Hybrid Framework for Multi-Robot Target Search in Unknown and Adversarial Environments

Tohid Kargar Tasooji Ramviyas Parasuraman

Abstract— **This paper presents a hybrid framework for multi-robot multi-target search in unknown, cluttered, and adversarial environments with limited sensing and intermittent communication. We employ a unified grid-based map encoding obstacles, coverage, target likelihood, and communication quality, together with a probabilistic adversarial model using layered Gaussian-like danger zones to capture the impact of varying adversary levels. On top of this representation, we develop an adaptive communication-aware, learning-based distributed MPC scheme that uses shared maps for cooperative trajectory planning under strong communication and learned inter-robot interaction models to predict neighbor states and maintain search efficiency when communication degrades, thereby reducing reliance on real-time data exchange. Extensive experiments on diverse maps with static and dynamic targets show that the proposed method reduces communication overhead and search time while improving prediction accuracy and overall resilience across different levels of adversaries.**

I. Introduction

Multi-robot systems, through collaboration, address the limitations inherent in individual robots, thereby improving resilience, scalability, and adaptability. This enhanced capability has attracted significant attention in the scientific and engineering communities. A central challenge in multi-robot systems is target search, which seeks to expedite the detection or identification of specific objects by leveraging the collective functionality of multiple autonomous agents. Recent field deployments demonstrate that multi-robot target search in unknown and communication-constrained environments is increasingly relevant to real-world applications. Representative examples include underground and tunnel exploration in GPS-denied settings, search-and-rescue operations in partially collapsed structures, and automated inspection of industrial facilities, where occlusions, clutter, and metallic infrastructure severely degrade wireless connectivity [1]–[3]. In safety-critical and adversarial scenarios, intentional or unintentional interference further introduces communication danger zones, requiring controllers that explicitly account for time-varying connectivity and system resilience.

Addressing the target search problem involves distinguishing between known and unknown environments. In known settings, the search domain is partitioned into segments to plan trajectories based on existing geographical data, although this approach may limit adaptability. In contrast, our work investigates target search in unknown environments, where predefined plans based on prior maps or fixed trajectories are less applicable, and online adaptation becomes essential. Multi-robot search systems are generally categorized into single-target and multitarget searches, with recent research increasingly leveraging intelligent algorithms and distributed systems inspired by cooperative behaviors observed in biological collectives. While these methods enhance fault tolerance and robustness, they continue to face significant challenges in multi-target searches, where additional coordination mechanisms are essential [4].

Environments for target search can be highly complex, and many studies simplify the problem by assuming obstacle-free spaces or simple obstacles [5]. However, real-world deployments face additional challenges beyond mere physical barriers, particularly in high-risk environments, where they encounter significant communication constraints. The Model Predictive Control (MPC) technique has been widely applied to this problem [6]–[8]. But in adversarial environments, factors such as jamming, interference, blockages, and system failures exacerbate communication constraints, resulting in delays, data loss, and transmission errors. These factors not only reduce search efficiency but also impose substantial obstacles to effective task execution [5].

In this work, we address multi-robot multi-target search in unknown, cluttered, and adversarial environments, where robots must locate and track targets under limited sensing and intermittent communication. The main contributions of this paper are as follows. *First,* we develop a hybrid learning-based distributed MPC (DMPC) framework that couples a learning-based Modular deep recurrent neural network (MoDeRNN) predictor with a communication-aware distributed MPC layer. Under strong communication, robots use shared maps for cooperative trajectory planning; under degraded links, a learned predictor estimates neighbor states from past inter-robot interactions. Its predictions are embedded in each robot's local distributed MPC problem to maintain search efficiency. This adaptive architecture enhances resilience, reduces communication overhead, preserves coordination as communication degrades. *Second,* we consider a unified grid-based map encoding obstacles, coverage, target likelihood augmented with a probabilistic adversarial model using layered Gaussian-like danger zones enabling adaptive, communication-aware decision making in dynamic environments. *Finally*, we validate the framework on diverse maps with static and dynamic targets and varying obstacle density, adversarial interference, and communication constraints, demonstrating robust performance, scalability with team size and environment complexity, and reduced search time and tracking error relative to baselines.

School of Computing, University of Georgia, Athens, GA 30602, USA. Corresponding author email: `ramviyas@uga.edu`

This research is supported by the Army Research Laboratory and is accomplished under DCIST Cooperative Agreement W911NF-17-2-0181.

II. Related Work

Design methodologies for multi-robot systems are often classified as automated or behavior-based [5]. Automated approaches based on deep RL and neural networks struggle in complex search tasks due to hardware limits, nonstationary environments, and high computational cost and still depend on rich global information, making them fragile under limited or unreliable communication [9]. In contrast, nature-inspired heuristics such as particle swarm optimization, group explosion strategies, and probabilistic finite-state machines offer lightweight, scalable exploration and multi-target tracking in cluttered environments [10].

Recent work has applied model-predictive control to multi-robot search and coverage. Zhao *et al.* [6] propose a hierarchical distributed MPC (HDMPC) scheme for multi-different-target search in unknown environments, where robots exchange model-based predictions over a grid map to reduce overlap and improve search efficiency. Yu *et al.* [11] extend this idea to an adaptive communication-conditions HDMPC (ACC–HDMPC) framework that augments HDMPC with communication-quality maps and event-triggered information exchange, but still relies on deterministic, model-based prediction of neighbor states. Riehl *et al.* [7] study cooperative UAV search over dynamic graphs, in which aerial robots perform receding-horizon planning over probabilistic target maps and time-varying interaction graphs under motion and communication constraints. Yu *et al.* [12] develop a multi-robot waypoint-planning system that combines local area decomposition with a global waypoint planner to jointly improve coverage rate and search efficiency in unknown environments.

In parallel, a large body of work employs learning and heuristic optimization for multi-robot target search. Ebert *et al.* [13] propose a hybrid PSO-based strategy for continuous-cue target search and decision awareness in robot collectives. Zhu *et al.* [14] develop a deep reinforcement learning framework for hierarchical role selection and adaptive area search, while Hou *et al.* [15] apply multi-agent reinforcement learning to large-scale UAV swarm cooperative target search. Wolek *et al.* [16] perform cooperative mapping and target search over unknown occupancy graphs using mutual information to select actions that maximize information gain while reducing uncertainty about both the environment and target locations. Yu *et al.* [17] employ an OR-LSTM–based trajectory predictor together with a bioinspired hunting strategy for cooperative target engagement. Peng *et al.* [18] design graph neural network policies for decentralized task allocation and coordination in multi-robot systems, while Zhou *et al.* [19] use GNN-based representations for cooperative tracking and target assignment. Ni *et al.* [20] propose bioinspired neural controllers for real-time cooperative hunting with dynamic target motion.

Hudson *et al.* present Team CSIRO Data61's heterogeneous ground–air platform, which combines decentralized SLAM, frontier-based exploration, and disruption-tolerant map sharing to enable distributed coordination via opportunistic communication [21]. Dang *et al.* develop local–global planning frameworks based on RRT and graph search to support scalable lookahead planning in complex subterranean (Sub-T) networks [22]. These SubT-style approaches rely primarily on model-based coordination and prediction, with limited use of learning for inter-robot interaction under degraded communication. In contrast, our hybrid framework embeds a learning-based MoDeRNN predictor within a communication-aware distributed MPC layer, enabling robust coordination under packet drops and delays.

Taken together, the above MPC-based approaches for multi-robot search and tracking [6]–[8], [11], [12] and the learning- and heuristic-based methods [13]–[20] address complementary aspects of the problem. However, the MPC-based approaches remain purely model-driven: they assume that neighboring states or shared maps are sufficiently available over the network and do not use learned predictors to reconstruct missing information or to explicitly reason about adversarial or highly variable communication conditions. Conversely, the learning-based methods leverage data-driven models for prediction, task allocation, and policy design, yet they are not formulated as communication-aware distributed MPC schemes and typically do not offer explicit constraint handling or guarantees on closed-loop coordination under degraded communication. In contrast, our proposed framework tightly couples a learning-based MoDeRNN predictor with a communication-aware distributed MPC layer: the predictor is trained on historical inter-robot interactions and communication metrics to estimate neighbor states under packet drops and delays, and these learned predictions are embedded directly into each robot's local DMPC optimization together with communication-quality-dependent costs and constraints. This yields a learning-enhanced DMPC architecture that (i) maintains feasibility and coordination when communication degrades, (ii) adapts search behavior to heterogeneous and adversarial link conditions, and (iii) unifies model-based MPC with data-driven interaction prediction for multi-robot target search in unknown and adversarial environments.

III. Resilient Multi-Robot Multi-target Search

We address the problem of efficient and resilient multi-robot, multi-target search in unknown and adversarial environments, where limited sensing, environmental obstacles, communication failures, and adversarial interference often degrade performance. We propose an *Adaptive Communication-Aware, Learning-Based Distributed Model Predictive Control* (ACC-LBDMPC) framework (Fig. 1).

Here, the **cooperative search layer** constructs and maintains obstacle, exploration, target, and communication maps, enabling robots to identify unsearched regions, targets, and communication quality. **The adaptive communication–prediction layer** manages map and state exchange, leverages a learning-based predictor trained on historical inter-robot interactions to recover missing or delayed data, and incorporates a probabilistic danger-zone model with Gaussian-like radial profiles to quantify adversarial severity and balance cooperation with autonomy.

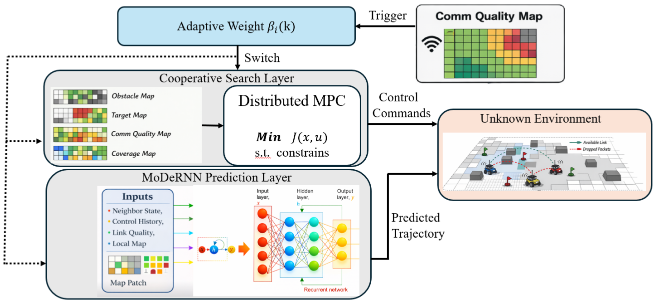

Fig. 1: Architectural overview of the proposed approach for resilient multi-robot multi-target search in unknown and adversarial environments.

Fig. 2: Multi-robot target search environment under communication constraints (left). Grid map-based modeling of robotic search operations and target detection (right).

The motion control layer synthesizes safe and efficient trajectories using real-time map and prediction outputs, ensuring adaptability to environmental changes and communication disruptions. Collectively, this layered framework improves search efficiency, reduces reliance on continuous communication, and enhances system resilience, advancing the state of the art in cooperative multi-robot search.

A. Modeling the Search Environment

Fig. 2 illustrates the search environment, modeled as a rectangular region of dimensions $x \times y$, discretized into an $M \times N$ grid with cell size d. The color gradient (blue) represents varying levels of communication interference, with darker shades indicating more severe degradation in communication quality. Obstacles that restrict robot mobility are marked in red, while red circular symbols denote potential target zones. The locations of obstacles and targets are not known a priori, increasing the complexity of the mission. This discrete grid representation, coupled with heterogeneous communication conditions and unknown terrain features, highlights the fundamental challenges of autonomous multi-robot search in uncertain environments.

Remark 1: The environment is represented by a uniform grid whose cells store occupancy, coverage, target-belief, and communication-quality information, while the robot dynamics and ACC–LBDMPC controller operate in continuous state space. Thus, the grid resolution mainly trades off spatial fidelity against computational and communication cost: finer grids provide more detailed maps but increase update and sharing overhead. Moderate changes in cell size lead to similar qualitative search behavior, with complexity growing approximately with the number of cells.

B. Modeling the Multi-Robot System

Each robot explores unknown regions using exteroceptive sensors and communication devices. The robots move in a two-dimensional plane, and their motion is described by:

$$\dot{x}_i(t) = v_i(t)\cos\phi_i(t), \quad \dot{y}_i(t) = v_i(t)\sin\phi_i(t), \tag{1}$$

where $v_i(t)$ is the linear velocity of the robot R_i at time t with bounds $\|v_i(t)\| \le v_{\max}$, $\phi_i(t)$ is the angle measured from the x-axis, and $v_{\max}$ denotes the maximum speed. If the continuous spatial position of a robot is denoted by $P_r\big(x_r(t), y_r(t)\big)$ and its corresponding grid position is given by $P_g\big(x_g(t), y_g(t)\big)$, then grid coordinates are computed as: $M = \left\lfloor \frac{X}{d} \right\rfloor, N = \left\lfloor \frac{Y}{d} \right\rfloor$, $x_g(t) = \left\lfloor \frac{x_r(t)}{d} \right\rfloor, y_g(t) = \left\lfloor \frac{y_r(t)}{d} \right\rfloor$, where X and Y are the overall dimensions of the environment. We assume that X, Y, d, and thus the resulting grid dimensions M and N, are known to all robots at the start of the mission (e.g., provided as part of a shared mission specification). Robots move in eight distinct directions, $A_1, \ldots, A_8$, uniformly distributed over $[0, 2\pi]$. Each robot is equipped with an exteroceptive sensor with an isotropic circular sensing footprint of radius d_{tec} (Fig. 2). This circular footprint is enclosed within a square region D_{tec} subdivided into an $L \times L$ grid aligned with the global map. The sensing model is discretized by marking as detectable those cells whose centers lie inside the sensing disc and non-detectable those whose centers lie outside.

The detection state over this local sensing area is captured by the binary matrix $C_d = [\theta_{ij}]_{L\times L}$,, where a value of 1 indicates that the grid cell at (i, j) falls within the detectable region $D_{\text{tec}}(t)$ and 0 indicates otherwise (cells in the non-detectable region $D_{\text{ntec}}(t)$), with the union of these regions equaling D_{tec}:

$$\theta_{ij} = \begin{cases} 1, & \text{if the center of cell } (i,j) \in D_{\text{tec}}(t), \\ 0, & \text{if the center of cell } (i,j) \in D_{\text{ntec}}(t). \end{cases}$$

In this work, we adopt an ideal binary sensor model, treating cells with $\theta_{ij} = 1$ as fully observed and cells with $\theta_{ij} = 0$ as unobserved, although the same structure can accommodate probabilistic detection by allowing $\theta_{ij} \in [0, 1]$.

The parameter L is determined by $L = 2\left\lfloor \frac{d_{\text{tec}}}{d} - \frac{1}{2} \right\rfloor + 1$, so that the $L \times L$ patch fully covers the sensing disc for grid-cell size d. For each robot, an L-order submatrix $C^i_{2LE}(t) = [\beta_{ij}(t)]_{L\times L}$ is extracted from its full coverage matrix $C^i_2(t) = [\beta_{ij}(t)]_{M\times N}$. This submatrix is updated using the Hadamard product with the detection matrix:

$$C^{i,\text{new}}_{2LE}(t) = C^i_{2LE}(t) \circ C_d = \big[\beta_{ij}(t)\,\theta_{ij}\big]_{L\times L}. \tag{2}$$

C. Modeling the Multi-Target Scenario

Targets are statically positioned and randomly distributed within the search area, forming the set $T = \{T_1, T_2, \ldots, T_n\}$. They are represented as red regions in Fig. 2 to indicate detectability. Each target T_ℓ is modeled as a point in the two-dimensional plane with coordinates (x_ℓ, y_ℓ). A target is considered detected when a robot enters a circular neighborhood of radius d_{obj} around it, i.e.,

$$\big\|x_i(t) - x_\ell\big\| < d_{\text{obj}}$$

This distance-based detection rule is consistent with the circular sensing footprint used in the grid-based sensor model, where targets located within the sensing region correspond to cells marked as detectable in the local $L \times L$

patch. Furthermore, we introduce the notion of *target affinity* to quantify the suitability of assigning a particular robot to a given target. This affinity is a task-allocation score (e.g., based on distance, expected travel time, or local map information). Note that all robots share the same circular detection radius d_{obj}, and a target T_ℓ is detected whenever $\|x_i(t) - x_\ell\| < d_{\text{obj}}$, consistent with the circular footprint discretized in the local $L \times L$ patch.

D. Algorithmic Implementation

Our framework is built using a hierarchical architecture by incorporating: (i) an explicit probabilistic adversary model, (ii) a learning-based residual prediction module, and (iii) a dynamic communication-aware switching strategy. At runtime, robots monitor communication quality (e.g., packet-loss rate and bit-error rate). When communication is reliable, they operate in a *top-level cooperative search layer*, executing distributed MPC while sharing grid maps of obstacles, coverage, targets, and link quality. Under degraded communication, they switch to a *bottom-level local search layer*, relying solely on a modular deep recurrent neural network (MoDeRNN) [23] predictions for decision-making.

1) Probabilistic Adversary Model: We model communication interference using a probabilistic framework with layered circular danger zones. Each physical adversarial zone, denoted by $\mathcal{A}_z$ for $z = 1, \ldots, M$, is defined in the two-dimensional space $\mathbb{R}^2$ as the region where the adversarial probability exceeds a threshold ϵ, i.e.,

$$\mathcal{A}_z(t) = \{\mathbf{x} \in \mathbb{R}^2 \mid P_z(\mathbf{x}, t) > \epsilon\}, \tag{3}$$

where $\epsilon \in (0, 1)$ specifies the boundary between safe and unsafe regions. The adversarial probability $P_z(\mathbf{x}, t)$ is modeled as a radially symmetric Gaussian-like distribution:

$$P_z(\mathbf{x}, t) = \alpha_z(t) \exp\Big(-\tfrac{\|\mathbf{x}-\mathbf{c}_z(t)\|^2}{2\sigma_z^2(t)}\Big), \tag{4}$$

where $\alpha_z(t)$ denotes the peak probability representing the maximum communication disruption in the zth zone, $\mathbf{c}_z(t) \in \mathbb{R}^2$ is the zone center, and $\sigma_z(t)$ characterizes the spatial spread, governing the decay of interference with distance from the center.To connect the abstract adversarial field $P_z(x, t)$ to packet loss, we interpret $P_z(x, t)$ as a local measure of communication disruption and define a spatially varying communication-quality function $q(x, t) \in [0, 1]$ via a monotone mapping $q(x, t) = q_{\min} + \big(q_{\max} - q_{\min}\big)\big(1 - P_z(x, t)\big)$, with $0 < q_{\min} \le q_{\max} \le 1$. Thus, points close to the center of a danger zone (high P_z) are assigned lower communication quality, while points in safe regions (low P_z) have $q(x, t) \approx q_{\max}$. In our communication model, $q(x, t)$ is interpreted as the packet-delivery probability for messages transmitted at position x, and the communication-quality metric is computed as the empirical packet-delivery ratio over a sliding time window.

2) Local MPC Formulation: In the *cooperative search layer*, each robot generates a short-term trajectory predictions using a model predictive control (MPC) scheme that balances exploration efficiency, motion smoothness, target detection, and cooperative awareness. At time step k, robot i optimizes its control inputs $\{u_i(\tau|k)\}_{\tau=k}^{k+T-1}$ over a prediction horizon of length T to minimize a cumulative cost

$$J_i(k) = \sum_{\tau=k}^{k+T-1} \big(\omega_1 J_C^i(\tau) + \omega_2 J_T^i(\tau) + \omega_3 J_B^i(\tau) + \omega_4 J_D^i(\tau)\big), \tag{5}$$

where the stage-cost components are defined as follows.

Let $x_i(\tau|k)$ denote the predicted state of robot i at time τ and $u_i(\tau|k)$ the corresponding control input. Let $g(x_i(\tau|k))$ denote the grid cell containing $x_i(\tau|k)$, and let $C(g) \in [0, 1]$ be the coverage value of that cell (with $C = 0$ for unexplored cells and $C \approx 1$ for fully explored cells). The coverage term

$$J_C^i(\tau) = 1 - C\big(g(x_i(\tau|k))\big) \tag{6}$$

penalizes revisiting already explored areas and encourages exploration of unvisited regions.

For smooth motion, we penalize changes in control inputs:

$$J_T^i(\tau) = \big\|u_i(\tau|k) - u_i(\tau-1|k)\big\|^2. \tag{7}$$

Obstacle avoidance is enforced via a smooth barrier term

$$J_B^i(\tau) = \phi_B\big(d_{\text{obs}}(x_i(\tau|k))\big), \tag{8}$$

where $d_{\text{obs}}(x)$ is the distance from x to the closest obstacle and $\phi_B(\cdot)$ is a decreasing function (e.g., $\phi_B(d) = 1/(d^2 + \varepsilon)$, $\varepsilon > 0$) that grows as the robot approaches obstacles. Finally, the target term

$$J_D^i(\tau) = \min_{\ell \in \mathcal{T}} \big\|x_i(\tau|k) - \hat{x}_\ell^{\text{tar}}(\tau)\big\|^2 \tag{9}$$

penalizes the distance to the nearest believed target location $\hat{x}_\ell^{\text{tar}}(\tau)$, encouraging trajectories that move toward and detect targets. The weights $\omega_1, \ldots, \omega_4$ balance coverage, smoothness, safety, and target-seeking behavior.

The predicted trajectories are subject to the discrete-time dynamics

$$x_i(\tau+1|k) = f\big(x_i(\tau|k), u_i(\tau|k)\big), \quad \tau = k, \ldots, k+T-1,$$

with $x_i(k|k) = x_i(k)$, and to state–input and interaction constraints

$$x_i(\tau|k) \in \mathcal{X}, \quad u_i(\tau|k) \in \mathcal{U}, \quad \forall \tau, \tag{10}$$

$$\big\|x_i(\tau|k) - \hat{x}_j(\tau|k)\big\| \ge d_{\min}, \quad \forall j \in \mathcal{N}_i(k),\ \forall \tau, \tag{11}$$

$$\big\|x_i(\tau|k) - \hat{x}_j(\tau|k)\big\| \le R_c, \quad \forall j \in \mathcal{N}_i(k),\ \forall \tau, \tag{12}$$

which ensure motion feasibility, collision avoidance, and maintenance of communication links with neighboring agents $j \in \mathcal{N}_i(k)$. Here, $\hat{x}_j(\tau|k)$ denotes the predicted states of neighbors, obtained from the communication-aware, learning-based prediction mechanism described next.

3) Learning-Based Neighbor-State Prediction: When communication is reliable, each robot receives predicted trajectories $x_j^{\text{MPC}}(\tau|k)$ from its neighbors $j \in \mathcal{N}_i(k)$ (e.g., the solutions of their local MPC problems) and uses these in (11)–(12). Under degraded communication, however, packets may be delayed or dropped, and robot i must estimate the future states of its neighbors. To this end, we employ a learning-based model to predict neighbor trajectories.

For a neighbor $j \in \mathcal{N}_i(k)$, the nominal prediction $x_j^{\mathrm{nom}}(\tau+1|k)$ (e.g., extrapolated from the last received state or trajectory) is augmented by a learned residual produced by a neural network model (MoDeRNN [23]):

$$\hat{x}_j^{\mathrm{RNN}}(\tau+1|k) = x_j^{\mathrm{nom}}(\tau+1|k) + \Delta x_{ij}(\tau|k), \tag{13}$$

$$\Delta x_{ij}(\tau|k) = \mathrm{MoDeRNN}_{ij}\big(z_{ij}(\tau|k), \mathcal{H}_{ij}(\tau|k)\big), \tag{14}$$

where $z_{ij}(\tau|k)$ stacks the inputs to the network (e.g., the last received neighbor state $x_j(k_{\text{last}})$ and control, robot i's own state $x_i(\tau|k)$, and communication-quality features such as packet-loss indicators), and $\mathcal{H}_{ij}(\tau|k) \in \mathbb{R}^h$ is the hidden state of a two-layer gated recurrent unit (GRU) network. The MoDeRNN models are updated online from trajectories collected under varying communication and adversarial conditions using truncated backpropagation through time over a sliding window of length W.

To adapt the relative trust in the MPC- and RNN-based neighbor predictions to the current link quality and the reliability of the learned model, we combine a communication-health metric with an on-line prediction error metric.

First, for each robot i we define a communication-health metric $\rho_i(k) \in [0,1]$ as the empirical packet delivery ratio (PDR) over a sliding window of W_c communication slots:

$$\rho_i(k) = \frac{1}{W_c} \sum_{\ell=k-W_c+1}^{k} \mathbb{I}\big\{\text{packet at time } \ell \text{ is received}\big\},$$

where $\mathbb{I}\{\cdot\}$ denotes the indicator function. A second quantity captures the online accuracy of the MoDeRNN predictions.

$$\text{Let } e_i(k) = \frac{1}{|\mathcal{N}_i(k)|} \sum_{j \in \mathcal{N}_i(k)} \left\| \hat{x}_j^{\mathrm{RNN}}(k|k-1) - x_j(k) \right\| \tag{15}$$

be the average one-step prediction error over neighbors when new states $x_j(k)$ are received. We normalize this error using a design upper bound $e_{\max} > 0$:

$$\tilde{e}_i(k) = \mathrm{sat}_{[0,1]}\left(\frac{e_i(k)}{e_{\max}}\right), \quad \eta_i(k) = 1 - \tilde{e}_i(k), \tag{16}$$

so that $\eta_i(k) \in [0,1]$ is a prediction-reliability score (close to 1 when the RNN is accurate; close to 0 when its error is large). We then form a combined trust score

$$s_i(k) = \lambda\, \rho_i(k) + (1-\lambda)\, \eta_i(k), \qquad \lambda \in [0,1], \tag{17}$$

and map it to a mixing weight $\beta_i(k) \in [0,1]$ via

$$\beta_i(k) = \mathrm{sat}_{[0,1]}\left(\frac{s_i(k) - s_{\min}}{s_{\max} - s_{\min}}\right), \tag{18}$$

with $0 \le s_{\min} < s_{\max} \le 1$ design thresholds. Thus, $\beta_i(k)$ is close to 1 when communication is good and/or the RNN predictions are unreliable (high PDR or large error), in which case the MPC-based neighbor prediction is trusted, and close to 0 when communication is poor, but the RNN predictions remain accurate.

The hybrid neighbor prediction used in the DMPC constraints (11)–(12) is then

$$\begin{aligned} \hat{x}_j(\tau+1|k) = {} & \beta_i(k)\, x_j^{\mathrm{MPC}}(\tau+1|k) \\ & + \big(1-\beta_i(k)\big)\, \hat{x}_j^{\mathrm{RNN}}(\tau+1|k), \end{aligned} \tag{19}$$

for all neighbors $j \in \mathcal{N}_i(k)$, where $x_j^{\mathrm{MPC}}(\tau+1|k)$ is the last communicated MPC prediction from robot j and $\hat{x}_j^{\mathrm{RNN}}$ is given by (13). In this way, the ACC–LBDMPC layer exploits communication-aware, learning-based neighbor predictions when they are reliable, while automatically reverting to a purely model-based DMPC behavior when communication is good or the RNN error becomes large.

Remark 2: MoDeRNN is trained online in a self-supervised fashion using trajectories generated during multi-robot operation. Each robot logs tuples $(x_i(k), x_j(k), u_j(k), q_{ij}(k))$ containing local states, last received neighbor information, and communication-quality features; when updated neighbor states arrive, the resulting prediction errors serve as training signals, and the network is updated via truncated backpropagation through time over a sliding window, using data collected across randomized maps, initial conditions, and communication regimes. When the predictor becomes unreliable or operates out-of-distribution, the gating mechanism drives $\beta_i(k)$ toward one, reverting to purely model-based DMPC behavior.

E. Theoretical Analysis

We summarize conditions under which the proposed ACC–LBDMPC scheme guarantees recursive feasibility, ISS-type stability, and mission-level convergence for multi-robot target search with hybrid framework.

Assumptions. (A1) **Bounded residual:** the closed-loop dynamics satisfy $x_i(k{+}1) = f(x_i(k), u_i(k)) + \Delta f_i(k)$ with a known bound $\|\Delta f_i(k)\| \le \delta$, enforced by the communication-/error-gated blending between the nominal model and MoDeRNN. (A2) **Regularity:** f is locally Lipschitz on $\mathcal{X} \times \mathcal{U}$, and the linearization on the operating region is stabilizable. (A3) **Robust terminal ingredients:** there exist a compact terminal set $\mathcal{X}_f$, a terminal law $u = k_f(x)$, and a Lyapunov-type terminal cost V_f such that $\mathcal{X}_f$ is robustly positively invariant for all $\|w\| \le \delta$ and V_f decreases along $f(x, k_f(x)) + w$ up to a term $\alpha(\delta)$ with $\alpha(\delta) \to 0$ as $\delta \to 0$. (A4) **Hybrid dwell time:** mode switching (between cooperative and local modes) obeys a minimum dwell time $T_d \ge 1$ chosen so that the net Lyapunov change over each mode interval is negative. (A5) **Initial feasibility:** the MPC problem is feasible at $k = 0$ for all robots. (A6) **Coverage cost:** the coverage term in the stage cost is strictly positive whenever unvisited cells lie in the reachable set over the horizon, and cooperative periods synchronize maps among connected robots.

Theorem 1: Under Assumptions (A1)–(A6), consider the finite-horizon MPC cost

$$J(x_i(k), \mathbf{u}_i) = \sum_{j=0}^{T-1} \ell\big(x_i(j), u_i(j)\big) + V_f\big(x_i(T)\big), \tag{20}$$

implemented with robust constraints or the terminal set of (A3) and switching satisfying (A4).

Then, for each robot R_i:

(i) Recursive feasibility: if the MPC problem is feasible at time k, it remains feasible for all $k \ge 0$.

(ii) ISS-type performance: there exist $\gamma > 0$ and a class-$\mathcal{K}$ function β such that the optimal value function $V(x_i(k))$ satisfies

$$V\big(x_i(k+1)\big) - V\big(x_i(k)\big) \leq -\gamma\|x_i(k)\|^2 + \beta(\delta), \quad (21)$$

so the closed loop is input-to-state stable with respect to the bounded residual $\Delta f_i(k)$, with an ultimate bound that scales with δ.

(iii) Mission convergence: under (A6), repeated cooperative intervals and the coverage cost yield a monotone decrease of the global uncovered area, so all reachable targets are detected in finite time.

Proof: Let $V(x_i(k))$ denote the optimal cost-to-go at time k. At time $k+1$, consider the standard shifted candidate

$$\tilde{\mathbf{u}}_i = \{u_i^*(1), \ldots, u_i^*(T-1), k_f(x_i^*(T))\}, \quad (22)$$

where $\mathbf{u}_i^*$ and $x_i^*(\cdot)$ are optimal at time k. By (A1)–(A3), the resulting terminal state

$$x_i^+ = f\big(x_i^*(T), k_f(x_i^*(T))\big) + \Delta f_i(T) \quad (23)$$

remains in $\mathcal{X}_f$ for all $\|\Delta f_i(T)\| \leq \delta$, so $\tilde{\mathbf{u}}_i$ is feasible; the dwell time (A4) prevents switching from violating this argument, which yields recursive feasibility.

Comparing the cost of $\tilde{\mathbf{u}}_i$ with $V(x_i(k))$ and using (A3) gives $V\big(x_i(k+1)\big) - V\big(x_i(k)\big) \leq -\ell\big(x_i(k), u_i(k)\big) + \alpha(\delta)$. Since ℓ is positive definite, there exist $\gamma > 0$ and a class-$\mathcal{K}$ function β such that

$$\ell(x, u) \geq \gamma\|x\|^2, \quad \alpha(\delta) \leq \beta(\delta), \quad (24)$$

which yields the ISS inequality. Finally, (A6) and the finiteness of the grid imply that synchronized cooperative intervals strictly reduce the cardinality of the unvisited set until all reachable targets are covered, proving mission convergence. ■

Remark 3: Theorem 1 guarantees recursive feasibility and input-to-state stability of the hybrid ACC–LBDMPC framework for any bound δ that upper-bounds the residual $\Delta f_i(k)$, which captures both model mismatch and learning-induced prediction error. In practice, δ is estimated in a data-driven manner: each robot monitors the one-step residuals $\|\Delta f_i(k)\| = \|x_i(k+1) - f(x_i(k), u_i(k))\|$ over a sliding window and sets δ as a safety-margined upper bound (e.g., a scaled high percentile across robots and scenarios), yielding an empirical disturbance bound consistent with Assumption (A1). This bound is then used to construct a robust terminal set and control law via local linearization and LQR, with $u = k_f(x) = Kx$, $V_f(x) = x^\top Px$, and $\mathcal{X}_f = \{x : x^\top Px \leq r_f\}$, where P solves the discrete-time Riccati equation and r_f is selected so that $\mathcal{X}_f$ is invariant for all $\|\Delta f_i(k)\| \leq \delta$. Under this design, the ISS bound in Theorem 1 implies graceful performance degradation as δ increases (e.g., under poorer or out-of-distribution learning), while feasibility and mission convergence are preserved as long as the estimated residual remains bounded.

Remark 4: The MPC cost in (5) is used as a tractable surrogate for the mission-level search objective. The target-attraction term promotes faster convergence toward detected target locations, serving as an alternative search-efficiency indicator alongside detection time, cumulative detection time, and final detection ratio. The coverage term encourages exploration, while the smoothness and obstacle avoidance terms ensure feasible and safe motion. The weights are fixed across all experiments and selected once through validation. Theorem 1 applies to static or reachable targets. For dynamic targets, we consider small deviations around the nominal predicted target position to capture target-location uncertainty; hence, robots are driven toward a local uncertainty region rather than a fixed point. Since coverage alone does not guarantee detection of moving targets, dynamic-target results are interpreted as empirical robustness evaluations. During prolonged communication loss, stale prediction-error estimates reduce confidence in MoDeRNN, and the controller gradually reverts to the model-based prediction.

IV. Experiments

We conducted experiments in the Robotarium simulation–hardware platform [24], which provides remote access to a multi-robot testbed. A team of 5 to 12 robots was deployed in a rectangular $3.2\,\mathrm{m} \times 2.0\,\mathrm{m}$ workspace, initialized from fixed poses specified in our implementation, and operated for 140 time steps per trial. In the proposed framework, each robot solves a local sampling-based MPC problem (N_{horizon}=5, N_{samples}=50) augmented with a GRU-based residual predictor trained from prior inter-robot interactions.

A grid-based map encodes obstacles, exploration coverage, target occupancy, and a communication-quality layer. This layer governs adaptive mode switching: under robust communication, robots share maps to generate coordinated search trajectories, whereas degraded communication (below a threshold) triggers prediction-based autonomy relying on learned forecasts. Adversarial effects are modeled as layered circular danger zones with radially decaying communication quality, which degrade information exchange and force greater reliance on local predictions.

To evaluate performance, we benchmark the proposed framework against two distributed baselines implemented under identical sensing, solver, and initialization conditions: (i) a hierarchical communication-aware distributed MPC controller inspired by ACC-HDMPC [11], and (ii) an OR-LSTM–based cooperative hunting method with trajectory prediction [17]. Comparative analysis is conducted using the different metrics by our implementation, namely *tracking cost*, *coverage cost*, *heading cost*, *collision cost*, and the resulting *total cost*, together with the *average communication quality* and the recorded *2D trajectories* of the robots. Table I provides a comprehensive summary of the results.

A. Static Targets and Adversarial Zone

Figure 3 reports the comparative performance of the proposed Hybrid framework, ACC–HDMPC, and OR–LSTM under an adversarial region of radius $R = 0.8$. In this setting, the communication-degraded zone occupies a substantial portion of the workspace, severely impairing inter-robot information exchange and challenging methods that rely

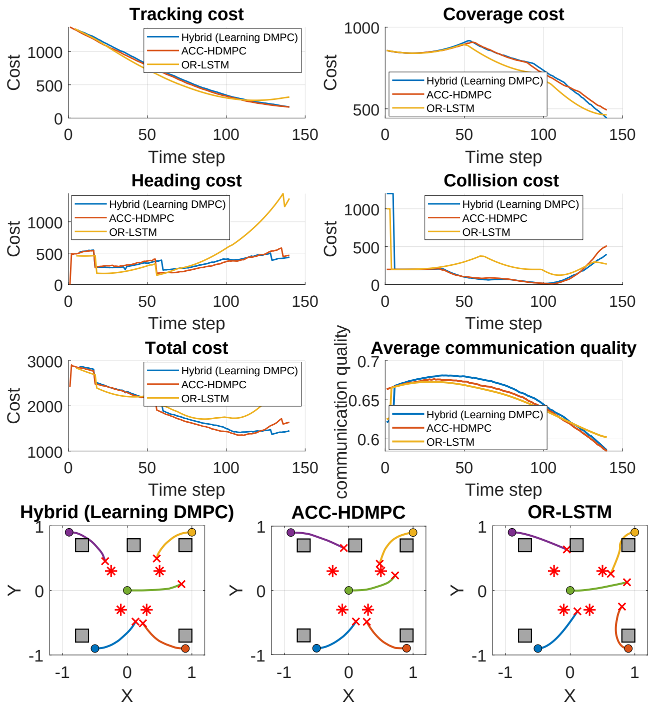

Fig. 3: Performance results of multi-robot target search in the presence of static targets and adversarial zones (R = 0.8).

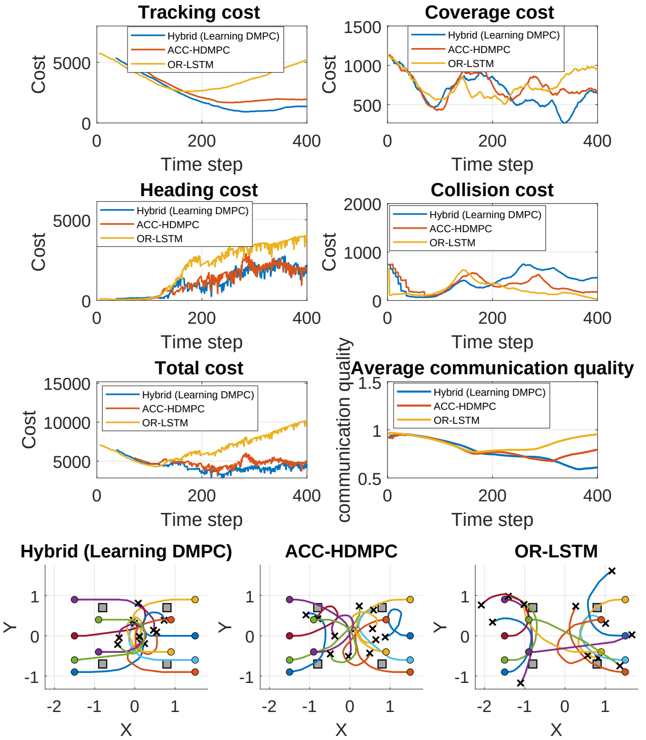

Fig. 4: Performance results in dynamic adversarial environments with 12 robots.

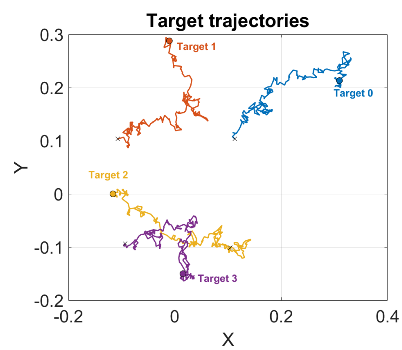

Fig. 5: Dynamic adversarial environments with 4 moving targets.

on frequent, reliable coordination. Under these conditions, ACC–HDMPC and OR–LSTM exhibit increased tracking, coverage, and collision costs, reflecting the impact of stale or missing information on global coordination. In contrast, the Hybrid framework allows each robot to autonomously switch from communication-enabled cooperation to a local prediction mode once link quality deteriorates. In prediction mode, the MoDeRNN-based residual predictor supplies short-horizon motion forecasts and residual corrections that inform a local sampling-based MPC. This per-robot RNN+MPC pipeline preserves collision avoidance, reduces redundant exploration, and maintains lower aggregate cost in the presence of a large adversarial zone.

Overall, the results in Fig. 3 indicate that the proposed Hybrid framework achieves an effective trade-off between tracking accuracy, coverage efficiency, and safety under severe communication degradation, outperforming ACC–HDMPC and OR–LSTM in this challenging scenario.

B. Dynamic Targets under Adversarial Communication

Fig. 4 reports the performance of the proposed Hybrid controller, ACC–HDMPC, and OR–LSTM for a 12-robot team pursuing four dynamic targets (movements of the targets are shown in Fig. 5) in the presence of an adversarial communication-degraded region. As the team size increases, communication becomes more congested, and coordination errors are amplified, causing ACC–HDMPC and OR–LSTM to exhibit higher tracking and total costs, along with more pronounced spikes in heading and collision costs.

In the Hybrid case, the collision cost is initially lower because robots operate in a communication-rich regime and the controller relies predominantly on model-based MPC predictions of neighbor motion. As robots enter the communication-degraded zone and link quality drops, the controller gradually shifts weight toward prediction-based coordination using the MoDeRNN residual model, which introduces modest uncertainty and leads to a slight increase in collision cost, while keeping it bounded. Overall, the Hybrid framework maintains lower tracking and total costs while keeping heading deviations and collision penalties controlled, indicating that the proposed learning-augmented MPC scales favorably to larger teams with dynamic targets and delivers robust, resilient performance compared to the ACC–HDMPC and OR–LSTM methods in adversarial environments.

C. Real-World Robot Demonstration

We demonstrate the proposed hybrid learning-based MPC framework on the Robotarium hardware, where a team of five robots executed cooperative search tasks in environments with obstacles and adversarial regions under limited actuator commands (see Fig. 6). The experiment confirmed the simulation results in enabling resilient and efficient multi-target search. More details of this demonstration are provided in

TABLE I: Comparison of various performance metrics for multi-robot target search across four scenarios. Our Hybrid approach is compared with ACC-HDMPC [11] and OR-LSTM [17]. Best results are highlighted in **bold face.**

Metric	Static targets, adversarial (R=0.4), N=5			Static targets, adversarial (R=0.8), N=5			Dynamic targets, adversarial (R=0.4), N=5			Dynamic targets, adversarial (R=0.4), N=12		
	Hybrid	ACC-HDMPC	OR-LSTM	Hybrid	ACC-HDMPC	OR-LSTM	Hybrid	ACC-HDMPC	OR-LSTM	Hybrid	ACC-HDMPC	OR-LSTM
Mean total cost	**1807**	1866	2197	**1890**	1911	2086	**3439**	3447	3580	**4352**	4762	6688
Peak total cost (last step)	**1280**	1389	3000	**1436**	1619	2506	**2082**	2882	2803	**4932**	6466	10189
Mean tracking error (m)	**626**	651	714	**636**	641	613	**2087**	2158	2225	**2194**	2422	3752
Mean coverage cost	**781**	766	780	**773**	773	738	**1042**	955	943	**666**	734	749
Mean collision count (run)	**0.6**	1.1	1.5	**0.7**	0.9	1.4	**2.2**	3.8	4.5	**5.0**	7.2	8.5
Avg. communication quality	0.78	0.80	**0.82**	0.75	0.77	**0.78**	0.70	0.68	**0.72**	**0.68**	0.65	0.66
Time-to-convergence (steps)	**120**	130	135	**125**	130	140	**220**	240	260	**300**	340	380

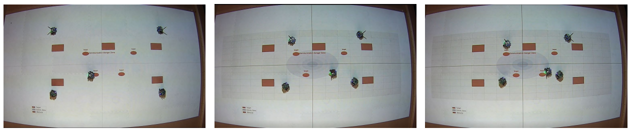

Fig. 6: Illustration of experiments conducted in the Robotarium hardware platform, showing the robots' trajectories over time, finally closing in on the targets.

the attached video.

V. Conclusion

This paper introduced a hybrid communication-aware, learning-augmented distributed MPC framework for resilient multi-robot multi-target search in unknown and adversarial environments. By integrating a grid-based map representation with adaptive mode switching between communication-enabled coordination and prediction-driven autonomy, and embedding a GRU-based residual predictor within a sampling-based MPC loop, the proposed approach effectively addresses sensing limitations, communication disruptions, and adversarial interference. Extensive Robotarium experiments demonstrated that the framework consistently outperforms ACC-HDMPC and OR-LSTM baselines by achieving lower tracking, coverage, and total costs, while preserving safety through bounded heading and collision costs. Notably, the hybrid controller maintained reliable performance even as adversarial zone size increased and in larger-scale deployments with up to 12 robots, confirming its scalability in both computation and communication.

References

[1] J. Stephan, J. Fink, V. Kumar, and A. Ribeiro, "Concurrent control of mobility and communication in multirobot systems," *IEEE Trans. Robot.*, vol. 33, no. 5, pp. 1248–1254, Oct. 2017.

[2] T. K. Tasooji and R. Parasuraman, "Distributed fault-tolerant multi-robot cooperative localization in adversarial environments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Hangzhou, China, 2025, pp. 4626–4633.

[3] A. Sagale, T. K. Tasooji, and R. Parasuraman, "Dcl-sparse: Distributed range-only cooperative localization of multi-robots in noisy and sparse sensing graphs," *arXiv preprint arXiv:2412.14793*, 2024.

[4] H. Chen, X. Wang, Y. Jiao, and Y. Li, "An algorithm of coverage flight path planning for uavs in convex polygon areas," *Acta Aeronautica et Astronautica Sinica*, vol. 31, no. 9, pp. 1802–1808, 2010.

[5] Z. H. Ismail and M. G. M. Hamami, "Systematic literature review of swarm robotics strategies applied to target search problem with environment constraints," *Applied Sciences*, vol. 11, no. 5, 2021.

[6] L. Zhao, R. Li, J. Han, and J. Zhang, "A distributed model predictive control-based method for multi-different-target search in unknown environments," *IEEE Trans. Evol. Comput.*, vol. 27, no. 1, pp. 111–125, Feb. 2023.

[7] J. R. Riehl, G. E. Collins, and J. P. Hespanha, "Cooperative search by uav teams: A model predictive approach using dynamic graphs," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 4, pp. 2637–2656, Oct. 2011.

[8] C. Liu and J. K. Hedrick, "Model predictive control-based target search and tracking using autonomous mobile robot with limited sensing domain," in *Proc. Amer. Control Conf. (ACC)*, Seattle, WA, USA, 2017, pp. 2937–2942.

[9] Y. Yang *et al.*, "Multi-agent determinantal q-learning," in *Proceedings of the International Conference on Machine Learning*, 2020, pp. 10 757–10 766.

[10] J. Li and Y. Tan, "The multi-target search problem with environmental restrictions in swarm robotics," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, 2014, pp. 2685–2690.

[11] X. Yu, B. Su, Z. Wang, J. Zhang, and C. Zhang, "Cognitive robotics: Enhancing multirobot target search in unknown environments through adaptive communication strategies," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 55, no. 5, pp. 3449–3463, 2025.

[12] L. Yu, X. Su, Y. Tian, J. Huang, and C. Shen, "Multirobot target searches in unknown environments via waypoint planning system," *IEEE Trans. Ind. Electron.*, vol. 71, no. 7, pp. 7511–7519, Jul. 2024.

[13] J. T. Ebert, F. Berlinger, B. Haghighat, and R. Nagpal, "A hybrid pso algorithm for multi-robot target search and decision awareness," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Kyoto, Japan, 2022, pp. 11 520–11 527.

[14] L. Zhu, J. Cheng, H. Zhang, Z. Cui, W. Zhang, and Y. Liu, "Autonomous and adaptive role selection for multi-robot collaborative area search based on deep reinforcement learning," *IEEE Trans. Autom. Sci. Eng.*, vol. 22, pp. 16 728–16 740, 2025.

[15] Y. Hou, J. Zhao, R. Zhang, X. Cheng, and L. Yang, "Uav swarm cooperative target search: A multi-agent reinforcement learning approach," *IEEE Trans. Intell. Veh.*, vol. 9, no. 1, pp. 568–578, Jan. 2024.

[16] A. Wolek, S. Cheng, D. Goswami, and D. A. Paley, "Cooperative mapping and target search over an unknown occupancy graph using mutual information," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 1071–1078, Apr. 2020.

[17] J. Yu, Z. Chen, Z. Zhao, H. Deng, J. Wu, and J. Xu, "A cooperative hunting method for multi-usvs based on trajectory prediction by or-lstm," *IEEE Trans. Veh. Technol.*, vol. 73, no. 12, pp. 18 087–18 101, Dec. 2024.

[18] J. Peng, H. Viswanath, and A. Bera, "Graph-based decentralized task allocation for multi-robot target localization," *IEEE Robot. Autom. Lett.*, vol. 9, no. 11, pp. 10 676–10 683, Nov. 2024.

[19] L. Zhou, V. D. Sharma, Q. Li, A. Prorok, A. Ribeiro, P. Tokekar, and V. Kumar, "Graph neural networks for decentralized multi-robot target tracking," in *Proc. IEEE Int. Symp. Saf., Secur., Rescue Robot. (SSRR)*, Sevilla, Spain, 2022, pp. 195–202.

[20] J. Ni and S. X. Yang, "Bioinspired neural network for real-time cooperative hunting by multirobots in unknown environments," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2062–2077, Dec. 2011.

[21] N. Hudson and et al., "Heterogeneous ground and air platforms, homogeneous sensing: Team csiro data61's approach to the darpa subterranean challenge," *Field Robotics*, vol. 2, pp. 595–636, 2022.

[22] T. Dang, F. Mascarich, S. Khattak, C. Papachristos, and K. Alexis, "Graph-based path planning for autonomous robotic exploration in subterranean environments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Macau, China, 2019, pp. 3105–3112.

[23] N. Mohajerin and S. L. Waslander, "Modelling a quadrotor vehicle using a modular deep recurrent neural network," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, 2015, pp. 376–381.

[24] S. Wilson, P. Glotfelter, L. Wang, S. Mayya, G. Notomista, M. Mote, and M. Egerstedt, "The robotarium: Globally impactful opportunities, challenges, and lessons learned in remote-access, distributed control of multirobot systems," *IEEE Control Syst. Mag.*, vol. 40, no. 1, pp. 26–44, 2020.