Towards Human-Centric Manufacturing: Task Planning Under Uncertainties in Human-Robot Collaborative Assembly⋆

Yingchao You[a,*], Ze Ji[b] and Changyun Wei[a]

[a]College of Mechanical and Electrical Engineering, Hohai university, Changzhou, 213200, China
[b]School of Engineering, Cardiff University, Queen's Buildings, The Parade, Cardiff, CF24 3AA, Wales, UK

ARTICLE INFO

Keywords:
human-robot collaboration
task planning
physical exertion
human-centric manufacturing

ABSTRACT

Task planning plays a pivotal role in ensuring the smooth collaboration between humans and robots by efficiently allocating tasks among agents and scheduling available resources. Although some recently proposed task planners incorporate human factors into their frameworks, few explicitly account for human-related uncertainties, which can potentially lead to task failures. To address this gap, this study introduces a physical exertion–aware task planner that explicitly considers uncertainties in both human factors and task execution time. The uncertainties associated with physical exertion and execution time are modelled using the Single-Valued Triangular Neutrosophic (SVTN) Number method. Furthermore, a reinforcement learning-based approach is developed to learn adaptive task allocation policies and scheduling under these uncertainties. The experimental results indicate that the reinforcement learning-based approach effectively reduces performance variability compared with the benchmark methods.

1. Introduction

Human–robot collaboration (HRC) represents a critical manufacturing paradigm that integrates the flexibility and decision-making capabilities of humans with the precision and efficiency of robotic systems. The overarching objective is to enable efficient collaboration while ensuring both human well-being and system stability Noormohammadi-Asl, Smith and Dautenhahn (2024). Within this context, adaptive task allocation and scheduling play a pivotal role in meeting the requirements of flexible production and stringent quality standards Maderna, Pozzi, Zanchettin, Rocco and Prattichizzo (2022). Unlike fully automated production lines, however, HRC systems comprise heterogeneous agents, humans and robots, with inherently distinct capabilities, work rhythms, and operational constraints. This heterogeneity introduces significant challenges for task planning and necessitates the development of advanced, adaptive scheduling methods Alirezazadeh and Alexandre (2022).

Despite the substantial research efforts devoted to HRC task planners, several critical challenges remain unresolved. First, existing studies predominantly emphasise traditional performance metrics such as makespan, often neglecting human factors such as physical exertion. In practice, excessive physical workload not only degrades human performance but also elevates health risks Pupa and Secchi (2021). Second, assembly tasks inherently involve multiple sources of uncertainty, including variable task durations and fluctuations in human physical states. If left unaddressed, these uncertainties may propagate throughout the system, resulting in suboptimal schedules or even production disruptions Djogdom, Meziane and Otis (2024). Therefore, there is a pressing need for uncertainty-aware and human-centred task planning methods tailored to HRC assembly environments.

To address these research gaps, we propose a physical exertion-aware task planning framework for HRC assembly under uncertainty. The proposed approach integrates: (i) Single-Valued Triangular Neutrosophic Numbers to model uncertain task durations and the accumulation–recovery dynamics of human physical exertion, explicitly capturing the truth, indeterminacy, and falsity aspects of human physical states; and (ii) a multi-objective Double Deep Q-Network with Single-Valued Triangular Neutrosophic Numbers (DDQN-SVTN) scheduling algorithm designed to minimise both the overall makespan and the risk of excessive human physical exertion. Specifically, the HRC environment incorporates a multi-objective reward function that penalises not only prolonged completion times but also unsustainable physical exertion levels. This formulation encourages balanced scheduling solutions that jointly improve production efficiency and worker well-being. Through iterative reinforcement learning, the proposed method

⋆

Youy4@cardiff.ac.uk (Y. You); JiZ1@cardiff.ac.uk (Z. Ji); c.wei@hhu.edu.cn (C. Wei)
ORCID(s):

adaptively learns task allocation strategies that respond to stochastic task durations, system dynamics, and evolving physical exertion states.

The key contributions of this study are summarised as follows:

1. Human-centric uncertainty modelling: We develop an SVTN-based framework to represent uncertainty in both production processes and human physical states, enabling more accurate characterisation of physical exertion and execution-time variability.
2. Multi-objective HRC scheduling: We design a reinforcement learning-based scheduling approach that jointly improves production efficiency and physical exertion, ensuring sustainable human–robot collaboration.
3. Scalable learning-based planning: By integrating uncertainty modelling with deep reinforcement learning, the proposed method provides adaptive, scalable task planning policies for complex, large-scale HRC assembly systems.

2. Related work

2.1. Human-robot collaborative task planning and scheduling

Human–robot collaboration has attracted growing attention due to its inherent flexibility advantages. Early research primarily focused on task allocation and sequencing under static conditions, where tasks were assigned to either humans or robots to minimise makespan Pupa, Landi, Bertolani and Secchi (2020); Wilcox (2013). To this end, mathematical programming and heuristic algorithms were widely employed to enhance scheduling efficiency Zhang, Chen, Xu, Gong and Meng (2020); Lamon, Fusaro, De Momi and Ajoudani (2023). However, these traditional optimisation approaches typically rely on precisely defined process parameters and exhibit limited adaptability to the dynamic and stochastic environments characteristic of modern HRC systems.

In recent years, research efforts have shifted toward dynamic and adaptive HRC scheduling, enabling the scheduler to respond to real-time variations in system state Alirezazadeh and Alexandre (2022); Wang, Yan, Hu, Yang and Zhang (2025b); Liu, Liu, Wang, Xu and Zhou (2021). Reinforcement learning-based methods have demonstrated promising performance in complex manufacturing scenarios by learning task allocation policies directly from interaction data, thus eliminating the need for predefined rules.

Although earlier HRC scheduling studies often treated human operators as deterministic resources and focused on productivity-oriented objectives, recent research has increasingly begun to incorporate human physical exertion and physical states into the task planning process. For example, Peternel, Tsagarakis, Caldwell and Ajoudani (2017) proposed adaptive robot control to account for human physical exertion during co-manipulation, while Messeri, Bicchi, Zanchettin and Rocco (2022) developed a dynamic allocation strategy to mitigate physical exertion. Moreover, Ghorbani, Keivanpour, Sekkay and Imbeau (2024) introduced fuzzy inference for adaptive workload management in human-centric robotic assembly. These advances highlight growing attention to operator well-being, though comprehensive uncertainty modelling of human states remains limited.

2.2. Human factors and physical exertion modelling in manufacturing

The human factor plays a pivotal role in HRC systems, as humans are constrained by physical exertion and recovery processes that directly influence both task performance and operational health Zhang and Li (2014); Peternel et al. (2017). Over the years, various modelling approaches have been proposed to quantify physical workload and physical exertion. For instance, work–rest scheduling models Messeri et al. (2022) and ergonomic risk assessment methods such as RULA Keshvarparast, Berti, Chand, Guidolin, Lu, Battaïa, Xu and Battini (2024) provide rule-based evaluations of physical workload. Other studies have incorporated physiological measurements (e.g., heart rate, muscle activity) into physical exertion estimation frameworks Kumar, Jujjavarapu and Esfahani (2021), while stochastic and fuzzy modelling techniques have been employed to capture variability and uncertainty in human states Ghorbani et al. (2024). An online physical exertion estimation method based on an upper-limb musculoskeletal model integrated with a 3D vision system was proposed to compute real-time muscle loading through motion capture data. The estimated muscle forces were further utilised to dynamically adjust human–robot task allocation, effectively reducing muscular physical exertion in industrial HRC scenarios Messeri et al. (2022). In addition, a Transformer-based multimodal fusion framework integrating visual, audio, and wearable sensor data was developed to assess cognitive workload and muscle physical exertion during collaborative robot interaction, thereby enhancing safety and reliability in physical human–robot collaboration Guo, Liu and Niu (2023).

Several physical exertion-aware scheduling models have been proposed. Certain studies have incorporated recovery dynamics and temporal physical exertion evolution. For example, Peternel et al. (2017) formulated a differential physical exertion model reflecting muscle recovery, and You, Cai, Pham, Liu and Ji (2025) introduced a human digital twin model to estimate physical exertion accumulation and recovery. Zhang, Li, Shang and Liu (2022a) proposed a physical exertion-aware human–robot collaborative scheduling model that integrates micro-breaks within job cycles and solves the cycle time–physical exertion trade-off using an improved chemical reaction optimisation algorithm. Chand and Lu (2023) developed a personalised physical exertion-aware dual scheduling strategy that considers individual worker differences and applies a customized NSGA-III algorithm to balance cycle time and worker physical exertion. Yao, Li, Ji, Xiao and Xu (2024) present a multimodal physical exertion assessment framework combined with dynamic task reallocation, where real-time physical exertion estimation triggers reinforcement learning-based optimisation for adaptive HRC scheduling. Baratta, Cimino, Longo, Mirabelli and Nicoletti (2024) introduced a simulation-based task allocation approach that evaluates different collaboration modes to jointly improve productivity and reduce operator physical exertion in assembly stations. Liau and Ryu (2020) designed a capability-driven task allocation method using analytic network process and genetic algorithms to optimise action-level assignment among human and robotic agents. Zheng, Chand, Keshvarparast, Battini and Lu (2023) proposed a non-contact physical exertion-aware task allocation framework that estimates operator physical exertion from video-based action recognition and integrates it into a cycle-time-constrained optimisation model.

Despite these developments, the integration of such stochastic physical exertion dynamics into large-scale, uncertainty-aware HRC scheduling frameworks remains scarce. In this work, a physical exertion estimation model from our previous work You et al. (2025) is integrated, which is based on a human digital twin model. Moreover, few studies explicitly incorporate physical exertion considerations into multi-objective optimisation frameworks, leaving a gap in balancing production efficiency and human well-being in HRC assembly systems.

2.3. Uncertainty modelling-based scheduling

Manufacturing processes inevitably involve multiple sources of uncertainty, including stochastic task durations and human performance fluctuations Casalino, Mazzocca, Giorgio, Zanchettin and Rocco (2019). Fuzzy set theory and its variants, such as triangular fuzzy numbers and intuitionistic fuzzy sets, have been widely used to represent imprecise information in scheduling problems Chen, Zou and Wang (2023). Building on this, Single-Valued Triangular Neutrosophic Numbers were recently introduced to capture truth, indeterminacy, and falsity degrees simultaneously, providing a more flexible representation of uncertainty Pregina, Vasugi and Kannan (2025). SVTN-based methods have shown advantages in decision-making problems under incomplete and inconsistent information Fujita (2024).

Parallel to advances in uncertainty modelling, reinforcement learning, especially Deep Q-Networks and policy gradient methods, has been applied to scheduling under uncertainty Amirnia and Keivanpour (2024); Altundas, Wang, Bishop and Gombolay (2022). These approaches enable the system to learn adaptive policies that improve long-term rewards without explicit enumeration of all possible scenarios. Recent reinforcement learning-based scheduling studies have started to integrate multi-objective optimisation into HRC systems, addressing trade-offs between efficiency, ergonomics, and safety. For example, multi-objective deep reinforcement learning has been used to learn non-dominated scheduling policies that balance operational efficiency and human workload fairness in stochastic collaborative environments Smit, Bukhsh, Pechenizkiy, Alogariastos, Hendriks and Zhang (2024). Similarly, RL-based task planners have been developed to jointly optimise safety and task efficiency in cooperative task-motion planning, enabling adaptive scheduling under uncertain human movements Liu, de Winter, Merckaert, Steckelmacher, Nowe and Vanderborght (2025).

Beyond productivity-oriented objectives, recent works have explicitly integrated ergonomic indicators into the learning loop. Model-free RL approaches have been proposed to dynamically improve workers' posture quality (e.g., RULA) during collaboration by adapting robot configurations in real time Xie, Lu, Wang, Su, Liu and Xu (2022, 2023). In addition, learning-based task allocation frameworks have modelled human cognitive states such as trust and physical exertion within Markov decision processes, enabling joint optimisation of task success probability and long-term operational cost under uncertainty Wu, Hu and Lin (2017).

More broadly, reinforcement learning has been shown to enhance adaptability to stochastic human performance variation in manufacturing settings Oliff, Liu, Kumar, Williams and Ryan (2020), while surveys highlight the growing emphasis on jointly achieving safety, ergonomics, and productivity in collaborative robotics control and scheduling frameworks Proia, Carli, Cavone and Dotoli (2022); Lorenzini, Lagomarsino, Fortini, Gholami and Ajoudani (2023). Nevertheless, although existing studies address efficiency–safety trade-offs, ergonomic optimisation, and stochastic

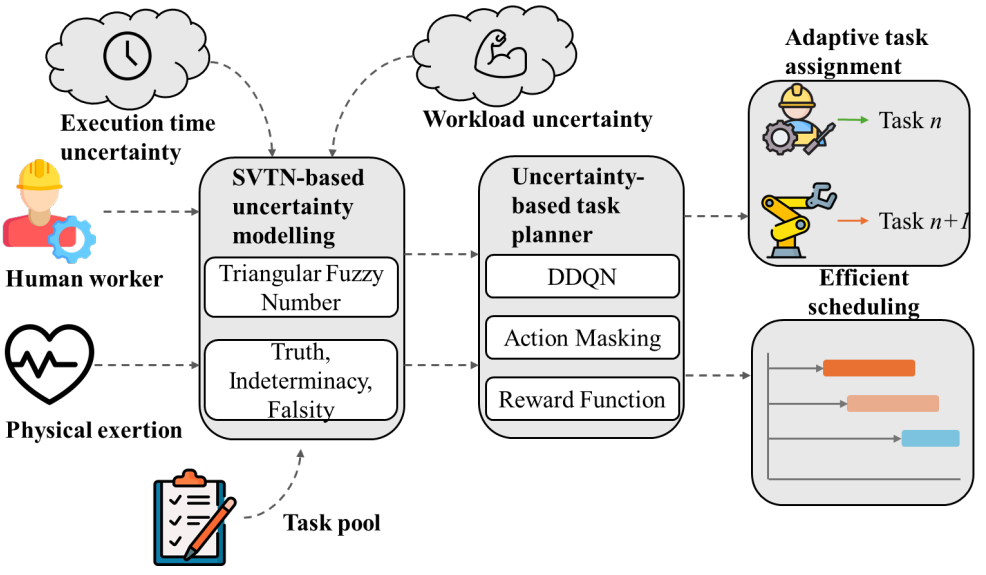


Figure 1: The overall framework of the task planner for human-robot collaborative assembly.

human behaviour separately, the explicit integration of uncertainty-aware physical exertion modelling into a unified multi-objective RL scheduling framework remains limited.

Existing uncertainty-aware HRC scheduling studies mainly focus on improving production efficiency. For example, Jia, Xie and Zhang (2025) proposed a flexible task assignment and scheduling model that minimises fuzzy cycle time using a genetic algorithm, where uncertainty is primarily associated with execution time variability. In contrast, our work adopts a human-centric perspective by explicitly modelling uncertainty in both task duration and human physical exertion. By integrating SVTN-based uncertainty modelling with reinforcement learning, we jointly optimise makespan and human well-being, addressing sustainability concerns that are insufficiently captured by efficiency-oriented scheduling models alone.

3. Methodology

3.1. Framework

Figure 1 illustrates the overall framework of the proposed physical exertion–aware task planning approach for human–robot collaborative assembly under uncertainty. The system is composed of two major modules: (1) SVTN-based uncertainty modelling, and (2) uncertainty-based task planner.

In the first module, the stochastic characteristics of both task execution time and human physical exertion are represented using the SVTN model. This model extends the conventional triangular fuzzy number by introducing three dimensions—truth, indeterminacy, and falsity. Such representation enables more comprehensive quantification of uncertainty arising from human performance variation, tool condition, and environmental factors.

The second module constitutes the decision-making core, where a Double Deep Q-Network agent learns optimal task allocation and sequencing policies. The agent interacts with the SVTN-driven environment and employs action masking to ensure that only feasible actions—consistent with task precedence and resource constraints—are selected during training. The reward function integrates both production efficiency and human well-being, penalising excessive makespan and accumulated physical exertion.

Finally, the learned policy outputs a coordinated schedule that dynamically allocates operations between human workers and robots, adapting to uncertainty in execution time and physical exertion. The resulting plan achieves a balanced trade-off between productivity and operator safety, demonstrating the synergy between uncertainty modelling and reinforcement learning-based task planning.

To clarify the logical flow of Section 3, we briefly outline how the subsequent subsections are connected within the proposed framework. Section 3.2 first identifies task execution time and human physical exertion as the main sources of uncertainty in human–robot collaborative assembly and models them using SVTNs (Section 3.2.1). These SVTNs are

then transformed into value and ambiguity indices, which provide compact representations of expected performance and associated uncertainty (Section 3.2.2). In Section 3.3, the resulting indices are embedded into the reinforcement learning formulation, where they are incorporated into both the state representation and the reward function. Through this pipeline, uncertainty is consistently propagated from modelling assumptions to learning-based task allocation and scheduling decisions, ensuring a human-centric and uncertainty-aware planning process.

3.2. Single-Valued Triangular Neutrosophic Modelling

Uncertainty is an inherent feature of human–robot collaborative assembly due to variability in human performance and stochastic task conditions. Traditional deterministic or probabilistic models are often insufficient to capture the ambiguity, imprecision, and indeterminacy present in real-world assembly environments. To address these issues, this study employs the SVTN framework to model the uncertainties in task execution time and human physical exertion.

3.2.1. Modelling Task Execution Time And Physical Exertion

An SVTN is defined as a triplet $\hat{A} =< T, I, F >$, where T, I and F represent the truth, indeterminacy and falsity, respectively. Each component is expressed as a triangular fuzzy number (a_1, a_2, a_3), capturing the lower, modal, and upper bounds of the corresponding dimension. Mathematically, the SVTN extends classical fuzzy logic by allowing explicit representation of the indeterminate region, thereby enhancing flexibility in describing uncertain phenomena.

In the proposed human–robot collaborative assembly framework, the task execution time (D_i) and human physical exertion (E_i) for each operation i are modeled as SVTNs:

$$D_i = (T_D^i, I_D^i, F_D^i), \quad E_i = (T_E^i, I_E^i, F_E^i) \tag{1}$$

Each component is parameterised using empirical quantiles of repeated observations and represents a different level of uncertainty. Specifically, the truth component T captures the most representative or typical range of task execution outcomes, corresponding to the central portion of the empirical distribution. The indeterminacy component I and falsity component F progressively expand this range to account for increasing levels of variability, forming uncertainty envelopes that reflect unmodelled disturbances and rare or extreme deviations.

It should be noted that the indeterminacy I_D^i and falsity F_D^i components do not correspond to directly measured variables. Instead, they are intentionally introduced to represent the aggregated effect of unmodelled execution-level disturbances that are difficult to observe or quantify explicitly.

Similarly, E_i accounts for human workload uncertainty, where the truth component corresponds to measured exertion, while indeterminacy captures the variability in physiological responses. These SVTNs are constructed from empirical physical exertion parameters using a percentile-based calibration method. The physical exertion is estimated using a muscle force-based method in our previous work You et al. (2025). An IK-BiLSTM-AM-based surrogate model, which integrates inverse kinematics (IK), bidirectional long short-term memory (BiLSTM), and an attention mechanism (AM), was proposed. This model approximates the complex biomechanical simulation for estimating muscle force. A muscle force-based physical exertion estimation model Peternel et al. (2017) is then applied, which estimates physical exertion using historical force data via a first-order differential equation. The mathematical formulation of this model is as follows:

$$\frac{de_m(t)}{dt} = \begin{cases} (1 - e_m(t))\frac{f_m(t)}{c_m} & f_m(t) \geq f_{\text{th}} \\ -e_m(t)\frac{R}{c_m} & f_m(t) < f_{\text{th}} \end{cases} \tag{2}$$

where $e_m(t)$ represents the physical exertion of human muscle m, while $f_m(t)$ denotes the force exerted by muscle m at time t, which is estimated using previous work You et al. (2025). The recovery coefficient R, set to 0.5, indicates the recovery rate from physical exertion. The threshold of muscle force for muscle m is denoted by f_{th}. The capability coefficient of muscle m, denoted by c_m, reflects the muscle's resistance to physical exertion.

By aggregating all operations:

$$D_{\text{total}} = \sum_i D_i, \quad E_{\text{total}} = \sum_i E_i \tag{3}$$

The planner can compute the expected task completion time and overall physical exertion considering uncertainty. During optimisation, the objective function can minimise the combined V_θ while constraining or penalising large A_θ values, thus balancing efficiency and robustness under uncertainty.

3.2.2. Data-driven construction of SVTNs

Let $\mathcal{S}_i = \{x_i^{(k)}\}_{k=1}^{N_i}$ denote a set of N_i repeated observations for operation i, where $x_i^{(k)}$ is either the realised task duration $d_i^{(k)}$ or the realised operation-level physical exertion $e_i^{(k)}$. We construct the three triangular fuzzy numbers in the SVTN $X_i = (T_X^i, I_X^i, F_X^i)$ $(X \in \{D, E\})$ using an empirical quantile calibration.

Denote by $Q_i(p)$ the empirical p-quantile of $\mathcal{S}_i$. Given three tail probabilities $0 < \gamma < \beta < \alpha < 1$, the truth, indeterminacy, and falsity components are parameterised as:

$$T_X^i = \big(Q_i(\alpha),\ Q_i(0.5),\ Q_i(1-\alpha)\big), \tag{4}$$

$$I_X^i = \big(Q_i(\beta),\ Q_i(0.5),\ Q_i(1-\beta)\big), \tag{5}$$

$$F_X^i = \big(Q_i(\gamma),\ Q_i(0.5),\ Q_i(1-\gamma)\big). \tag{6}$$

By construction, the three components satisfy a nested uncertainty hierarchy $[Q_i(\alpha), Q_i(1-\alpha)] \subseteq [Q_i(\beta), Q_i(1-\beta)] \subseteq [Q_i(\gamma), Q_i(1-\gamma)]$, i.e., T captures the most typical range, while I and F progressively expand to represent increasingly conservative uncertainty envelopes associated with unmodelled disturbances and rare/extreme outcomes.

3.2.3. Quantitative Measures of SVTN: Value and Ambiguity

To quantify an SVTN for optimisation or comparison, two scalar indices are introduced: the value degree (V_θ) and the ambiguity degree (A_θ) Jia et al. (2025).

(1) Value degree The value degree measures the representative or centroidal value of the SVTN:

$$V_\theta = \theta \cdot c(T) + (1-\theta) \cdot \frac{1}{2}\left[c(I) + c(F)\right] \tag{7}$$

where $c(\cdot)$ denotes the centroid of a triangular fuzzy number:

$$c(T) = \frac{T_L + T_M + T_U}{3} \tag{8}$$

where $T = (T_L, T_M, T_U)$ denotes a triangular fuzzy number. T_L, T_M and T_U represent the lower, modal, and upper values, respectively, and $c(T)$ is its centroid. $\theta \in [0, 1]$ controls the relative importance between determinacy (T) and indeterminacy (I, F).

(2) Ambiguity degree The ambiguity degree reflects the spread or uncertainty width of the SVTN:

$$A_\theta = \theta \cdot s(T) + (1-\theta) \cdot \frac{1}{2}\left[s(I) + s(F)\right] \tag{9}$$

where $s(\cdot)$ is the spread of a triangular fuzzy number:

$$s(T) = T_U - T_L \tag{10}$$

A smaller A_θ implies higher confidence in the value, while a larger one indicates greater uncertainty.

3.2.4. Aggregation of SVTNs

In practical modelling, each task or operator state may be represented by multiple SVTN quantities. For instance, an assembly process with several operations can be expressed as

$$\tilde{X}_{\text{total}} = \sum_{i=1}^{n} \tilde{X}_i \tag{11}$$

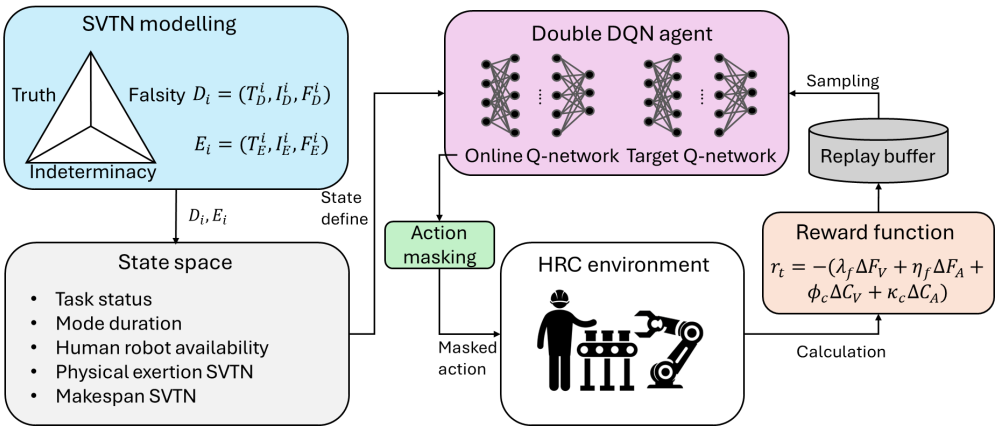


Figure 2: The DDQN-SVTN-based task planner

Given two SVTNs $\tilde{X}_1 = (T_1, I_1, F_1)$ and $\tilde{X}_2 = (T_2, I_2, F_2)$, their summation $\tilde{X}_{\text{total}}$ is defined by the component-wise addition of the corresponding TFNs:

$$\tilde{X}_1 + \tilde{X}_2 = (T_1 \oplus T_2, I_1 \oplus I_2, F_1 \oplus F_2) \tag{12}$$

where

$$(a_1, b_1, c_1) \oplus (a_2, b_2, c_2) = (a_1 + a_2,\ b_1 + b_2,\ c_1 + c_2) \tag{13}$$

where (a_1, b_1, c_1) and (a_2, b_2, c_2) are triangular fuzzy numbers corresponding to two SVTN components (i.e., T, I, or F). For each triangular fuzzy number, a, b, and c denote the lower bound, modal, and upper bound values, respectively. The operator $\oplus$ represents the addition of two triangular fuzzy numbers, which is performed component-wise to preserve the uncertainty structure. This additive property enables the cumulative estimation of total task duration or total physical exertion under uncertainty, preserving the neutrosophic semantics of truth, indeterminacy, and falsity.

3.3. Double DQN-Based Task Planning

3.3.1. DDQN-SVTN

To enable efficient and stable learning in the stochastic assembly environment, this study adopts a Double Deep Q-Network with Single-Valued Triangular Neutrosophic policy to learn task allocation policies and sequencing decisions under uncertainty. The DDQN-SVTN-based task planner is shown in Fig. 2. Unlike the traditional DQN that tends to overestimate state–action values, the DDQN-SVTN decouples the action selection and evaluation processes to reduce value bias. Specifically, the next action a^* is chosen using the online network, while the corresponding target value is evaluated by the target network:

$$a^* = \arg\max_a Q_{\text{online}}(s', a), \quad y = r + (1 - d)\gamma Q_{\text{target}}(s', a^*) \tag{14}$$

where s' is the next state, r is the immediate reward, d is the termination flag, and γ is the discount factor. The online network Q_{online} and target network Q_{target} share identical architectures. The target network parameters are softly updated at each step using a factor τ:

$$\theta_{\text{target}} \leftarrow (1 - \tau)\theta_{\text{target}} + \tau\theta_{\text{online}} \tag{15}$$

This update strategy maintains training stability by preventing oscillations in the value estimates. The DDQN-SVTN agent interacts with the human–robot collaborative assembly environment to learn optimal decisions that minimise total completion time and human physical exertion simultaneously.

3.3.2. Action Masking Technique

To handle dynamically changing task precedence and feasibility constraints, an action masking mechanism is integrated into the DDQN-SVTN policy. At each decision step, the environment computes a binary mask $M \in \{0,1\}^{|\mathcal{A}|}$, where each element indicates whether a specific task–mode pair is currently valid:

$$M_i = \begin{cases} 1, & \text{if task } i \text{ is ready and all predecessors are completed;} \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

During action selection, the DDQN-SVTN agent applies the mask to the Q-values to suppress illegal actions:

$$Q^{\text{masked}} = \begin{cases} Q(s,a), & M_a = 1, \\ -\infty, & M_a = 0. \end{cases} \tag{17}$$

This ensures that the agent only explores feasible actions, thereby accelerating convergence and preventing invalid scheduling moves that violate precedence or resource constraints. Such masking is critical in human–robot collaborative assembly, where each operation may be executed in multiple modes (human, robot, or hybrid), and infeasible choices can lead to deadlocks or unsafe configurations.

3.3.3. State, Action, and Reward Function

The state space characterises the system configuration of the human–robot collaborative assembly at each decision step, including task progress, resource availability, execution uncertainty, and human physical condition.

Formally, the state at time step t is defined as

$$s_t = \left[\mathbf{x}_t,\ \mathbf{d}_t,\ \mathbf{r}_t,\ \mathbf{SVTN}(C_{\max,t}), \mathbf{f}_t \right], \tag{18}$$

where each component is defined as follows.

- $\mathbf{x}_t = \{(x_i^t, \rho_i^t, p_i^t) \mid t_i \in \text{Tasks}\}$ denotes the task status, where $x_i^t \in \{0,1\}$ indicates whether task t_i is completed, $\rho_i^t \in \{0,1\}$ indicates whether task t_i is ready for execution, and $p_i^t \in \mathbb{N}$ represents the number of unfinished predecessors of task t_i.
- $\mathbf{d}_t = \{\hat{D}_{i,m}^t \mid t_i \in \text{Tasks},\ m \in \text{Modes}(t_i)\}$ represents the expected execution duration of task t_i under feasible execution mode $m \in \{\text{H}, \text{R}, \text{HR}\}$.
- $\mathbf{r}_t = \{r_k^t \mid k \in \mathcal{H} \cup \mathcal{R}\}$ denotes the availability times of all human operators and robots.
- $\mathbf{SVTN}(C_{\max,t}) = \big(V_\theta(C_{\max,t}),\ A_\theta(C_{\max,t})\big)$ denotes the neutrosophic value and ambiguity of the current makespan using Eq. 7 and Eq. 9.
- $\mathbf{f}_t = \big(V_\theta(F_t),\ A_\theta(F_t)\big)$ denotes the human physical exertion state, where $F_t \in [0,1]$ is the normalised cumulative physical exertion derived from the muscle force–based physical exertion model, and (V_θ, A_θ) represent its SVTN-based expectation and uncertainty.

The **action space** comprises all possible combinations of task identifiers and operation modes:

$$\mathcal{A} = \{(t_i, m_j) \mid t_i \in \text{Tasks},\ m_j \in \text{Modes}(t_i)\}, \tag{19}$$

where $m_j \in \{\text{H}, \text{R}, \text{HR}\}$ corresponds to human, robot, and collaborative execution modes, respectively.

The **reward function** is designed to balance production efficiency and human well-being by penalising increases in makespan and physical exertion:

$$r_t = -\big(\lambda_f \Delta F_V + \eta_f \Delta F_A + \phi_{cV} \Delta C_V + \kappa_{cA} \Delta C_A\big), \tag{20}$$

where F_V and F_A are the neutrosophic value and ambiguity of cumulative physical exertion, and C_V and C_A denote the corresponding terms for makespan ($C_{\max}$).

Through continuous interaction and experience replay, the DDQN-SVTN agent learns to schedule tasks adaptively, achieving a robust trade-off between productivity and operator physical exertion under multi-source uncertainty.

3.3.4. Model Training

The proposed DDQN-SVTN is trained through iterative interaction with the SVTN-based human–robot collaborative assembly environment. During each episode, the agent observes the current state, selects a feasible action according to the ϵ-greedy policy with action masking, and receives feedback in terms of the immediate reward and the next state. The replay buffer stores past experiences, allowing the DDQN-SVTN to break temporal correlations and stabilise training. The learning process continues until convergence of the cumulative reward or until the completion time and physical exertion indicators reach steady performance. The training process is shown in algorithm 1.

Algorithm 1 Training Procedure of DDQN-SVTN-Based Planner

Require: Environment $\mathcal{E}$, replay buffer $\mathcal{B}$, learning rate α, discount factor γ, update rate τ

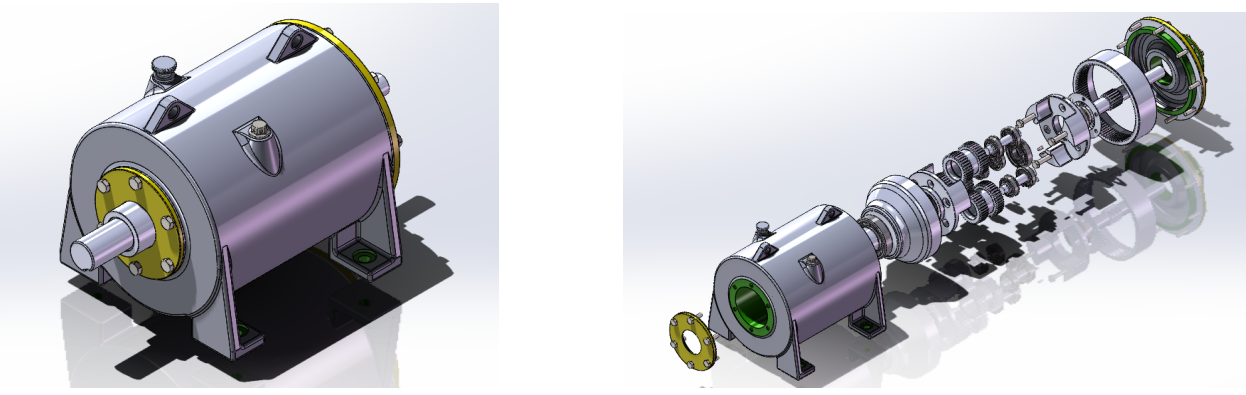
Initialize online network Q_{online} and target network Q_{target}
Initialize $\mathcal{B} \leftarrow \emptyset$, exploration rate $\epsilon \leftarrow 1.0$
for each episode $e = 1 \dots N_{\text{episodes}}$ **do**
    Reset environment: $s_0 \leftarrow \mathcal{E}$.reset()
    while not terminal **do**
        Compute valid mask $M_t \leftarrow \mathcal{E}$.valid_mask()
        Select action a_t using ϵ-greedy policy on masked Q-values
        Execute a_t in $\mathcal{E}$, observe (r_t, s_{t+1}, M_{t+1})
        Store transition $(s_t, a_t, r_t, s_{t+1}, M_{t+1})$ in $\mathcal{B}$
        if size($\mathcal{B}$) > warm-up threshold **then**
            Sample minibatch $(s_i, a_i, r_i, s'_i, M'_i) \sim \mathcal{B}$
            Compute target: $y_i = r_i + (1 - d_i)\gamma Q_{\text{target}}(s'_i, \arg\max_a Q_{\text{online}}(s'_i, a|M'_i))$
            Update Q_{online} by minimizing $(y_i - Q_{\text{online}}(s_i, a_i))^2$
            Soft update target network: $\theta_{\text{target}} \leftarrow (1 - \tau)\theta_{\text{target}} + \tau\theta_{\text{online}}$
        end if
    end while
    Decay exploration rate: $\epsilon \leftarrow \max(\epsilon_{\min}, \epsilon \cdot \epsilon_{\text{decay}})$
end for

During training, the agent progressively learns to balance task efficiency and operator well-being through reward feedback. The SVTN-based environment provides stochastic task durations and physical exertion dynamics, enabling the DDQN-SVTN to generalise across uncertain assembly conditions. The use of experience replay and target network decoupling mitigates overestimation bias and improves convergence stability. The final policy yields adaptive task scheduling that minimises makespan while maintaining acceptable physical exertion levels for human workers.

In all experiments, both the online and target Q-networks adopt a fully connected architecture with two hidden layers of 128 neurons, using ReLU activation functions. The networks are trained with the Adam optimiser and a learning rate of 1×10^{-3}. An ϵ-greedy exploration strategy with action masking is employed, where ϵ is initialised to 1.0 and exponentially decayed to a minimum value of 0.05 during training. The replay buffer has a fixed capacity of 1×10^5 transitions, and minibatches of 64 samples are randomly drawn for network updates after a warm-up phase. SVTNs are transformed into scalar values and ambiguity indices, which are embedded into the state representation and reward function, while stochastic realisations sampled from the SVTNs are used to simulate actual task execution and physical exertion accumulation. Each model is trained for 2,000 episodes, after which the learned policy is evaluated.

4. Experiment

To validate the proposed task planner under uncertainty, two experimental studies were carried out. First, a numerical study based on a two-stage planetary gearbox assembly was performed, evaluating the proposed method from three aspects: (1) ablation analysis, (2) uncertainty assessment, and (3) comparative analysis with multiple humans and multiple robots. Second, a real-world proof-of-concept experiment was conducted in a personal computer (PC) tower human–robot collaborative assembly cell, involving multiple participants to further demonstrate the effectiveness of the proposed method.



(a) (b)

Figure 3: The isometric view and exploded view of the gearbox used in the experiments.

4.1. Numerical study

4.1.1. Task Description

To evaluate the proposed physical exertion–aware task planning framework, a realistic two-stage planetary gearbox assembly process is constructed as the benchmark problem. The gearbox is shown in Fig. 3. The assembly consists of 24 sequential and parallel tasks, encompassing multiple operation levels from initial inspection to final packaging. Each task can be performed by a human (H), a robot (R), or a human–robot collaborative mode (HR), depending on its complexity, precision requirement, and ergonomic load.

The assembly procedure is hierarchically organised into eight layers (L0–L7):

- L0 includes incoming inspection (INSP) and housing preparation (CASE), which serve as the foundation for subsequent operations. Mode: H, R. (Here, **H** and **R** denote *human-only* and *robot-only* execution modes, respectively, while **HR** represents human–robot collaborative execution.)
- L1 covers bearing and ring gear installation (BR1, BR2, RG1, RG2), enabling the mechanical interfaces of the gearbox. Mode: H, R, HR.
- L2 involves the first-stage gear and planetary assembly (SG1, PL1A–C, CA1). Mode: H, R, HR.
- L3 performs the integration of the first-stage subassembly with the housing (M1). Mode: H, R.
- L4–L5 focus on the second-stage gear and carrier installation (SG2, PL2A–C, CA2, M2, SHIM). Mode: H, R, HR.
- L6 includes lubrication, sealing, and torque tightening (LUB, TORQ, SEAL). Mode: H, R.
- L7 completes end-of-line testing and packaging (EOL, PKG). Mode: H, R.

Each task's execution time and human exertion level are modelled as SVTN, characterised by truth, indeterminacy, and falsity components. SVTNs are constructed by generating synthetic samples of task duration and physical exertion for human and robot operations and mapping these samples into the truth, indeterminacy, and falsity components using the empirical quantile formulation in Eqs. (4)–(6), where $\alpha = 0.2, \beta = 0.1$ and $\gamma = 0.05$. This representation captures the inherent stochasticity of assembly durations caused by process variability, human physical exertion, and environmental fluctuations. Task precedence relationships are explicitly defined. All computational tasks were performed on an NVIDIA GeForce RTX 3060 GPU to leverage its robust processing capabilities.

4.1.2. Ablation study

Experimental Objective To investigate the contribution of different objective components in the proposed DDQN-SVTN-based task planner, an ablation study is conducted focusing on the trade-off between **workload minimisation** and **makespan optimisation**. The full model jointly considers both human physical exertion and production efficiency through the multi-objective reward function, while two ablated variants disable parts of this formulation:

Table 1
Ablation study on workload and makespan performance

Method	C_V^{end}	C_A^{end}	F_V^{end}	F_A^{end}
Proposed (full model)	92.14	22.02	12.14	4.10
Makespan only (λ_f=0, η_f=0)	90.20	21.54	15.82	5.61
physical exertion only (ϕ_{cV}=0, κ_{cA}=0)	125.38	29.46	11.63	3.77

- Makespan only, where physical exertion-related coefficients (λ_f, η_f) are set to zero;
- physical exertion only, where time-related coefficients (ϕ_{cV}, κ_{cA}) are removed.

This setup allows isolating the impact of each optimisation term on both assembly duration and physical exertion accumulation.

In the ablation study, only one human and one robot were involved in the experiments. All key parameters used in the experiments were fixed before training and kept identical across all experimental variants to ensure a fair comparison. Specifically, the number of training episodes was set to `episodes` = 2000, which was sufficient to guarantee stable convergence of the learning process.

The coefficient λ_f was set to $\lambda_f = 10$ to control the relative importance of the physical exertion-related objective, while $\eta_f = 1$ was used as a normalisation factor to balance the numerical scale of physical exertion accumulation across tasks. In addition, the efficiency-related penalty coefficients were set to $\kappa_{cA} = 2$ and $\phi_{cV} = 2$.

These values were selected based on preliminary pilot experiments to ensure comparable magnitudes between makespan- and physical exertion-related reward components, rather than to optimise performance for any specific variant. In the ablation study, individual objectives were disabled solely by setting the corresponding coefficients to zero, without further parameter tuning.

Results and Discussion. As shown in Table 1, the proposed full model achieves a balanced performance between efficiency and ergonomic safety. When only the makespan objective is considered, the planner produces a slightly shorter completion time ($C_V^{end} = 90.2$) but at the cost of significantly higher workload levels ($F_V^{end} = 15.82$), indicating increased physical exertion accumulation for human operators. Conversely, the physical-exertion-only variant substantially reduces human workload ($F_V^{end} = 11.63$) but leads to a longer assembly duration ($C_V^{end} = 125.38$), reflecting a conservative scheduling policy that prioritises rest and recovery over efficiency. The complete model maintains a reasonable makespan (92.14) while effectively mitigating physical exertion growth ($F_V^{end} = 12.14$), demonstrating that integrating both objectives yields an optimal trade-off between productivity and worker well-being. In addition, the Gantt charts of the optimal schedules generated by the three planners are shown in Fig. 5. These results confirm the effectiveness of the proposed multi-objective reward design in guiding the DDQN-SVTN agent toward human-aware and uncertainty-resilient scheduling decisions.

4.1.3. Ablation Study on Uncertainty Modelling: SVTN versus TFN

Experimental Objective This set of experiments aims to conduct an ablation study on the uncertainty modelling component of the proposed DDQN-SVTN framework. Specifically, the SVTN-based uncertainty representation is replaced by a conventional Triangular Fuzzy Number (TFN) model. For a fair comparison, the TFN is constructed using the same empirical observation set. The single triangular fuzzy number is defined as

$$X_i^{\mathrm{TFN}} = \big(Q_i(\beta),\ Q_i(0.5),\ Q_i(1-\beta)\big), \tag{21}$$

where $Q_i(p)$ denotes the empirical p-quantile of the observed samples. $\beta = 0.05$.

All other components of the framework, including the DDQN architecture, state representation, reward formulation, and training procedure, are kept unchanged. By isolating the uncertainty modelling mechanism in this manner, the experiment explicitly evaluates the capability of integrating SVTN with deep reinforcement learning and its effect on uncertainty propagation in task planning.

Both variants are applied to the same gearbox assembly problem under identical Double DQN configurations. The evaluation focuses on the makespan ($C_{\max}$) and the cumulative human physical exertion (F). The mean value and standard deviation are used to quantify uncertainty amplification and stability across 100 runs.

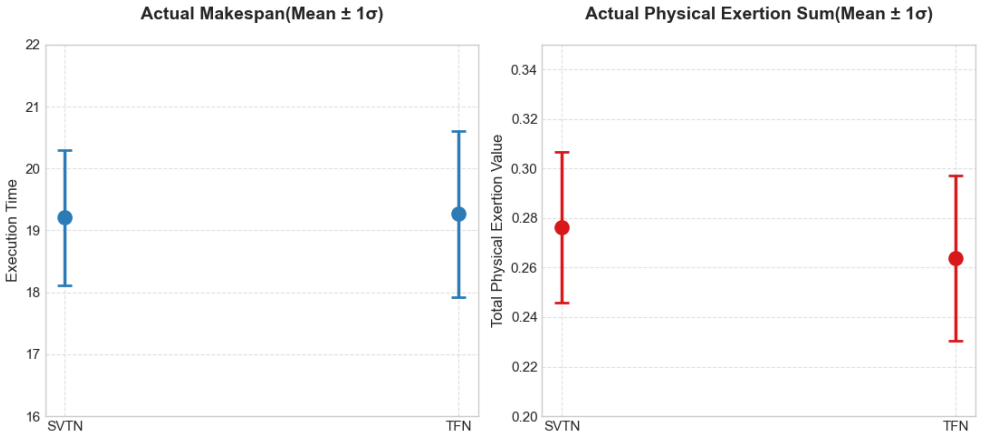


Figure 4: Ablation results on uncertainty modelling: SVTN versus TFN

In these experiments, a one-human–one-robot (1H1R) configuration is adopted to collaboratively perform the gearbox assembly task. The experimental and control groups use identical parameter settings, with $\lambda_f = 10$, $\eta_f = 1$, $\kappa_{cA} = 2$, and $\phi_{cV} = 2$. All models are trained for 2000 episodes to ensure convergence. The only difference between the two variants lies in the uncertainty modelling approach, namely SVTN versus TFN.

Results and Discussion Fig. 4 compares the actual makespan and accumulated physical exertion between the SVTN-based and TFN-based planners, with error bars representing the mean and standard deviation across repeated trials. As illustrated in the left panel, the average makespan is 19.21 ± 1.09 for SVTN and 19.27 ± 1.364 for TFN. Although the mean makespan values are nearly identical (with a difference of only 0.31%), a clearer distinction emerges in terms of variability. The standard deviation under SVTN is reduced from 1.364 to 1.09, corresponding to a **20.1% decrease in execution-time dispersion**.

A similar pattern is observed in the physical exertion results shown in the right panel. The accumulated physical exertion is 0.276 ± 0.030 for SVTN and 0.264 ± 0.033 for TFN. While the mean physical exertion differs slightly (approximately 4.5%), the variability under SVTN decreases from 0.033 to 0.030, representing a **9.1% reduction in physical exertion dispersion**. This demonstrates that the SVTN-based planner maintains comparable ergonomic performance while achieving improved stability. Overall, although the two methods exhibit similar average performance, the SVTN-based approach consistently reduces variance in both makespan and physical exertion.

4.1.4. Algorithm Comparison Study

Experimental Objective To further validate the effectiveness of the proposed uncertainty-aware DDQN-SVTN planner, this section compares it with a series of baseline algorithms widely used in human–robot collaborative scheduling. The baseline methods include: (1) rule-based scheduling (*Priority Rule* Haupt (1989)); (2) search-based optimisation (*Beam Search* Steinbiss, Tran and Ney (1994)); (3) evolutionary algorithms (*Genetic Algorithm, GA* Lambora, Gupta and Chopra (2019) and *NSGA-IIVerma, Pant and Snasel (2021)*); and (4) deep reinforcement learning approaches (*PPO* Yu, Velu, Vinitsky, Gao, Wang, Bayen and Wu (2022), *SAC* Haarnoja, Zhou, Abbeel and Levine (2018), and *GNN-based planner* Alablani and Alenazi (2023)). All methods are evaluated under four collaboration configurations: **1H1R**, **1H2R**, **2H2R**, and **2H4R**, representing increasing levels of task complexity and cooperation density. The comparison focuses on four metrics: the value degree of expected makespan (C_V^{end}), its uncertainty term (C_A^{end}), the value degree of expected cumulative physical exertion (F_V^{end}), and physical exertion uncertainty (F_A^{end}). Lower values indicate more efficient and stable task plans. All methods employ the same SVTN-based modeling approach, with no differences in any other aspects.

Results and Discussion. Table 2 summarises the quantitative results. Across all collaborative configurations, the proposed DDQN-SVTN consistently achieves the smallest or near-smallest values in both C_V^{end} and F_V^{end}, indicating superior optimisation of makespan and human workload. Compared with heuristic approaches (Priority and Beam

Table 2
Performance comparison of different algorithms under varying human–robot configurations

Method	1H1R		1H2R		2H2R		2H4R	
	C_V^{end}, C_A^{end}	F_V^{end}, F_A^{end}	C_V^{end}, C_A^{end}	F_V^{end}, F_A^{end}	C_V^{end}, C_A^{end}	F_V^{end}, F_A^{end}	C_V^{end}, C_A^{end}	F_V^{end}, F_A^{end}
Priority Rule	100.80, 23.28	13.71, 4.52	98.98, 22.86	11.69, 3.84	77.18, 17.88	7.40, 2.32	76.68, 17.88	6.66, 2.19
Beam Search	107.10, 24.54	13.40, 4.44	103.04, 23.88	11.63, 3.78	74.04, 17.40	7.48, 2.48	78.00, 17.94	6.14, 2.03
GA	96.54, 22.26	12.08, 4.04	95.72, 22.08	11.63, 3.77	75.14, 17.52	6.71, 2.13	76.34, 17.64	5.92, 1.99
NSGA-II	96.54, 22.26	12.08, 4.04	95.72, 22.08	11.63, 3.77	74.22, **17.34**	7.82, 2.37	73.30, 17.04	6.83, 2.18
PPO	102.44, 24.12	14.86, 4.97	98.24, 22.50	14.56, 4.84	78.10, 18.18	6.32, 1.98	71.20, 16.80	6.34, 2.07
SAC	98.70, 23.04	**11.63, 3.77**	92.30, 21.42	11.63, 3.77	75.74, 17.28	6.38, 2.07	73.90, 16.92	5.97, 1.95
GNN	93.62, **21.84**	12.62, 4.34	96.14, 22.50	11.63, 3.77	76.54,17.4	6.60,2.02	76.68,17.88	**5.69,1.77**
Proposed	**92.14**, 22.02	12.14, 4.10	**91.36, 21.30**	**11.63, 3.77**	**74.14**, 17.40	**6.16, 1.91**	**70.68, 16.38**	7.76, 2.32

Search), the proposed DDQN-SVTN method achieves up to a 13.97 % reduction in average makespan and a 17.64% reduction in cumulative physical exertion. Evolutionary algorithms (GA, NSGA-II) perform competitively but exhibit lower adaptability in stochastic tasks. In contrast, DDQN-SVTN effectively leverages experience replay and action masking to learn adaptive scheduling strategies that handle precedence and feasibility constraints. Compared to PPO and SAC, the DDQN-SVTN's value decoupling mechanism prevents overestimation and improves convergence stability, especially under high uncertainty. Overall, the DDQN-SVTN demonstrates strong scalability from 1H1R to 2H4R configurations, achieving balanced performance across efficiency and ergonomics.

Figure 5 illustrates the final task allocation schedule generated by the DDQN-SVTN planner under the 1H1R configuration. Tasks are grouped by execution mode (H, R, HR), and each colored bar represents a distinct assembly operation. The resulting Gantt chart shows a clear temporal coordination pattern between humans and robots, achieving an optimised makespan of approximately 95.5 units.

Figures 6 show the learning curves of the DDQN-SVTN, GNN, PPO, and SAC algorithms across different human–robot configurations. Each curve represents the best accumulated reward over 2000 training episodes. Across all scenarios, the DDQN-SVTN exhibits the fastest and most stable convergence behaviour, achieving near-optimal rewards within approximately 500–800 episodes. GNN-based learning follows a similar trend but converges slightly slower due to graph aggregation overhead. Although PPO and SAC exhibit fast convergence, their final performance consistently falls short of that achieved by the proposed method.

These results confirm the superior convergence efficiency and generalisation capability of the proposed DDQN-SVTN planner for complex multi-agent, multi-mode human–robot collaborative assembly systems.

4.2. Real-world HRCA experiment

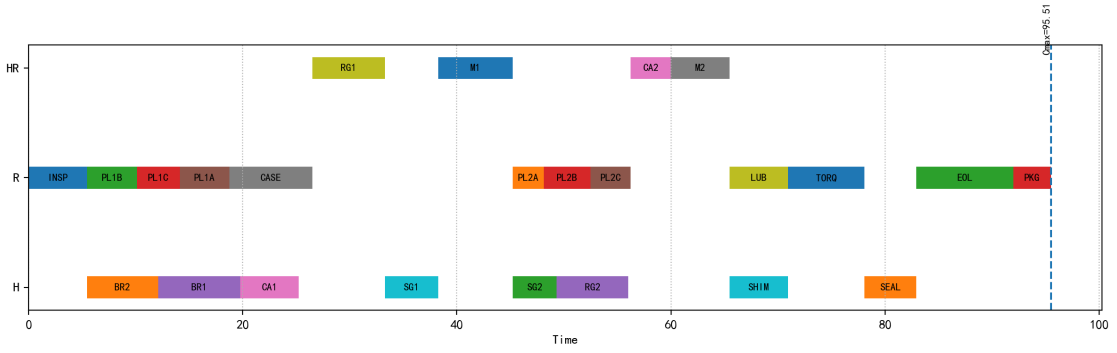
4.2.1. Experiment setup

A proof-of-concept, laboratory-scale PC tower human–robot collaborative assembly experiment was designed to validate whether incorporating SVTN-based uncertainty modelling improves the robustness of scheduling performance compared with a deterministic RL baseline.

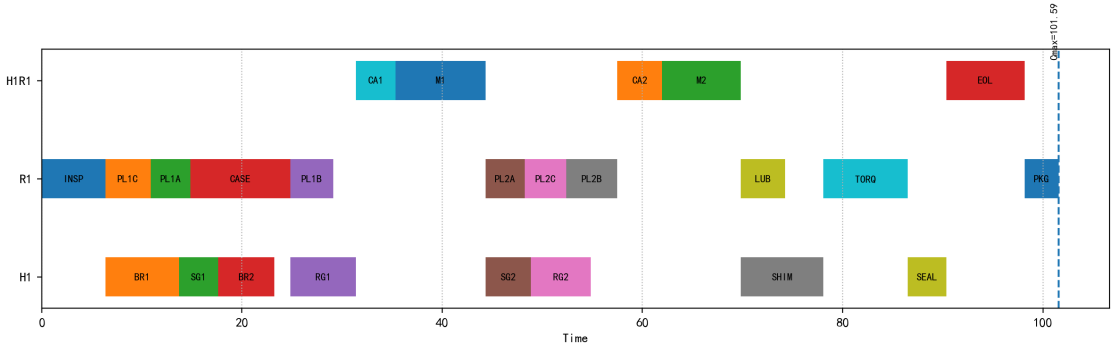
The PC tower used in the experiment, as shown in Fig. 7a, consists of eight core components, including the motherboard, CPU, GPU, and other key parts. The assembly precedence constraints of the PC tower are represented using an AND–OR graph, as illustrated in Fig. 7b. During the experiment, the assembly task is completed collaboratively by one human and one robot, with both human and robot operations scheduled by the proposed method.

Four participants were recruited as assembly workers (two male and two female, aged 25–32). Their operational data serve as an important data source for SVTN modelling. Each participant performed five independent PC tower assembly trials. The physical exertion states of the participants were analysed using Eq. (2). Subsequently, Eqs. (4)–(6) were employed to construct SVTNs for subtask execution time and physical exertion. In addition, SVTNs were constructed to model the robot execution time for each subtask. Based on these data, Algorithm 1 was used to train the policy for 2000 episodes. The trained policy was then deployed on a computer and executed via the ROS framework to control the robotic manipulator. The training parameters were kept identical to those used in the numerical study.

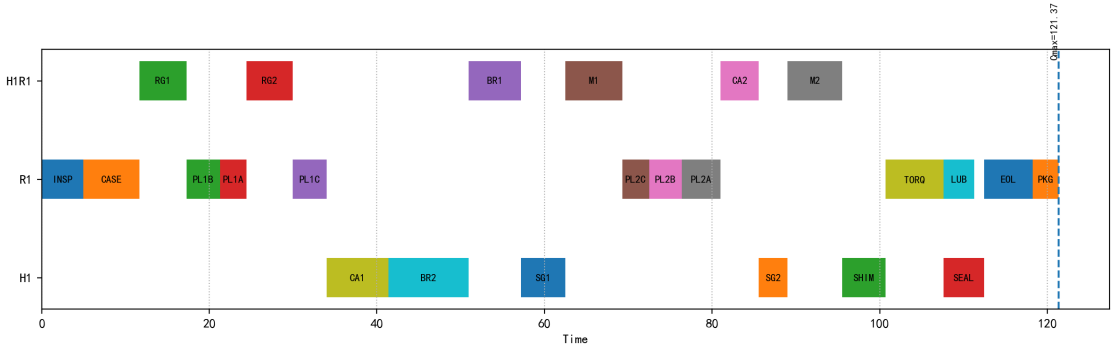
The experiment was designed with an experimental group and a control group. Both groups employed reinforcement-learning-based planners. The experimental group adopted the proposed DDQN-SVTN planner, while the control group used a standard DDQN planner. Specifically, the experimental group modelled uncertainty using SVTNs, whereas no uncertainty modelling was incorporated in the control group. For the control group, the state representation was derived



(a) Full model



(b) Makespan only



(c) physical exertion only

Figure 5: Gantt charts of the optimised 1H1R human–robot collaborative assembly schedules generated by the full model, makespan-only, and physical-exertion-only planners.

from Eq. (18) by removing the SVTN-related components, and the reward function aimed to minimise both makespan and cumulative physical exertion. All other parameters were kept identical to those of the experimental group. Each participant took part in both the experimental and control groups, performing five trials in each setting. The task makespan and accumulated physical exertion were recorded and analysed to evaluate the performance advantages of the proposed method.

4.2.2. Experiment results

A human–robot collaborative assembly process of a PC tower is used as an illustrative example. The assembly procedure consists of eight sequential tasks, shown in Fig. 8. Specifically, the human operator is responsible for

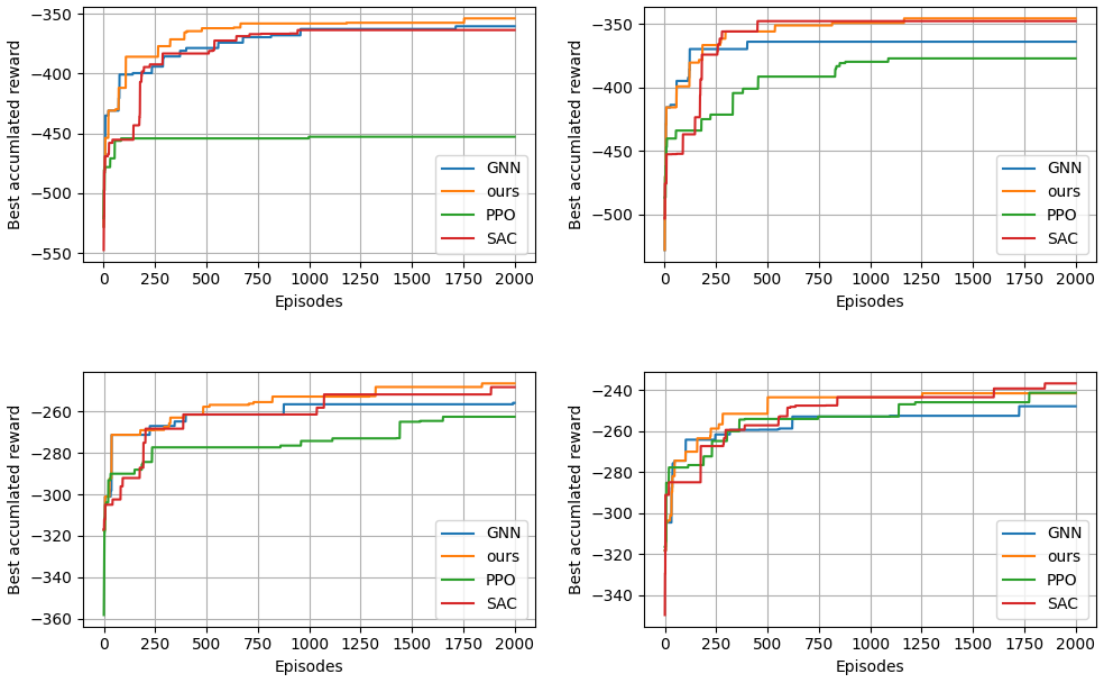
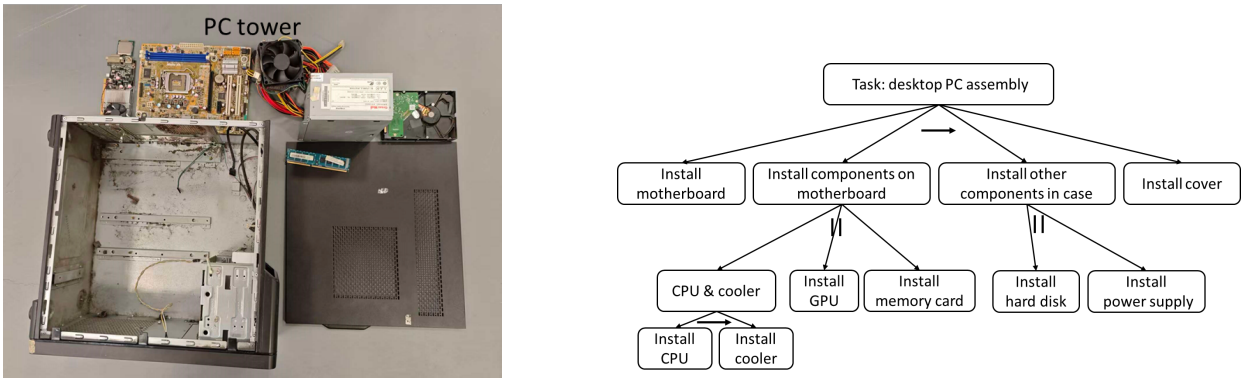


Figure 6: Training convergence of the proposed planner, GNN, PPO, and SAC under four configurations: (a) 1H1R, (b) 1H2R, (c) 2H2R, (d) 2H4R.



(a) PC tower (b) And–Or graph

Figure 7: The PC tower used in the real human–robot collaborative (HRC) experiment and the corresponding And–Or graph representing the assembly requirements.

installing the motherboard, the CPU, and the GPU, as well as the hard disk, while the robot installs the cooler, the memory card, the power supply, and the case cover.

Table 3 compares the proposed DDQN-SVTN with the baseline DDQN in real-world PC tower assembly experiments. While the two methods exhibit comparable mean performance in task duration and physical exertion, a clearer distinction emerges in terms of variability across repeated trials. For all participants, DDQN-SVTN consistently yields smaller standard deviations in task duration, indicating more stable execution outcomes. A similar trend is observed for physical exertion, where DDQN-SVTN generally reduces dispersion compared with DDQN. Quantitatively, for task duration, the average standard deviation across participants decreases from 17.76 (DDQN) to 12.00 (DDQN-SVTN), corresponding to a 32.4% reduction in execution-time variability. Similarly, for accumulated physical exertion,

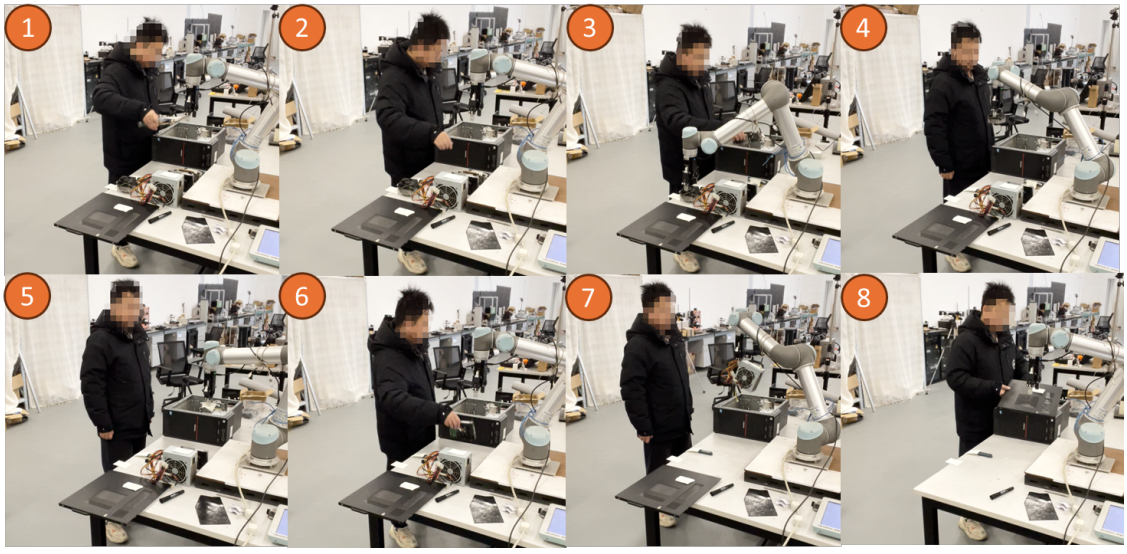


Figure 8: A representative human–robot collaborative assembly process for a PC tower is illustrated as follows: (1) the human installs the motherboard; (2) the human installs the CPU; (3) the human installs the GPU; (4) the robot installs the cooler; (5) the robot installs the memory card; (6) the human installs the hard disk drive; (7) the robot installs the power supply; and (8) the robot installs the case cover.

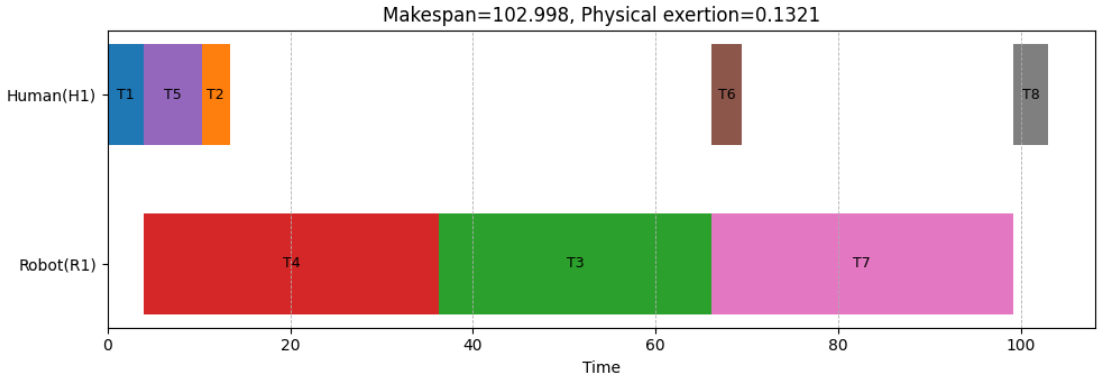
Table 3
Comparison of task duration and physical exertion between DDQN-SVTN and DDQN

Participant	DDQN-SVTN (Ours)		DDQN	
	Duration	physical exertion	Duration	physical exertion
Participant 1	109.60 ± 12.29	0.2024 ± 0.051	95.33 ± 17.29	0.331 ± 0.072
Participant 2	103.36 ± 13.76	0.170 ± 0.034	88.14 ± 17.38	0.240 ± 0.045
Participant 3	77.86 ± 11.48	0.234 ± 0.091	102.05 ± 19.25	0.204 ± 0.119
Participant 4	74.19 ± 10.48	0.159 ± 0.040	92.02 ± 17.12	0.120 ± 0.058

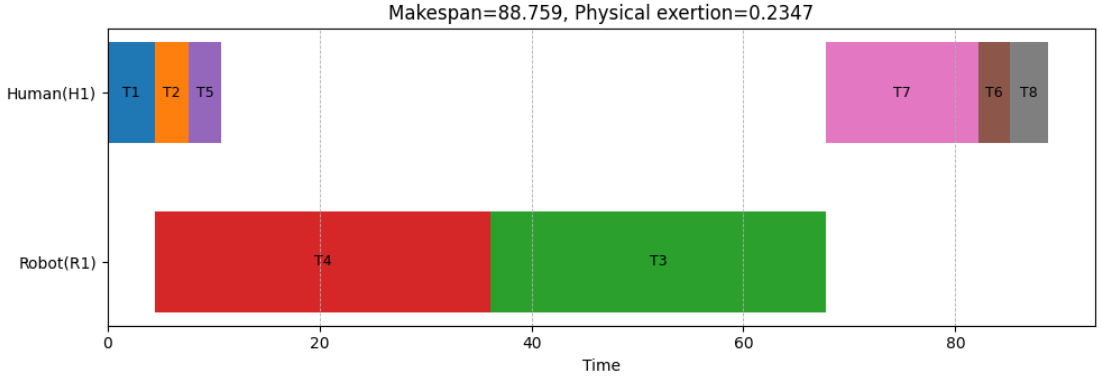
the average standard deviation decreases from 0.0735 to 0.054, representing a 26.5% reduction in physical exertion variability.

The uncertainty-resilient behaviour of the proposed DDQN-SVTN is further illustrated by the scheduling difference observed for task T7 (power supply installation) in Fig. 9. This task exhibits duration variability because it requires precise alignment within a confined workspace. In the DDQN-SVTN solution, T7 is preferentially assigned to the robot, whereas the baseline DDQN allocates it to the human operator in order to exploit the higher average execution efficiency.

This contrast reveals the underlying mechanism driving the robustness of the proposed method. The DDQN primarily optimises expected performance and therefore selects the seemingly faster human execution mode, but it becomes susceptible to the large variance associated with this task. In contrast, DDQN-SVTN explicitly incorporates execution-time uncertainty through SVTN modelling and ambiguity-aware reward design, which biases the learned policy toward more stable task–agent pairings. By assigning high-variance operations such as T7 to the robot, the proposed approach avoids amplifying fluctuations, leading to schedules with lower dispersion and improved repeatability. These results indicate that DDQN-SVTN enhances robustness not by maximising short-term efficiency, but by systematically mitigating uncertainty propagation in human–robot collaborative assembly.



(a) DDQN-SVTN



(b) DDQN

Figure 9: Gantt charts of the optimised 1H1R human-robot collaborative assembly schedules generated by DDQN-SVTN and DDQN planners. T1: motherboard, T2: CPU, T3:cooler, T4: memory card, T5: GPU, T6: hard disk, T7: power supply, T8: cover.

5. Conclusion

This study proposes a physical exertion–aware task planning framework for human–robot collaborative assembly under uncertainty, integrating Single-Valued Triangular Neutrosophic modelling with a DDQN optimisation approach. The SVTN model effectively captured multi-source uncertainties arising from task duration variability and human physical exertion, while the DDQN-based planner achieved adaptive and balanced scheduling decisions through a multi-objective reward design. Experimental results demonstrated that the proposed method significantly reduced uncertainty propagation compared with the traditional fuzzy approach and achieved superior performance over heuristic, evolutionary, and reinforcement learning baselines. Specifically, the DDQN-SVTN framework achieved shorter makespan and lower cumulative physical exertion, ensuring both efficiency and operator well-being. Overall, this research contributes to advancing human-centric manufacturing by providing an uncertainty-resilient, learning-based planning paradigm that bridges the gap between productivity optimisation and sustainable human–robot collaboration.

References

Alablani, I., Alenazi, M.J., 2023. DQN-GNN-based user association approach for wireless networks. Mathematics 11, 4286.

Alirezazadeh, S., Alexandre, L., 2022. Dynamic Task Scheduling for Human-Robot Collaboration. IEEE Robotics and Automation Letters 7, 8699–8704. doi:10.1109/lra.2022.3188906.

Altundas, B., Wang, Z., Bishop, J., Gombolay, M., 2022. Learning Coordination Policies over Heterogeneous Graphs for Human-Robot Teams via Recurrent Neural Schedule Propagation. 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) , 11679–11686doi:10.1109/IROS47612.2022.9981748.

Amirnia, A., Keivanpour, S., 2024. Real-time sustainable cobotic disassembly planning using fuzzy reinforcement learning. International Journal of Production Research 63, 3798 – 3821. doi:10.1080/00207543.2024.2431172.

Baratta, A., Cimino, A., Longo, F., Mirabelli, G., Nicoletti, L., 2024. Task Allocation in Human-Robot Collaboration: A Simulation-based approach to optimize Operator's Productivity and Ergonomics. Procedia Computer Science 232, 688–697.

Casalino, A., Mazzocca, E., Giorgio, M.G.D., Zanchettin, A., Rocco, P., 2019. Task scheduling for human-robot collaboration with uncertain duration of tasks: a fuzzy approach. 2019 7th International Conference on Control, Mechatronics and Automation (ICCMA) , 90–97doi:10.1109/ICCMA46720.2019.8988735.

Chand, S., Lu, Y., 2023. Dual task scheduling strategy for personalized multi-objective optimization of cycle time and fatigue in human-robot collaboration. Manufacturing Letters 35, 88–95.

Chen, Z., Zou, J., Wang, W., 2023. Digital twin-oriented collaborative optimization of fuzzy flexible job shop scheduling under multiple uncertainties. Sādhanā 48, 1–15. doi:10.1007/s12046-023-02133-z.

Djogdom, G.V.T., Meziane, R., Otis, M.J.D., 2024. Robust dynamic robot scheduling for collaborating with humans in manufacturing operations. Robotics Comput. Integr. Manuf. 88, 102734. doi:10.1016/j.rcim.2024.102734.

Ehrlich, S.K., Dean-Leon, E., Tacca, N., Armleder, S., Dimova-Edeleva, V., Cheng, G., 2023. Human-robot collaborative task planning using anticipatory brain responses. Plos one 18, e0287958.

Fujita, T., 2024. Advancing Uncertain Combinatorics through Graphization, Hyperization, and Uncertainization: Fuzzy, Neutrosophic, Soft, Rough, and Beyond. ArXiv abs/2411.17411. doi:10.48550/arXiv.2411.17411.

Ghorbani, E., Keivanpour, S., Sekkay, F., Imbeau, D., 2024. Human-centric robotic assembly line design: a fuzzy inference system approach for adaptive workload management. The International Journal of Advanced Manufacturing Technology doi:10.1007/s00170-024-14282-4.

Guo, B., Liu, H., Niu, L., 2023. Safe physical interaction with cobots: a multi-modal fusion approach for health monitoring. Frontiers in Neurorobotics 17. doi:10.3389/fnbot.2023.1265936.

Haarnoja, T., Zhou, A., Abbeel, P., Levine, S., 2018. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor, in: International conference on machine learning, Pmlr. pp. 1861–1870.

Haupt, R., 1989. A survey of priority rule-based scheduling. Operations-Research-Spektrum 11, 3–16.

Hou, W., Xiong, Z., Yue, M., Chen, H., 2024. Human-robot collaborative assembly task planning for mobile cobots based on deep reinforcement learning. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 238, 11097–11114.

Jia, Z., Xie, S., Zhang, W., 2025. Flexible task assignment and assembly scheduling for human-robot collaboration cell considering uncertainty. International Journal of Production Research , 1–21.

Keshvarparast, A., Berti, N., Chand, S., Guidolin, M., Lu, Y., Battaïa, O., Xu, X., Battini, D., 2024. Ergonomic design of Human-Robot collaborative workstation in the Era of Industry 5.0. Comput. Ind. Eng. 198, 110729. doi:10.1016/j.cie.2024.110729.

Kumar, R.S., Jujjavarapu, S., Esfahani, E., 2021. Fatigue Detection for Human Aware Adaptation in Human-Robot Collaboration. Volume 2: 41st Computers and Information in Engineering Conference (CIE) doi:10.1115/detc2021-70975.

Lambora, A., Gupta, K., Chopra, K., 2019. Genetic algorithm-A literature review, in: 2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon), IEEE. pp. 380–384.

Lamon, E., Fusaro, F., De Momi, E., Ajoudani, A., 2023. A unified architecture for dynamic role allocation and collaborative task planning in mixed human-robot teams. arXiv preprint arXiv:2301.08038 .

Li, X., Xu, W., Yao, B., Ji, Z., Liu, X., 2022. Dynamic task reallocation in human-robot collaborative workshop based on online biotic fatigue detection, in: 2022 IEEE 18th International Conference on Automation Science and Engineering (CASE), IEEE. pp. 116–122.

Liau, Y.Y., Ryu, K., 2020. Task allocation in human-robot collaboration (HRC) based on task characteristics and agent capability for mold assembly. Procedia manufacturing 51, 179–186.

Liu, G., de Winter, J., Merckaert, K., Steckelmacher, D., Nowe, A., Vanderborght, B., 2025. A Task-Efficient Reinforcement Learning Task-Motion Planner for Safe Human-Robot Cooperation. arXiv preprint arXiv:2510.12477 .

Liu, Z., Liu, Q., Wang, L., Xu, W., Zhou, Z., 2021. Task-level decision-making for dynamic and stochastic human-robot collaboration based on dual agents deep reinforcement learning. The International Journal of Advanced Manufacturing Technology 115, 3533 – 3552. doi:10.1007/s00170-021-07265-2.

Lorenzini, M., Lagomarsino, M., Fortini, L., Gholami, S., Ajoudani, A., 2023. Ergonomic human-robot collaboration in industry: A review. Frontiers in Robotics and AI 9. doi:10.3389/frobt.2022.813907.

Maderna, R., Pozzi, M., Zanchettin, A., Rocco, P., Prattichizzo, D., 2022. Flexible scheduling and tactile communication for human-robot collaboration. Robotics Comput. Integr. Manuf. 73, 102233. doi:10.1016/J.RCIM.2021.102233.

Messeri, C., Bicchi, A., Zanchettin, A., Rocco, P., 2022. A Dynamic Task Allocation Strategy to Mitigate the Human Physical Fatigue in Collaborative Robotics. IEEE Robotics and Automation Letters 7, 2178–2185. doi:10.1109/LRA.2022.3143520.

Munzer, T., Toussaint, M., Lopes, M., 2017. Preference learning on the execution of collaborative human-robot tasks, in: 2017 IEEE International Conference on Robotics and Automation (ICRA), IEEE. pp. 879–885.

Noormohammadi-Asl, A., Smith, S.L., Dautenhahn, K., 2024. To Lead or to Follow? Adaptive Robot Task Planning in Human-Robot Collaboration. ArXiv abs/2401.01483. doi:10.48550/arXiv.2401.01483.

Oliff, H., Liu, Y., Kumar, M., Williams, M., Ryan, M., 2020. Reinforcement learning for facilitating human-robot-interaction in manufacturing. Journal of Manufacturing Systems 56, 326–340.

Peternel, L., Tsagarakis, N., Caldwell, D., Ajoudani, A., 2017. Robot adaptation to human physical fatigue in human–robot co-manipulation. Autonomous Robots 42, 1011 – 1021. doi:10.1007/s10514-017-9678-1.

Pregina, K., Vasugi, K., Kannan, M.R., 2025. Neutrosophic fuzzy-based graphical evaluation and review technique for construction scheduling for residential buildings. Intelligent Decision Technologies doi:10.1177/18724981251333021.

Proia, S., Carli, R., Cavone, G., Dotoli, M., 2022. Control Techniques for Safe, Ergonomic, and Efficient Human-Robot Collaboration in the Digital Industry: A Survey. IEEE Transactions on Automation Science and Engineering 19, 1798–1819. doi:10.1109/tase.2021.3131011.

Pupa, A., Landi, C.T., Bertolani, M., Secchi, C., 2020. A dynamic architecture for task assignment and scheduling for collaborative robotic cells, in: International workshop on human-friendly robotics, Springer. pp. 74–88.

Pupa, A., Secchi, C., 2021. A Safety-Aware Architecture for Task Scheduling and Execution for Human-Robot Collaboration*. 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) , 1895–1902doi:10.1109/IROS51168.2021.9636855.

Pupa, A., Van Dijk, W., Brekelmans, C., Secchi, C., 2022. A resilient and effective task scheduling approach for industrial human-robot collaboration. Sensors 22, 4901.

Smit, I.G., Bukhsh, Z., Pechenizkiy, M., Alogariastos, K., Hendriks, K., Zhang, Y., 2024. Learning Efficient and Fair Policies for Uncertainty-Aware Collaborative Human-Robot Order Picking. ArXiv abs/2404.08006. doi:10.48550/arxiv.2404.08006.

Steinbiss, V., Tran, B.H., Ney, H., 1994. Improvements in beam search., in: ICSLP, pp. 2143–2146.

Verma, S., Pant, M., Snasel, V., 2021. A comprehensive review on NSGA-II for multi-objective combinatorial optimization problems. IEEE access 9, 57757–57791.

Wang, J., Yan, Y., Hu, Y., Yang, X., 2025a. A reinforcement learning from human feedback based method for task allocation of human robot collaboration assembly considering human preference. Advanced Engineering Informatics 66, 103497.

Wang, J., Yan, Y., Hu, Y., Yang, X., Zhang, L., 2025b. A transfer reinforcement learning and digital-twin based task allocation method for human-robot collaboration assembly. Eng. Appl. Artif. Intell. 144, 110064. doi:10.1016/j.engappai.2025.110064.

Wilcox, R.J., 2013. Flexible schedule optimization for human-robot collaboration. PhD Thesis. Massachusetts Institute of Technology.

Wu, B., Hu, B., Lin, H., 2017. A learning based optimal human robot collaboration with linear temporal logic constraints. arXiv preprint arXiv:1706.00007 .

Xie, Z., Lu, L., Wang, H., Su, B., Liu, Y., Xu, X., 2022. Mitigating the risk of musculoskeletal disorders during human robot collaboration: a reinforcement learning approach. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 66, 1543 – 1547. doi:10.1177/1071181322661151.

Xie, Z., Lu, L., Wang, H., Su, B., Liu, Y., Xu, X., 2023. Improving Workers' Musculoskeletal Health During Human-Robot Collaboration Through Reinforcement Learning. Human Factors 66, 1754 – 1769. doi:10.1177/00187208231177574.

Yao, B., Li, X., Ji, Z., Xiao, K., Xu, W., 2024. Task reallocation of human-robot collaborative production workshop based on a dynamic human fatigue model. Computers & Industrial Engineering 189, 109855.

You, Y., Cai, B., Pham, D., Liu, Y., Ji, Z., 2025. A Human Digital Twin Approach for Fatigue-Aware Task Planning in Human-Robot Collaborative Assembly. Computers & Industrial Engineering .

You, Y., Ji, Z., 2025. Towards human-centric manufacturing: A reinforcement learning method for physical exertion alleviation in HRCA. Robotics and Computer-Integrated Manufacturing .

Yu, C., Velu, A., Vinitsky, E., Gao, J., Wang, Y., Bayen, A., Wu, Y., 2022. The surprising effectiveness of ppo in cooperative multi-agent games. Advances in neural information processing systems 35, 24611–24624.

Yu, T., Huang, J., Chang, Q., 2020. Mastering the working sequence in human-robot collaborative assembly based on reinforcement learning. IEEE Access 8, 163868–163877.

Yu, T., Huang, J., Chang, Q., 2021. Optimizing task scheduling in human-robot collaboration with deep multi-agent reinforcement learning. Journal of Manufacturing Systems 60, 487–499.

Zhang, K., Li, X., 2014. Human-robot Team Coordination That Considers Human Fatigue. International Journal of Advanced Robotic Systems 11. doi:10.5772/58228.

Zhang, M., Li, C., Shang, Y., Liu, Z., 2022a. Cycle time and human fatigue minimization for human-robot collaborative assembly cell. IEEE Robotics and Automation Letters 7, 6147–6154.

Zhang, R., Lv, Q., Li, J., Bao, J., Liu, T., Liu, S., 2022b. A reinforcement learning method for human-robot collaboration in assembly tasks. Robotics and Computer-Integrated Manufacturing 73, 102227.

Zhang, Z.Y., Chen, Y.F., Xu, D., Gong, S.Y., Meng, Y.L., 2020. Research on Multi-robot Task Allocation Algorithm Based on HADTQL, in: 2020 International Workshop on Electronic Communication and Artificial Intelligence (IWECAI), IEEE. pp. 164–169.

Zheng, H., Chand, S., Keshvarparast, A., Battini, D., Lu, Y., 2023. Video-based fatigue estimation for human-robot task allocation optimisation, in: 2023 IEEE 19th International Conference on Automation Science and Engineering (CASE), IEEE. pp. 1–6.

Zhou, J., Zheng, L., Fan, W., 2024. Multirobot collaborative task dynamic scheduling based on multiagent reinforcement learning with heuristic graph convolution considering robot service performance. Journal of Manufacturing Systems 72, 122–141.