase is generally not very significant. However, the behavior of the exergy efficiency curve is completely different. The changes in this parameter due

to increasing temperature are sharply reduced. Similar to the previous figure, changes in the stochastic variables are quite irregular and oscillating.

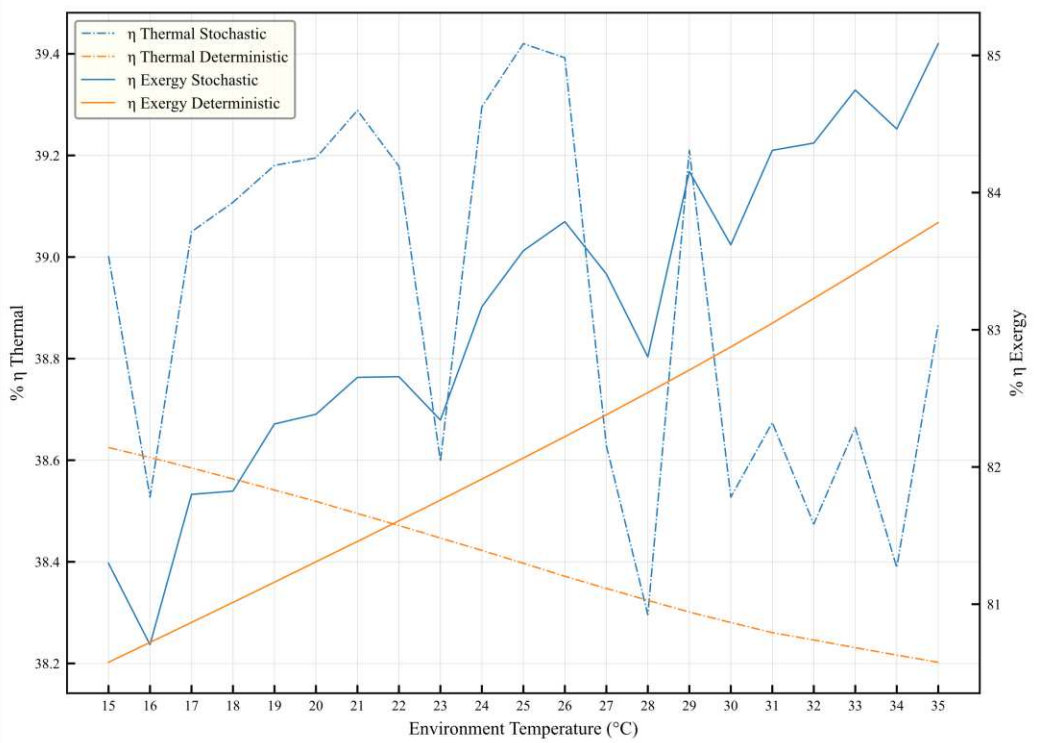

Fig. 9: The variation of the thermal and exergy efficiencies as stochastic and deterministic variables with T_0.

In another study, previously trained DRL neural networks were used to plot a continuous 3D layer of important cycle parameters including specific network, thermal efficiency and exergy efficiency against the outlet pressure of third pump (P_2), the outlet pressure of second pump (P_4) and the outlet pressure of first pump (P_6). The trained neural network was used to accurately predict the data at points where EES data was not available. Therefore, in addition to the EES data, data was obtained and then used to draw the layers. Thus, the layers are drawn with more accuracy.

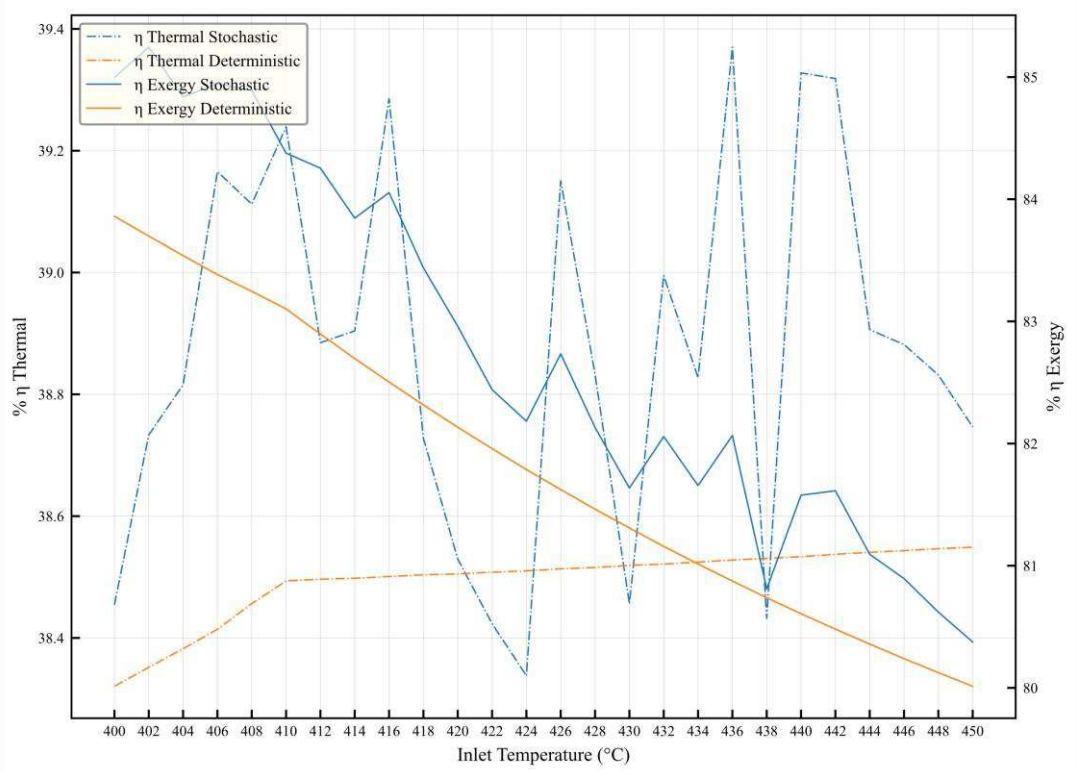

Fig. 10: The variation of the thermal and exergy efficiencies as stochastic and deterministic variables with T_{g-in}.

The continuous 3D layer of specific network (w_{net}) of cycle for changes in P_2, P_4 and P_6 can be seen in Fig. 11. A more precisely investigation of this figure shows that the specific network decreases at a significant slope as outlet pressure of the third pump increases. Also, increasing outlet pressure of the second pump leads to a decrease in the specific network of cycle. But increasing outlet pressure of the first pump pressure results an increase in the specific network of cycle. In other words, by reducing the pressure ratio of each of these three pumps, the amount of specific network of the cycle decreases and vice versa. Therefore, one way to increase this specific network is to increase the pressure ratio of the pumps. However, it is clear that increasing the pump pressure ratio excessively will involve justifiable costs and this ratio cannot be increased excessively without study.

The variation of thermal efficiency with pressure values P_2, P_4 and P_6 is presented in Fig. 12. In this 3D figure, it is clear that the thermal efficiency has a global maximum value between the upper and lower boundaries P_2 and P_4. But with increasing the output pressure of the first pump, the thermal efficiency value continuously increases. This can be a noteworthy result because by simultaneously considering Figs. 11, 12, it can be concluded that increasing P_6 simultaneously increases the work and efficiency values. However, as mentioned earlier, excessively increasing the outlet pressure of the first pump (meaning increasing the pump pressure ratio) is not possible due to economic and technical considerations.

Finally, Fig. 13 uses the results of the trained artificial neural network to plot a continuous 3D plot of exergy efficiency versus the outlet pressure of third pump (P_2), the outlet pressure of second pump (P_4) and the outlet pressure of first pump (P_6).

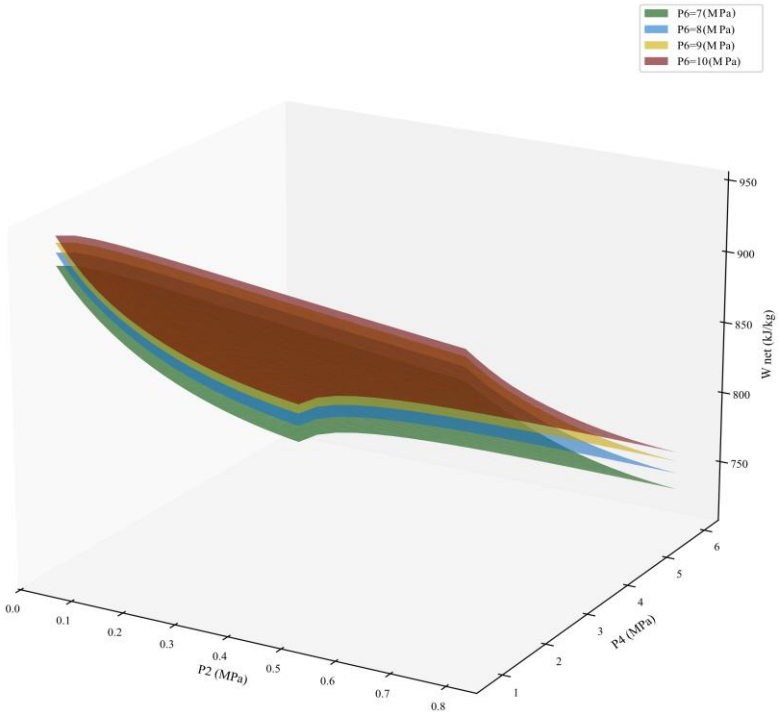

Fig. 11: The continuous 3D layer of w_{net} for changes in P_2, P_4 and P_6.

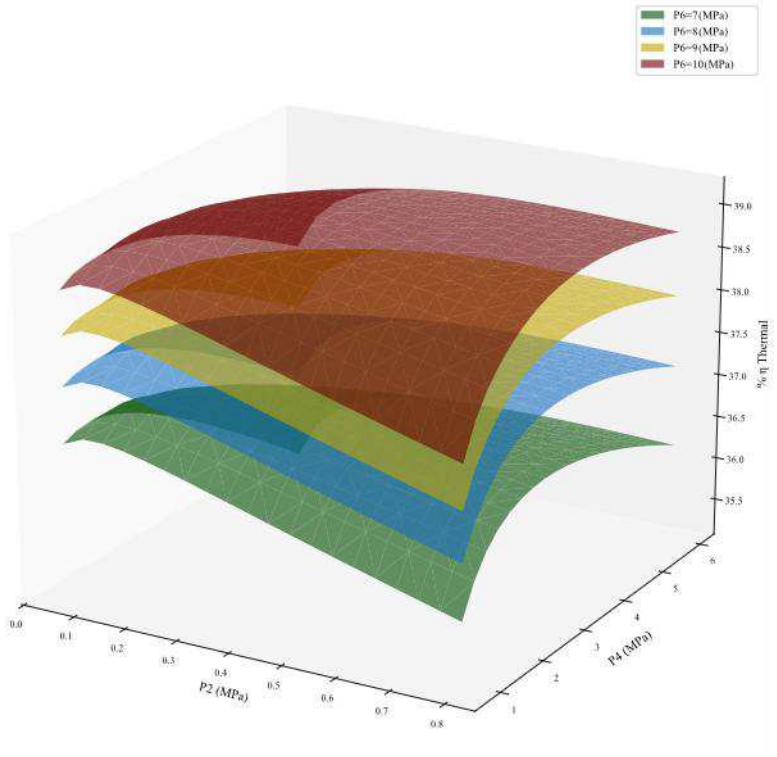

Fig. 12: The continuous 3D layer of $\eta_{Thermal}$ for changes in P_2, P_4 and P_6.

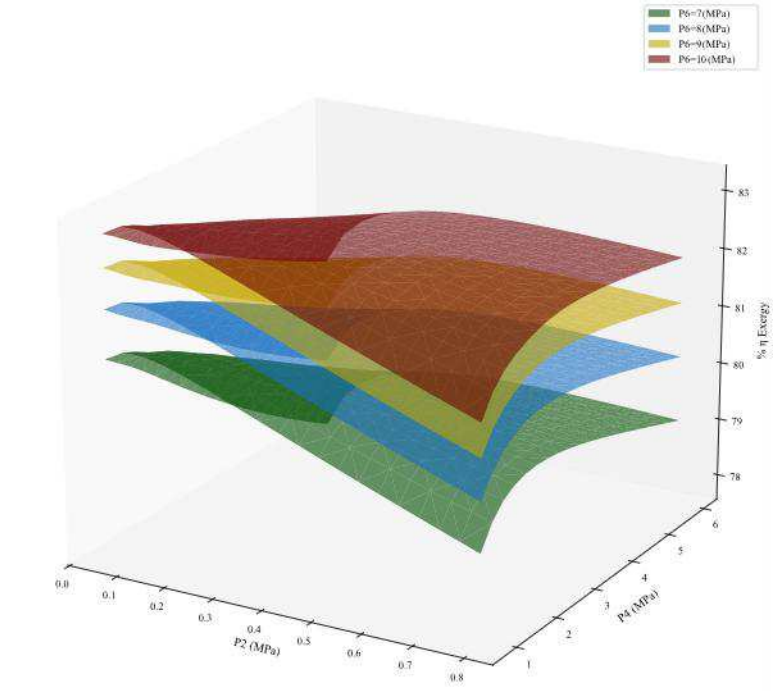

Fig. 13: The continuous 3D layer of η_{Exergy} for changes in P_2, P_4 and P_6.

DRL uses the neural network model to find the values of P_2, P_4 and P_6 that lead to maximum exergy efficiency. It is clear from this figure that the exergy efficiency, like the thermal efficiency, has a global maximum value between the upper and lower boundaries of P_2 and P_4. It is also clear that the exergy efficiency increases monotonically with increasing, although this increase must P_6 be accompanied by technical considerations.

Tables 4-9 are based on the EES output data and the optimal states obtained from DRL. Table 4 shows the optimal values of P_2, P_4 and P_6 that maximize the specific network (w_{net}), the thermal efficiency $(\eta_{Thermal})$ and the exergy efficiency (η_{Exergy}) based on the weight functions introduced in Table 3 against different values of the environment temperature. It is clear that with increasing environment temperature, the optimal values of P_2, P_4 and P_6 decrease, increase and decrease, respectively. Also, environment temperature has an decreasing effect on the maximum values of specific network and thermal efficiency and an increasing effect on the exergy efficiency.

Table 4: Optimization results for different values of T_0.

Environment Temperature (°C)	$\eta_{Thermal}$ (%)	η_{Exergy} (%)	w_{net} (kJ/kg)	P_2 (Mpa)	P_4 (Mpa)	P_6 (Mpa)
15	38.62	80.57	827.38	0.660	3.294	8.567
20	38.52	81.30	823.81	0.635	3.378	8.400
25	38.39	82.06	819.65	0.609	3.475	8.221
30	38.28	82.87	815.49	0.579	3.574	8.057
35	38.20	83.78	811.45	0.532	3.686	7.945

The optimal values of P_2, P_4 and P_6 that maximize the specific network, thermal efficiency, and exergy efficiency based on the weight functions introduced in Table 3 against different values

of the inlet temperature of air heat source are presented in Table 5. According to this table, with increasing inlet temperature, the optimal values of P_2, P_4 and P_6 increase, decrease, and increase, respectively. Also, inlet temperature has an increasing effect on the maximum values of specific network and thermal efficiency and a decreasing effect on exergy efficiency.

Table 5: Optimization results for different values of T_{g-in}.

Inlet Temperature (°C)	$\eta_{Thermal}$ (%)	η_{Exergy} (%)	w_{net} (kJ/kg)	P_2 (Mpa)	P_4 (Mpa)	P_6 (Mpa)
400	38.32	83.86	807.23	0.536	3.874	8.138
412	38.49	82.89	822.59	0.604	3.416	8.346
424	38.51	81.78	823.38	0.625	3.391	8.380
436	38.52	80.87	824.27	0.645	3.364	8.420
448	38.54	80.12	825.30	0.666	3.333	8.463

Similarly, the optimal values of values of P_2, P_4 and P_6 that maximize the specific network, thermal efficiency, and exergy efficiency based on the weighting functions introduced in Table 3 against different values of f_g, η_{Pumps} and $\eta_{Turbines}$ are presented in Tables 6-8. The presented information in these tables contains interesting results. As shown in Table 6, increasing the flow rate ratio from 6.0 to 10.0 leads to a monotonic increase in thermal efficiency (from 37.60% to 38.62%) and specific work (from 759.31 kJ/kg to 831.66 kJ/kg). In contrast, exergy efficiency first decreases from 83.40% to 81.36% (at ratio 9.0) and then slightly rises to 81.79% at ratio 10.0, indicating the presence of a local minimum. Regarding the optimal operating pressures, P_2 and P_6 increase monotonically, while P_4 decreases monotonically as the flow rate ratio increases. The

Table 7 reveals that increasing pump isentropic efficiency from 60% to 90% causes a gradual decrease in thermal efficiency (from 38.60% to 38.26%) and specific work (from 825.18 kJ/kg to 819.22 kJ/kg), while exergy efficiency increases slightly (from 81.25% to 81.49%). The optimal values of P_2, P_4, and P_6 remain nearly constant, indicating that pump efficiency has a weak effect on the optimum pressure settings. This suggests that once the pump efficiency exceeds 60%, further improvements do not significantly alter the optimal operating points. Also, Table 8 demonstrates a strong positive impact of turbine efficiency on all three objective functions. As turbine efficiency rises from 60% to 90%, thermal efficiency jumps from 29.41% to 40.54% (an increase of nearly 38%), exergy efficiency from 66.07% to 84.60% (a 28% increase), and specific work from 646.22 kJ/kg to 857.27 kJ/kg (a 33% increase). Notably, the optimal outlet pressure of the first pump (P_6) decreases with higher turbine efficiency, while P_4 increases. This indicates that more efficient turbines allow lower pump outlet pressures to achieve better overall cycle performance.

Table 6: Optimization results for different values of f_g.

The ratio of heat source mass flow rate to basic mass flow rate	$\eta_{Thermal}$ (%)	η_{Exergy} (%)	w_{net} (kJ/kg)	P_2 (Mpa)	P_4 (Mpa)	P_6 (Mpa)
6.0	37.60	83.40	759.31	0.161	5.280	7.939
7.0	38.31	82.21	806.36	0.559	3.900	8.154
8.0	38.50	81.39	822.01	0.630	3.429	8.368
9.0	38.56	81.36	827.54	0.645	3.273	8.469
10.0	38.62	81.79	831.66	0.654	3.160	8.554

Table 7: Optimization results for different values of η_{Pumps}.

Pump Efficiency (%)	$\eta_{Thermal}$ (%)	η_{Exergy} (%)	w_{net} (kJ/kg)	P_2 (Mpa)	P_4 (Mpa)	P_6 (Mpa)
60	38.60	81.25	825.18	0.635	3.376	8.402
68	38.47	81.33	823.09	0.635	3.379	8.399
74	38.40	81.40	821.81	0.635	3.381	8.397
82	38.33	81.44	820.39	0.634	3.384	8.395
90	38.26	81.49	819.22	0.634	3.386	8.393

Table 8: Optimization results for different values of $\eta_{Turbines}$.

Turbine Efficiency (%)	$\eta_{Thermal}$ (%)	η_{Exergy} (%)	w_{net} (kJ/kg)	P_2 (Mpa)	P_4 (Mpa)	P_6 (Mpa)
60	29.41	66.07	646.22	0.767	2.482	9.422
68	32.65	71.52	712.41	0.731	2.690	9.191
74	34.96	75.40	757.91	0.705	2.899	8.952
82	37.83	80.18	811.77	0.655	3.265	8.523
90	40.54	84.60	857.27	0.511	3.776	8.097

Finally, Table 9 summarizes the important optimization results. In this table, the specific values of P_2, P_4 and P_6 for maximizing the specific network, thermal efficiency and exergy efficiency can be seen. According to these results, the maximum amount of specific network is $w_{net} = 944.1 kJ/kg$, which occurred at pressures values of $P_2 = 0.0733\,Mpa$, $P_4 = 0.5136\,Mpa$ and

$P_6 = 6200\,Mpa$. Also, the maximum amount of thermal efficiency is $\eta_{Thermal} = 37.13\%$, which occurred at pressure values of $P_2 = 0.1513\,Mpa$, $P_4 = 1.5852\,Mpa$ and $P_6 = 6200\,Mpa$.

Table 9: Summary of optimization results.

Method	$\eta_{Thermal}$ (%)	η_{Exergy} (%)	w_{net} (kJ/kg)	P_2 (Mpa)	P_4 (Mpa)	P_6 (Mpa)
Maximum of w_{net}	36.70	80.48	944.1	0.0733	0.5136	6.200
Maximum of $\eta_{Thermal}$	37.13	80.16	876.9	0.1513	1.5852	6.200
Maximum of of η_{Exergy}	36.75	80.42	939.8	0.0774	0.5624	6.200
SAC Deterministic	38.52	81.30	823.8	0.635	3.378	8.400
SAC Stochastic	38.99	82.35	888.2	0.103	1.822	8.971

The maximum amount of exergy efficiency is also $\eta_{Exergy} = 36.75\%$, which occurred at pressure values of $P_2 = 0.0774\,Mpa$, $P_4 = 0.5624\,Mpa$ and $P_6 = 6200\,Mpa$.

6. Conclusion

Parametric analysis of regenerative Rankine cycle with two feedwater heaters, as well as their multi objective optimization with a novel procedure based on a DRL algorithm with the soft actor-critic architecture, has been presented. The maximum variation in the target parameters is obtained to be less than 0.58% through the optimization procedure. Convergence to an almost constant value of thermal efficiency for a wide variation of the heat source mass flow rate to basic mass flow rate and the inlet temperature of air heat source is also achieved. In addition, an ascending trend for the thermal and exergy efficiencies at a maximum point as a function of the outlet pressure of third and second pumps is obtained. In contrast, the variation of these efficiencies with respect to the

first pressure has a completely descending pattern. Here, the simultaneous optimized values of the thermal efficiency, exergy efficiency and specific work are obtained to be 38.99%, 82.35% and 888.2 kJ/kg, respectively.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request. The EES code and the Python implementation of the SAC algorithm are not publicly available due to ongoing development but can be shared with qualified researchers upon request. All data used to generate the figures and tables in this paper are included in the manuscript.

Author Contributions Statement

Amir Basiriparsa: Conceptualization, Methodology, Software (EES), Formal analysis, Writing–original draft, Supervision. **MohammadJavad Faraji:** Methodology (DRL algorithm), Software (Python), Validation, Writing-review & editing. **Alireza Kokabi:** Data curation, Investigation, Visualization, Writing–review & editing. All authors have read and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Funding Declaration

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

References

1. Ding T, Liu S, Yuan W, et al (2015) A two-stage robust reactive power optimization considering uncertain wind power integration in active distribution networks. IEEE Transactions on Sustainable Energy 7:301–311

2. Tian H, Liu P, Shu G (2021) Challenges and opportunities of Rankine cycle for waste heat recovery from internal combustion engine. Progress in Energy and Combustion Science 84:100906

3. Bejan A, Tsatsaronis G, Moran MJ (1995) Thermal design and optimization. John Wiley & Sons

4. Moran MJ, Shapiro HN, Boettner DD, Bailey MB (2010) Fundamentals of engineering thermodynamics. John Wiley & Sons

5. Nowak W, Borsukiewicz-Gozdur A, Stachel A (2008) Using the low-temperature Clausius–Rankine cycle to cool technical equipment. Applied Energy 85:582–588

6. Fischer J (2011) Comparison of trilateral cycles and organic Rankine cycles. Energy 36:6208–6219

7. Yamamoto T, Furuhata T, Arai N, Mori K (2001) Design and testing of the organic Rankine cycle. Energy 26:239–251

8. Rashidi M, Galanis N, Nazari F, et al (2011) Parametric analysis and optimization of regenerative Clausius and organic Rankine cycles with two feedwater heaters using artificial bees colony and artificial neural network. Energy 36:5728–5740

9. Shakya AK, Pillai G, Chakrabarty S (2023) Reinforcement learning algorithms: A brief survey. Expert Systems with Applications 231:120495

10. Chen X, Yao L, McAuley J, et al (2021) A survey of deep reinforcement learning in recommender systems: A systematic review and future directions. arXiv preprint arXiv:210903540

11. Ruan Y, Liang Z, Qian F, et al (2023) Operation strategy optimization of combined cooling, heating, and power systems with energy storage and renewable energy based on deep reinforcement learning. Journal of Building Engineering 65:105682

12. Zhang B, Hu W, Cao D, et al (2020) Economical operation strategy of an integrated energy system with wind power and power to gas technology–a DRL-based approach. IET Renewable Power Generation 14:3292–3299

13. Wang X, Yao H, Mai T, et al (2023) Reinforcement learning-based particle swarm optimization for end-to-end traffic scheduling in TSN-5G networks. IEEE/ACM Transactions on Networking 31:3254–3268

14. Yi Z, Luo Y, Westover T, et al (2022) Deep reinforcement learning based optimization for a tightly coupled nuclear renewable integrated energy system. Applied Energy 328:120113

15. Adams D, Oh D-H, Kim D-W, et al (2021) Deep reinforcement learning optimization framework for a power generation plant considering performance and environmental issues. Journal of Cleaner Production 291:125915

16. Lin R, Luo Y, Wu X, et al (2024) Surrogate empowered Sim2Real transfer of deep reinforcement learning for ORC superheat control. Applied Energy 356:122310

17. Beeler C, Yahorau U, Coles R, et al (2021) Optimizing thermodynamic trajectories using evolutionary and gradient-based reinforcement learning. Phys Rev E 104:064128. https://doi.org/10.1103/PhysRevE.104.064128

18. Kim S, Jang M-G, Kim J-K (2023) Process design and optimization of single mixed-refrigerant processes with the application of deep reinforcement learning. Applied Thermal Engineering 223:120038. https://doi.org/10.1016/j.applthermaleng.2023.120038

19. Zhou S, Hu Z, Gu W, et al (2020) Combined heat and power system intelligent economic dispatch: A deep reinforcement learning approach. International Journal of Electrical Power & Energy Systems 120:106016. https://doi.org/10.1016/j.ijepes.2020.106016

20. Egan D, Xu B, Zhu Q, Prucka R (2022) Reinforcement Learning Based Control of an Organic Rankine Cycle Waste Heat Recovery System Over a Drive Cycle for Heavy-Duty Diesel Engines. American Society of Mechanical Engineers Digital Collection

21. Zhang L, Lin R, Xie L, et al (2024) Event-Triggered Constrained Optimal Control for Organic Rankine Cycle Systems via Safe Reinforcement Learning. IEEE Transactions on Neural Networks and Learning Systems 35:7126–7137. https://doi.org/10.1109/TNNLS.2022.3213825

22. Wang X, Wang R, Jin M, et al (2020) Control of superheat of organic Rankine cycle under transient heat source based on deep reinforcement learning. Applied Energy 278:115637. https://doi.org/10.1016/j.apenergy.2020.115637

23. Engineering Equation Solver Ees Software. https://reviewbooku.com/review/engineering-equation-solver-ees-software-4974125. Accessed 28 May 2026

24. Sepehrzad R, Faraji MJ, Al-Durra A, Sadabadi MS (2024) Enhancing Cyber-Resilience in Electric Vehicle Charging Stations: A Multi-Agent Deep Reinforcement Learning Approach. IEEE Transactions on Intelligent Transportation Systems 25:18049–18062. https://doi.org/10.1109/TITS.2024.3408238

25. Haarnoja T, Zhou A, Abbeel P, Levine S (2018) Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. In: Proceedings of the 35th International Conference on Machine Learning. PMLR, pp 1861–1870