Digital Twin and Semantic-Aware Multi-Agent RL for Maritime Search and Rescue Operations

(Invited Paper)

Bishmita Hazarika*, Octavia A. Dobre*, and Trung Q. Duong*,†
*Memorial University, Canada, e-mail:{bhazarika, odobre, tduong}@mun.ca
†Queen's University Belfast, UK, e-mail: trung.q.duong@qub.ac.uk

***Abstract*—Effective maritime search and rescue (SAR) requires fast, coordinated action from Internet of Maritime Things (IoMT) nodes operating under extreme communication, energy, and environmental constraints. Existing solutions treat semantic sensing, digital twin modeling, and decentralized control in isolation, limiting their responsiveness and scalability. We propose SEMADT-RL, a unified framework that integrates semantic-driven communication, predictive digital twin forecasting, and decentralized multi-agent deep reinforcement learning with graph attention networks (MADRL-GNN). The semantic layer enables lightweight, anomaly-triggered updates, significantly reducing bandwidth while preserving critical detection cues. The digital twin assimilates these updates using an extended Kalman filter to forecast survivor drift and node dynamics. These forecasts guide decentralized agents that collaboratively optimize mobility, processing, and transmission policies under dynamic and constrained maritime conditions. Simulation results demonstrate that SEMADT-RL achieves faster survivor detection, lower communication overhead, and higher energy efficiency than state-of-the-art baselines, providing a scalable solution for next-generation IoMT-assisted SAR operations.**

I. INTRODUCTION

MARITIME search and rescue (SAR) operations face significant challenges in locating and assisting survivors dispersed over vast oceanic regions. Traditional sensor networks and manual search efforts are often limited by communication bottlenecks, dynamic environmental factors, and constrained energy resources. With the emergence of the Internet of Maritime Things (IoMT), a network of semi-autonomous buoys, surface vehicles, and drones, a new opportunity has arisen to coordinate distributed assets for timely and efficient SAR missions [1]. However, IoMT deployments must address stringent requirements on communication efficiency, decentralized control, and environmental adaptability under highly dynamic maritime conditions. Despite offering expanded sensing and actuation capabilities, IoMT nodes inherently suffer from limitations in bandwidth, energy, and onboard processing power. In particular, transmitting full-resolution sensor data in real-time can quickly exhaust wireless capacity and battery resources, leading to significant delays or even mission failures [2]. This motivates a shift toward more efficient communication paradigms that prioritize critical information over raw data transfer.

This work was supported by the Canada Excellence Research Chair (CERC) Program CERC-2022-00109. The work of O. A. Dobre was supported by the Canada First Research Excellence Fund, Transforming Climate Action, TCA-LRP-20230-1.3-5, and Canada Research Chairs Program CRC-2022-00187

In this regard, semantic communication has emerged as a promising strategy for efficient and task-driven information exchange. Unlike conventional systems that aim to transmit raw bit sequences, semantic communication focuses on transmitting only the meaning-bearing elements most critical to a given task [3]. Although originally inspired by human language communication, the concept of semantics has since evolved to encompass task-relevant features extracted from non-linguistic sources such as sensor readings, video frames, and radar data [4]. In autonomous and vision-based systems, this typically involves transmitting compact descriptors—such as detected anomalies, object positions, or latent semantic embeddings—instead of raw sensor streams. Recent works demonstrate that semantic-aware designs can significantly reduce bandwidth, lower energy consumption, and improve task completion rates across edge computing, vehicular networks, and UAV swarms [5]. However, directly applying semantic compression in dynamic maritime environments presents unique challenges: high mobility, uncertain drift dynamics, strict survivor detection deadlines, and severe energy constraints.

Another critical challenge lies in the lack of predictive situational awareness among IoMT nodes. While semantic encoding compresses information, nodes often act only on delayed or incomplete local observations, leading to suboptimal decisions and wasted movements [6]. To mitigate this, digital twin (DT) technology, a virtual replica that mirrors and forecasts the physical state of the environment, has recently gained attention [7]. Works such as [8] have shown that DTs can enhance coordination in vehicular and industrial IoT settings. Yet, in maritime rescue scenarios, existing DT implementations either assume high-fidelity real-time data availability or neglect the sparse, noisy observation patterns typical of open-sea operations [9].

Moreover, decentralized decision-making is paramount for the scalability and robustness of large IoMT fleets. While centralized reinforcement learning (RL) can optimize control policies with full-state information [10], it suffers from poor scalability and delayed responsiveness under real-world communication constraints [11]. On the other hand, multi-agent deep reinforcement learning (MADRL) has shown success in decentralized robotics and UAV networks [12], enabling agents to learn local policies based on partial information [13]. Nevertheless, standard MADRL methods are often handicapped by noisy sensing, lack of semantic abstraction, and an inability to anticipate future environmental changes [14].

However, although semantic communication, DTs, and MADRL have each shown independent potential for enhancing efficiency, situational awareness, and decentralized decision-making, existing research has largely pursued these technologies in isolation or in limited pairwise combinations. Crucially, none of the existing works has addressed the deep interdependence between semantic information extraction, predictive environment modeling, and distributed policy learning—an interdependence that becomes especially critical in dynamic, resource-constrained IoMT-based maritime search and rescue. In such environments, semantic compression without predictive forecasting risks missing critical survivor drift patterns, while digital twins without real-time semantic updates suffer from outdated or noisy state estimation. Similarly, decentralized control without semantic abstraction leads to inefficient sensing and excessive energy expenditure. Thus, the absence of a unified architecture that jointly optimizes compression, prediction, and policy learning not only limits performance but fundamentally constrains the scalability and resilience of IoMT SAR systems.

Motivated by the challenges of real-time survivor detection, constrained resources, and decentralized coordination in maritime environments, we propose SEMADT-RL: a unified framework that integrates semantic sensing, predictive digital twin forecasting, and decentralized MADRL based on graph attention networks. By coupling these components, the framework enables IoMT nodes to reason predictively about survivor dynamics, energy usage, and communication tradeoffs under uncertainty. The key contributions of this work are summarized as follows:

- We formulate a maritime search-and-rescue (SAR) problem that jointly optimizes survivor detection, energy efficiency, semantic fidelity, and communication reliability under mobility, energy, and sensing constraints, leading to a structured non-linear constrained optimization problem.
- We develop a lightweight semantic sensing layer where IoMT nodes encode high-dimensional sensor inputs into low-rate latent descriptors using a variational autoencoder (VAE), triggering transmission based on reconstruction-based anomaly detection to preserve bandwidth and energy.
- We construct a DT model using an Extended Kalman Filter (EKF) that fuses semantic observations to forecast survivor drift, environmental evolution, and energy dynamics, allowing anticipatory node coordination rather than reactive control.
- We design a decentralized actor–critic control layer using graph attention mechanisms, where each node aggregates local and neighbor information along with DT forecasts to optimize its mobility, computation, and communication decisions in real-time.
- Through simulations in a dynamic maritime environment, we validate that SEMADT-RL significantly improves survivor detection rates, reduces communication overhead, and sustains energy-efficient operations compared to non-predictive or non-semantic baselines.

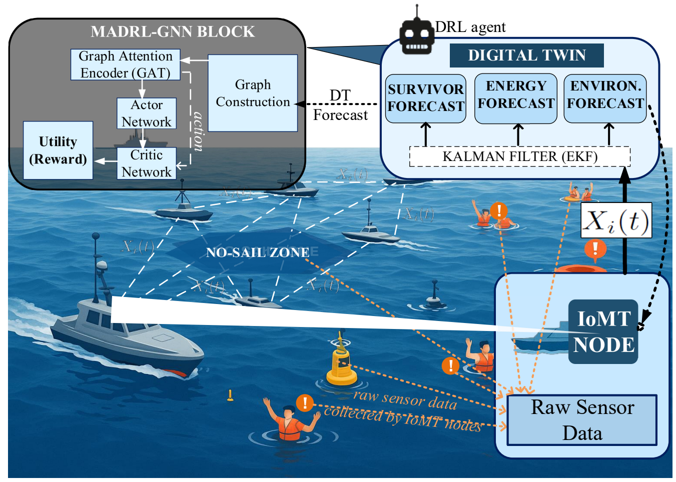

Fig. 1: Illustration of the maritime IoMT system.

II. SYSTEM MODEL

We propose a Search & Rescue (SAR) framework in a maritime domain, where a swarm of IoMT nodes—such as autonomous surface vehicles or specialized buoys—collaborates to locate and assist survivors following a distress event at sea. The operational environment is modeled as a two-dimensional region with potential no-sail zones (e.g., reefs or hazardous waters) and time-varying environmental factors (wave height, currents) that can displace survivors. Each IoMT node has limited energy, computation, and communication resources, compelling an intelligent approach to onboard data processing and node collaboration. Fig. 1 provides a schematic overview of the overall system architecture, highlighting the flow of raw sensing, semantic compression, wireless communication, and predictive coordination.

A. Maritime Environment

We consider a two-dimensional maritime region $\mathcal{R} \subset \mathbb{R}^2$ of size $\mathcal{M}_x \times \mathcal{M}_y$ meters. Let $\mathbf{p} = (x, y)$ denote coordinates within $\mathcal{R}$. Time is discretized into intervals $t \in \{0, 1, \ldots, T\}$ of duration Δt seconds. The region $\mathcal{R}$ may include no-sail zones $\mathcal{Z}(t) \subset \mathcal{R}$ that represent dangerous or restricted areas such as reefs or debris fields, and these zones can evolve over time due to changing conditions.

A distress event originates at a location $\mathbf{s}_{\text{init}} \in \mathcal{R}$. From this event, a set of survivors $\mathcal{S} = \{1, \ldots, N_{\text{surv}}\}$ is introduced, each with an initial position $\mathbf{s}_j(0) = \mathbf{s}_{\text{init}}$ or near it. Over time, survivor j drifts according to:

$$\mathbf{s}_j(t + \Delta t) = \mathbf{s}_j(t) + \mathbf{v}_{\text{drift}}(t)\,\Delta t + \boldsymbol{\eta}_j(t), \qquad (1)$$

where $\mathbf{v}_{\text{drift}}(t)$ captures prevailing sea currents and winds, and $\boldsymbol{\eta}_j(t)$ is a random offset representing local turbulence. Each survivor j must be detected or assisted within a critical survival time τ_j, beyond which rescue probabilities drop substantially.

B. IoMT Nodes

A set of N_{node} IoMT nodes, indexed by $i \in \{1, \ldots, N_{\text{node}}\}$, operates within $\mathcal{R}$. Each node may be an autonomous surface vessel, a buoy with partial mobility, or a coastal station acting as an edge server. The position of node i at time t is $\mathbf{p}_i(t)$, subject to

$$\|\mathbf{p}_i(t+\Delta t) - \mathbf{p}_i(t)\| \le v_{\max,i}\,\Delta t, \quad \mathbf{p}_i(t) \notin \mathcal{Z}(t), \tag{2}$$

where $v_{\max,i}$ is node i's maximum velocity, and $\mathcal{Z}(t)$ is the no-sail zone at time t. Each node has a maximum energy $E_{\max,i}$, a CPU frequency limit $f_{\max,i}$, and a communication power budget that governs how data is exchanged with other nodes or a central aggregator. The total energy consumed by node i from time t to $t+\Delta t$ is denoted $E_{\text{total},i}(t)$ and must not exceed $E_{\max,i}$.

C. Digital Twin Framework

A DT maintains an on-shore replica of both the maritime environment and the IoMT fleet. Its state snapshot at time t is

$$\mathcal{DT}(t) = \Big\{\mathbf{E}(t),\ \mathbf{S}_1(t), \ldots, \mathbf{S}_{N_{\text{node}}}(t)\Big\}, \tag{3}$$

where $\mathbf{E}(t) = \big[W(t), \Omega(t), \mathcal{Z}(t), \ldots\big]$ stacks the global environmental fields: significant wave height $W(t)$, wind vector $\Omega(t)$, and the current no-sail zone $\mathcal{Z}(t)$ (see Section II-A); $\mathbf{S}_i(t) = \big[\widehat{\mathbf{p}}_i(t), \widehat{E}_i(t), \widehat{\mathbf{v}}_i(t), \widehat{\gamma}_{i,j}(t), \ldots\big]$ holds node-specific forecasts: recommended future position $\widehat{\mathbf{p}}_i(t)$, projected residual energy $\widehat{E}_i(t)$, velocity advice $\widehat{\mathbf{v}}_i(t)$, and predicted SINR $\widehat{\gamma}_{i,j}(t)$ to neighbour j. Whenever new information $\mathbf{z}(t)$ (e.g. semantic packets, GNSS tracks) arrives, the DT advances according to

$$\mathcal{DT}(t+\Delta t) = \Phi\big(\mathcal{DT}(t), \mathbf{z}(t)\big) + \boldsymbol{\epsilon}(t), \tag{4}$$

where $\Phi(\cdot)$ implements physics-based prediction of environmental fields, survivor drift, and battery decay, and $\boldsymbol{\epsilon}(t)$ captures modelling error. Recommended actions such as $\widehat{\mathbf{v}}_i(t)$ and off-loading schedules are broadcast to the fleet; each node may follow or locally adapt the guidance.

D. Node Mobility and Energy Consumption

Node i updates its position according to (2), ensuring that it remains outside $\mathcal{Z}(t)$. The cost of moving from $\mathbf{p}_i(t)$ to $\mathbf{p}_i(t+\Delta t)$ is

$$E_{\text{mov},i}(t) = \mathcal{P}_{\text{mov},i}\,\big\|\mathbf{p}_i(t+\Delta t) - \mathbf{p}_i(t)\big\|, \tag{5}$$

where $\mathcal{P}_{\text{mov},i}$ is a movement power coefficient. The computation energy for local data processing at frequency $f_i(t)$ and processing time $T_{\text{proc},i}(t)$ is

$$E_{\text{comp},i}(t) = \kappa_i\,\big(f_i(t)\big)^2\,T_{\text{proc},i}(t), \tag{6}$$

with κ_i capturing hardware efficiency. Communication energy arises when node i transmits data at power $P_i(t)$ over time $T_{\text{tr},i}(t)$:

$$E_{\text{com},i}(t) = P_i(t)\,T_{\text{tr},i}(t). \tag{7}$$

The total energy for node i is

$$E_{\text{total},i}(t) = E_{\text{mov},i}(t) + E_{\text{comp},i}(t) + E_{\text{com},i}(t), \tag{8}$$

which must satisfy $E_{\text{total},i}(t) \le E_{\max,i}$.

E. Semantic Communication Model

Each node i carries sensors that capture high-volume data $\mathbf{o}_i(t)$, such as images or radar sweeps of the local ocean surface. Instead of transmitting raw data, node i uses an encoder

$$X_i(t) = \text{sem-enc}\big(\mathbf{o}_i(t)\big) \tag{9}$$

that extracts only the essential semantic content, for instance bounding boxes around detected objects or textual descriptors of anomalies. A channel encoder $\text{ch-enc}(\cdot)$ then modulates $X_i(t)$ for wireless transmission. If node j is the receiver, the received signal is

$$Y_{j,i}(t) = G_{i,j}(t)\,X_i'(t) + I_{j,i}(t) + N_j(t), \tag{10}$$

where $G_{i,j}(t)$ is the channel gain, $I_{j,i}(t)$ represents interference, and $N_j(t)$ is noise. The signal-to-interference-plus-noise ratio $\gamma_{j,i}(t)$ governs the achievable data rate

$$r_{j,i}(t) = B\,\log_2\!\big(1 + \gamma_{j,i}(t)\big). \tag{11}$$

If the semantic descriptor has size $D_i(t)$ bits, node i needs transmission time

$$T_{\text{tr},i}(t) = \frac{D_i(t)}{r_{j,i}(t)}. \tag{12}$$

At the receiver, the decoding process inverts ch-enc and applies $\text{sem-dec}(\cdot)$ to recover a compact representation $\hat{\mathbf{o}}_i^{(\text{sem})}$. A high overlap or cosine similarity between $\mathbf{o}_i(t)$ and $\hat{\mathbf{o}}_i^{(\text{sem})}$ indicates that the semantic meaning was effectively preserved.

F. Multi-Objective Utility Function

Each IoMT node aims to promptly detect survivors, efficiently conserve energy, maintain balanced computational workloads, and closely follow recommendations provided by the DT. We thus define the per-node utility $U_i(t)$ as the sum of four mathematical objectives:

$$U_i(t) = U_{\text{det},i}(t) + U_{\text{en},i}(t) + U_{\text{bal},i}(t) + C_{\text{DT},i}(t). \tag{13}$$

The survivor detection utility, $U_{\text{det},i}(t)$, quantifies how rapidly survivors within node i's sensing range at time t are detected. It explicitly rewards early detection and penalizes missed deadlines as follows:

$$U_{\text{det},i}(t) = \sum_{j \in \mathcal{S}_i(t)} \exp\left(\frac{\tau_j - \delta_{i,j}}{\tau_j}\right) - \mathbb{1}\{\delta_{i,j} > \tau_j\} \cdot C_{\text{plty}}, \tag{14}$$

where $\mathcal{S}_i(t)$ is the set of survivors detectable by node i at time t, $\delta_{i,j}$ denotes the elapsed time between the distress event and the detection of survivor j by node i, and τ_j is the critical survival deadline for survivor j. The indicator function $\mathbb{1}\{\delta_{i,j} > \tau_j\}$ applies a fixed penalty C_{plty} when survivor detection occurs after the deadline, strongly incentivizing prompt detection. The energy efficiency utility, $U_{\mathrm{en},i}(t)$, incentivizes minimal energy consumption and is normalized by node i's maximum onboard energy budget $E_{\max,i}$:

$$U_{\mathrm{en},i}(t) = 1 - \frac{E_{\mathrm{total},i}(t)}{E_{\max,i}}, \tag{15}$$

where $E_{\mathrm{total},i}(t)$ represents the total energy consumed by node i at time t (as detailed in (8)). The load-balancing utility, $U_{\mathrm{bal},i}(t)$, promotes equitable distribution of computational and detection workloads. It measures the alignment of node i's current load, $L_i(t)$, with the fleet-wide average load, $\overline{L}(t)$:

$$U_{\mathrm{bal},i}(t) = 1 - \frac{|L_i(t) - \overline{L}(t)|}{L_{\max}}, \tag{16}$$

where $L_{\max}$ normalizes the load difference, encouraging each node to maintain a balanced workload relative to the fleet average. Lastly, the DT compliance term, $C_{\mathrm{DT},i}(t)$, penalizes deviations between node i's actual velocity $\mathbf{v}_i(t)$ and the recommended velocity $\widehat{\mathbf{v}}_i(t)$ from the DT:

$$C_{\mathrm{DT},i}(t) = -\alpha_{\mathrm{DT}} \, \|\mathbf{v}_i(t) - \widehat{\mathbf{v}}_i(t)\|^2, \tag{17}$$

where the scaling parameter $\alpha_{\mathrm{DT}} > 0$ determines the penalty severity for deviations, ensuring nodes closely follow predictive guidance provided by the DT for globally coordinated mission execution.

III. Problem Formulation & Solution Approach

Over the finite horizon of T discrete decision steps, every IoMT node updates its planned positions $\{\mathbf{p}_i(t)\}_{t=0}^{T}$, the CPU frequency $f_i(t)$ applied to on-board semantic processing, and the transmit power $P_i(t)$ used for wireless links. These control variables jointly govern the rate at which survivors are detected, the energy consumed by propulsion, computation, and communication, and the reliability with which compressed semantic descriptors reach neighbouring nodes and the DT. Maximising the time-aggregated node utility defined in Section II-F therefore yields the operating policy that best balances detection speed, energy efficiency, load distribution. The problem can be formulated as:

$$(\mathcal{P}): \quad \max_{\mathbf{p}_i(\cdot),\, f_i(\cdot),\, P_i(\cdot)} \sum_{t=0}^{T} \sum_{i=1}^{N_{\text{node}}} U_i(t) \tag{18}$$

$$\begin{aligned}
&\text{(C.1)} \quad \frac{\|\mathbf{p}_i(t+\Delta t) - \mathbf{p}_i(t)\|}{\Delta t} \le v_{\max,i}, \quad \forall i,\ t,\\
&\text{(C.2)}\ \mathbf{p}_i(t) \notin \mathcal{Z}(t), \quad \forall i,\ t,\\
&\text{(C.3)}\ E_{\text{total},i}(t) \le E_{\max,i}, \quad \forall i,\ t,\\
&\text{(C.4)}\ \delta_{i,j} \le \tau_j, \quad \forall i,\ \forall j \in \mathcal{S},\\
&\text{(C.5)}\ f_i(t) \le f_{\max,i}, \quad \forall i,\ t,\\
&\text{(C.6)}\ \xi_{i,k}(t) \ge \xi_{\min}, \quad \forall\, k, i,\ t,\\
&\text{(C.7)}\ \Pr\big(\gamma_{j,i}(t) < \gamma_{\min}\big) \le \epsilon_{\text{mar}}, \quad \forall (i,j),\ t.
\end{aligned}$$

In the problem $(\mathcal{P})$, (C.1) limits node speed, (C.2) enforces avoidance of no-sail zones, (C.3) bounds total energy usage, (C.4) guarantees timely survivor detection, (C.5) respects CPU limits, (C.6) maintains semantic descriptor fidelity, and (C.7) ensures reliable communication via SINR outage control.

Solving $(\mathcal{P})$ requires joint optimization of mobility, computation, and communication under partial observability and strict constraints. Centralized methods are intractable, and standard RL lacks predictive context. To address this, in this study, we propose SemaDT-RL a scalable, three-layer framework combining semantic compression, digital twin forecasting, and decentralized multi-agent DRL for efficient, foresight-driven control in maritime IoMT networks.

IV. Proposed SemaDT-RL Framework

The goal of SemaDT-RL is to enable decentralized coordination of maritime IoMT nodes to perform efficient SAR under resource, safety, and latency constraints. The framework addresses the intractability of solving $(\mathcal{P})$ centrally by decomposing it into three tightly coupled components: (1) semantic sensing and anomaly detection, (2) predictive DT forecasting, and (3) decentralized control via MAGAC-GNN.

A. Preliminaries and Notation

Each node $i \in \{1, \ldots, N_{\text{node}}\}$ captures a raw observation $\mathbf{o}_i(t)$ at timestep t, e.g., an image or radar frame. The compressed latent representation $\mathbf{z}_i(t) \in \mathbb{R}^{64}$ is computed using a VAE, with anomaly detection triggering transmission of a semantic packet $X_i(t) = (\boldsymbol{\mu}_i(t), \mathbf{p}_i(t), t)$ to the DT. The global state of the environment is represented as $\mathbf{x}(t)$, including node positions $\mathbf{p}_i(t)$, energy $E_i(t)$, and survivor locations $\mathbf{s}_j(t)$. Each node's local state for decision-making is

$$\mathbf{s}_i(t) = [\mathbf{p}_i(t), E_i(t), L_i(t), \widehat{\mathbf{v}}_i(t), \widehat{E}_i(t)], \tag{19}$$

where $L_i(t)$ is processing load and $\widehat{\mathbf{v}}_i(t), \widehat{E}_i(t)$ are DT forecasts. The control action taken by each agent is:

$$\mathbf{a}_i(t) = [\Delta\mathbf{p}_i(t), f_i(t), P_i(t)], \tag{20}$$

where $\Delta\mathbf{p}_i(t)$ is the movement vector, $f_i(t)$ the CPU frequency, and $P_i(t)$ the transmit power.

B. Semantic Compression and Anomaly-Driven Updates

Each node trains a VAE encoder–decoder pair $\theta_i = \{\theta_i^{\text{enc}}, \theta_i^{\text{dec}}\}$ using the ELBO loss:

$$\mathcal{L}_{\text{VAE},i}(t) = \|\mathbf{o}_i - \hat{\mathbf{o}}_i\|^2 + \beta D_{\mathrm{KL}}(q_{\theta_i}(z|\mathbf{o}) \| p(z)), \tag{21}$$

where the KL divergence regularizes the latent space. An observation is flagged anomalous if

$$\text{err}_i(t) = \|\mathbf{o}_i(t) - \hat{\mathbf{o}}_i(t)\|^2 > \mathcal{T}_{\text{ano}}. \tag{22}$$

In that case, a semantic packet $X_i(t)$ is sent. Model updates $\Delta\theta_i(t)$ are transmitted if:

$$\|\Delta\theta_i(t)\|_2 > \varepsilon_{\text{sync}}, \tag{23}$$

and merged by the DT via FedAsync:

$$\theta_{\text{glob}} \leftarrow (1-\lambda)\theta_{\text{glob}} + \lambda \Delta\theta_i(t). \tag{24}$$

Algorithm 1 SEMADT-RL Algorithm

Initialize: θ_i, θ_{glob}, EKF state, π_θ, Q_ϕ
for $t = 0$ to T **do**
  for each node i **do**
    Capture $\mathbf{o}_i(t)$, compute anomaly via (22)
    if $\text{err}_i(t) > \mathcal{T}_{\text{ano}}$ **then**
      Transmit $X_i(t)$ to DT
    end if
    if (23) holds **then**
      Transmit $\Delta\theta_i(t)$, update via (24)
    end if
  end for
  DT runs EKF ((25)–(28)) and broadcasts forecasts
  for each node i **do**
    Form $\mathbf{s}_i(t)$, $\mathbf{h}_i(t)$ via (29)
    Select action via (30), apply it, compute $U_i(t)$
    Compute $Y_i(t)$, Q_ϕ, π_θ using (31), (32), (33).
  end for
end for

C. Extended Kalman Filter-Based Digital Twin Forecasting

The DT maintains and updates a latent system state $\mathbf{x}(t)$ using EKF. The process and measurement models are:

$$\mathbf{x}(t+\Delta t) = \mathbf{f}(\mathbf{x}(t)) + \mathbf{w}(t), \quad \mathbf{w}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}), \tag{25}$$

$$\mathbf{z}_i(t) = \mathbf{h}(\mathbf{x}(t)) + \mathbf{v}_i(t), \quad \mathbf{v}_i(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{R}). \tag{26}$$

The posterior estimate $\widehat{\mathbf{x}}(t|t)$ is used to forecast per-node control targets:

$$\widehat{\mathbf{v}}_i(t) = \arg \min_{\|\mathbf{v}\| \leq v_{\max}} \|\mathbf{p}_i(t) + \mathbf{v}\Delta t - \mathbf{s}_{\text{nn}}(t)\|, \tag{27}$$

$$\widehat{E}_i(t) = E_i(t) - (P_i(t) + P_{\text{mov}}\|\widehat{\mathbf{v}}_i(t)\|)\Delta t. \tag{28}$$

D. Decentralized Multi-Agent Actor-Critic with GAT

Each agent constructs a graph-augmented state from local and neighborhood inputs:

$$\mathbf{h}_i(t) = \text{GAT}(\{\mathbf{s}_j(t) : j \in \mathcal{N}_i(t)\}), \quad \mathbf{x}_i(t) = [\mathbf{s}_i(t), \mathbf{h}_i(t)]. \tag{29}$$

The actor selects an action:

$$\mathbf{a}_i(t) \sim \pi_\theta(\mathbf{x}_i(t)), \tag{30}$$

and the critic evaluates it via Bellman backup:

$$Y_i(t) = U_i(t) + \gamma Q'_\phi(\mathbf{x}_i(t+1), \pi'_\theta(\mathbf{x}_i(t+1))). \tag{31}$$

The loss functions are:

$$\mathcal{L}_{\text{critic}} = \big(Q_\phi(\mathbf{x}_i(t), \mathbf{a}_i(t)) - Y_i(t)\big)^2, \tag{32}$$

$$\nabla_\theta J(\theta) = \nabla_\theta \pi_\theta(\mathbf{x}_i(t)) \nabla_{\mathbf{a}_i} Q_\phi(\mathbf{x}_i(t), \mathbf{a}_i(t)). \tag{33}$$

V. PERFORMANCE EVALUATION

In this section, we evaluate the performance of the proposed SEMADT-RL framework for autonomous coordination in maritime SAR operations. The evaluation includes details of the adapted dataset and the simulation setup.

TABLE I: Simulation Parameters

Parameter	Value	Parameter	Value
$\mathcal{M}_x, \mathcal{M}_y$	1000×1000 m	Δt	2 s
T	200 steps	N_{node}	24
$v_{\max,i}$	3 m/s	$E_{\max,i}$	1000 J
$P_i(t)$	$[0, 2]$ W	$D_i(t)$	256 bytes
Replay Buffer	50,000	Batch Size	128
GNN Hidden Dim	128	Attention Heads	8
Learning Rate	5×10^{-4}	γ (discount)	0.98
Training Epochs	1000	β (VAE)	1

A. Simulation Setup

We simulate a 1000×1000 m maritime region with $N_{\text{node}} = 24$ IoMT agents over a $T = 200$-step horizon, using timestep $\Delta t = 2$ s. We adapt the UAVDT dataset [15] where frames are resized to 640×480 and passed through a semantic encoder to extract 64-dimensional latent vectors. The DT runs asynchronously, maintaining global forecasts via an extended Kalman filter. Simulations are implemented in Python using PyTorch, PyTorch Geometric, and Stable-Baselines3, on a system with an Intel i7-14700 CPU and NVIDIA T400 GPU. The simulation parameters are given in Tables I.

To evaluate the effectiveness of the proposed SEMADT-RL framework, we compare it against the following representative baselines:

- **Semantic-MADRL**, which omits DT forecasting to assess the impact of prediction-free semantic control;
- **Central-DRL**, a centralized policy with full global state access; and
- **Flat-MAPPO**, a decentralized MARL baseline without semantic encoding or DT support. These baselines highlight the importance of semantic compression and predictive coordination in maritime SAR.

B. Experimental Results

Fig. 2 shows the evolution of average utility $U_i(t)$, which captures task performance, energy use, and DT adherence. SEMADT-RL achieves consistent and high utility growth by aligning actions with forecasted survivor drift and energy availability, enabling faster convergence to effective policies. Semantic-MADRL shows early gains but flattens due to the lack of foresight, leading to reactive and less coordinated behavior. Central-DRL initially performs well with global knowledge but suffers from poor scalability. Flat-MAPPO struggles throughout due to unstructured sensing and energy inefficiency in the absence of semantic or predictive guidance.

Fig. 3 shows cumulative survivor detection over time. SEMADT-RL leads with early and sustained growth, driven by forecast-aware mobility that accelerates coverage and anomaly-based detection. Central-DRL starts strong due to full observability but stagnates without drift prediction. Semantic-MADRL lags due to reliance on local cues, missing survivors in delayed or occluded regions. Flat-MAPPO performs worst, with unstructured exploration leading to inefficient detection.

Fig. 4 shows average energy consumption per node. SEMADT-RL achieves the lowest and smoothest consumption

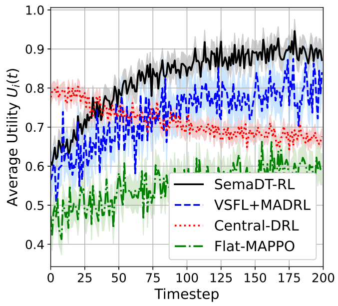

Fig. 2: Average utility vs timestep.

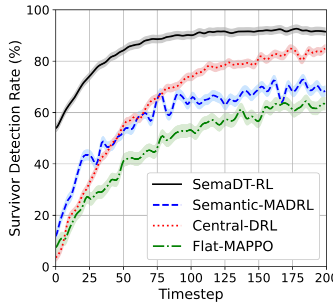

Fig. 3: Survivor detection rate.

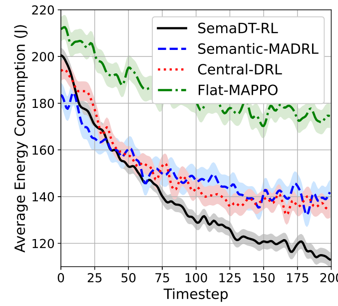

Fig. 4: Energy consumption vs timestep.

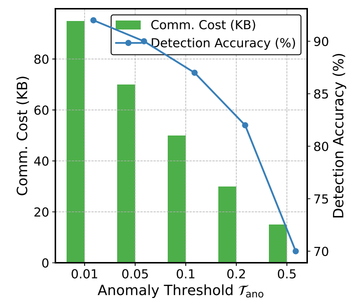

Fig. 5: Impact of $\mathcal{T}_{\rm ano}$.

profile, enabled by DT-guided control that avoids redundant movement and adapts power use based on forecasts. Semantic-MADRL consumes more due to reactive behavior without foresight. Central-DRL performs moderately well but fluctuates as centralized decisions fail to scale. Flat-MAPPO is the least efficient, with frequent, uncoordinated actions driving high energy usage.

Fig. 5 shows how the anomaly detection threshold $\mathcal{T}_{\rm ano}$ affects communication cost and survivor detection accuracy. As $\mathcal{T}_{\rm ano}$ increases, fewer frames exceed the anomaly score, reducing the number of transmitted semantic packets. The communication cost is computed as:

$$\text{Comm. Cost} = \sum_{i=1}^{N_{\rm node}} \sum_{t=0}^{T} \mathbb{1}\{\text{err}_i(t) > \mathcal{T}_{\rm ano}\} \times D_{\rm sem}, \quad (34)$$

where $D_{\rm sem}$ is the packet size (typically 256 bytes). However, higher thresholds risk missing subtle anomaly cues under challenging conditions, degrading detection performance. SEMADT-RL uses $\mathcal{T}_{\rm ano} = 0.1$ to reduce communication by over 50% while retaining survivor detection above 85%.

VI. Conclusions

In this paper, we presented SEMADT-RL, a three-layer framework that tightly integrates semantic compression, predictive digital twin modeling, and decentralized graph-based reinforcement learning to address the challenges of autonomous coordination in maritime IoMT-based search-and-rescue operations. The proposed approach enables IoMT nodes to efficiently perceive the environment, anticipate future dynamics, and adapt actions collaboratively under energy, mobility, semantic fidelity, and communication constraints. Through extensive simulations, we demonstrated that SEMADT-RL significantly outperforms baseline methods in terms of survivor detection rates, energy conservation, communication efficiency, and overall utility, particularly under dynamic and uncertain maritime conditions.

VII. Acknowledgments

This work was supported in part by the Canada Excellence Research Chair (CERC) Program CERC-2022-00109, in part by NSERC Discovery Grant, and in part through the Canada First Research Excellence Fund (CFREF) for the project Transforming Climate Action (TCA).

References

[1] T. Yang, Z. Jiang, R. Sun, N. Cheng, and H. Feng, "Maritime search and rescue based on group mobile computing for unmanned aerial vehicles and unmanned surface vehicles," *IEEE Trans. Ind. Inform.*, vol. 16, no. 12, pp. 7700–7708, 2020.

[2] W. Zhou, F. Zheng, G. Yin, Y. Pang, and J. Yi, "YOLOTrashCan: A deep learning marine debris detection network," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023.

[3] J. Fan, J. Zhou, X. Wang, and J. Wang, "A self-supervised transformer with feature fusion for sar image semantic segmentation in marine aquaculture monitoring," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–15, 2023.

[4] K. Lu, Q. Zhou, R. Li, Z. Zhao, X. Chen, J. Wu, and H. Zhang, "Rethinking modern communication from semantic coding to semantic communication," *IEEE Trans. Wirel. Commun.*, vol. 30, no. 1, pp. 158–164, 2023.

[5] Y. Wang, H. Han, Y. Feng, J. Zheng, and B. Zhang, "Semantic communication empowered 6g networks: Techniques, applications, and challenges," *IEEE Access*, vol. 13, pp. 28 293–28 314, 2025.

[6] C. Liu, C. Guo, Y. Yang, and N. Jiang, "Adaptable semantic compression and resource allocation for task-oriented communications," *IEEE Trans. Cogn. Commun. Netw.*, vol. 10, no. 3, pp. 769–782, 2024.

[7] Z. Zhang, Y. Liu, Z. Peng, M. Chen, D. Xu, and S. Cui, "Digital twin-assisted data-driven optimization for reliable edge caching in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 11, pp. 3306–3320, 2024.

[8] B. Hazarika, K. Singh, A. Paul, and T. Q. Duong, "Hybrid machine learning approach for resource allocation of digital twin in uav-aided internet-of-vehicles networks," *IEEE Trans Intell Veh*, vol. 9, no. 1, pp. 2923–2939, 2024.

[9] L. P. Qian, M. Li, P. Ye, Q. Wang, B. Lin, Y. Wu, and X. Yang, "Secrecy-driven energy minimization in federated-learning-assisted marine digital twin networks," *IEEE Internet Things J.*, vol. 11, no. 3, pp. 5155–5168, 2024.

[10] Q. Liu, Y. Guo, L. Deng, H. Liu, D. Li, H. Sun, and W. Huang, "Two-critic deep reinforcement learning for inverter-based volt-var control in active distribution networks," *IEEE Trans. Sustain. Energy*, vol. 15, no. 3, pp. 1768–1781, 2024.

[11] B. Hazarika, K. Singh, S. Biswas, and C.-P. Li, "Drl-based resource allocation for computation offloading in iov networks," *IEEE Trans. Ind. Inform.*, vol. 18, no. 11, pp. 8027–8038, 2022.

[12] B. Hazarika, K. Singh, S. Biswas, S. Mumtaz, and C.-P. Li, "Multi-agent drl-based task offloading in multiple ris-aided iov networks," *IEEE Trans Veh Technol.*, vol. 73, no. 1, pp. 1175–1190, 2024.

[13] B. Hazarika, P. Singh, K. Singh, S. L. Cotton, H. Shin, O. A. Dobre, and T. Q. Duong, "Generative ai-augmented graph reinforcement learning for adaptive uav swarm optimization," *IEEE Internet Things J.*, vol. 12, no. 8, pp. 9508–9524, 2025.

[14] H. Mazandarani, M. Shokrnezhad, and T. Taleb, "A semantic-aware multiple access scheme for distributed, dynamic 6g-based applications," in *Proc. IEEE WCNC 2024)*, 2024, pp. 1–6.

[15] D. Du, Y. Qi, H. Yu, Y. Yang, K. Duan, G. Li, W. Zhang, Q. Huang, and Q. Tian, "The unmanned aerial vehicle benchmark: Object detection and tracking," in *Proc. Springer ECCV 2018*, pp. 370–386.