Scilight

Intelligence & Control

https://www.sciltp.com/journals/ic



Article

Check for updates

Spatio-Temporal Coupled Assessment of UAV Swarm Regularity: Spectral Feature Fusion and Anti-Jitter Decision

Ping Zhang [1,†], Rui Zhang [1,†], Vladimir Gainutdinov [2], Yury Furletov [3], Hongwei Liu [1] and Jianwen Huo [1,*]

[1] School of Information and Control Engineering, Southwest University of Science and Technology, Mianyang 621000, China
[2] Institute of Aviation, Land Transportation and Power Engineering, Kazan National Research Technical University (named after A. N. Tupolev—KAI), Kazan 420111, Russia
[3] Chair of Land Vehicles, Moscow Polytechnic University, Moscow 107023, Russia
* Correspondence: huojianwen2008@hotmail.com
† These authors contributed equally to this work.

How To Cite: Zhang, P.; Zhang, R.; Gainutdinov, V.; et al. Spatio-Temporal Coupled Assessment of UAV Swarm Regularity: Spectral Feature Fusion and Anti-Jitter Decision. *Intelligence & Control* **2026**, *2*(4), 5. https://doi.org/10.53941/ic.2026.100019

Received: 14 July 2026
Revised: 27 September 2026
Accepted: 10 October 2026
Published: 16 November 2026

Abstract: Automated regularity assessment of UAV swarm spatial structure is essential for safe swarm control, yet existing distance-based methods are vulnerable to kinematically coupled false positives and unstable decisions near regime boundaries. To address these challenges, this paper first introduces a Spatiotemporal Coupled Dynamic Association Matrix (SCDAM), which incorporates velocity-direction consistency into pairwise distance relationships to distinguish geometrically similar but dynamically different swarm states, thereby improving the reliability of spatial representation. Building upon this dynamic representation, an FFT-based spectral analysis framework is further developed to extract multidimensional regularity descriptors from autocorrelation spectra, achieving robust feature characterization using FFT-based computation. To improve sequence-level decisions, a dual-threshold hysteresis mechanism with an emergency bypass is proposed to suppress unnecessary state switching while preserving rapid responses to abrupt behavioral changes. Experiments on the UCI Swarm Behaviour dataset and physically parameterized offline trajectory/interpolation simulations show higher static accuracy, improved temporal stability, and robustness to the tested coordinate transformations and measurement disturbances. These results support the feasibility of the proposed assessment pipeline under the evaluated conditions; real-flight, onboard sensing, communication, and actuator validation remain future work.

Keywords: UAV swarms; regularity assessment; spatiotemporal coupled dynamic association matrix; FFT-accelerated spectral features; hysteresis decision logic

1. Introduction

A deceptive phenomenon usually emerges when densely flying UAV swarms traverse or temporarily interleave in airspace: two sub-swarms with completely opposite motion trends happen to form a geometrically near-symmetric, evenly spaced arrangement at a certain instant. Recent comprehensive reviews on UAV swarm coordination [1] have highlighted that real-time, reliable spatial state assessment under such kinematically complex scenarios remains an open challenge for onboard control systems, especially as swarm sizes scale into the hundreds with strict collision-avoidance constraints [2]. For evaluators based purely on instantaneous distance, this

Copyright: © 2026 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).
Publisher's Note: Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

configuration is nearly indistinguishable from a stable formation on the "spatial regularity" metric, and is thus misjudged as an "ordered state". Once this erroneous signal is uploaded to the upper-level controller, the cooperative strategy will execute formation-holding commands precisely when emergency avoidance should be triggered, leading to sharply increased collision risk, delayed strategy switching, and even swarm disintegration [3,4]. In large-scale, highly dynamic airspace missions, such "geometrically ordered but dynamically divergent" pseudo-regular states are the fatal flaw that onboard assessment systems must identify from the root. Related autonomy studies cover task allocation, patrolling, and multi-UAV path planning [5–8], illustrating the operational breadth in which assessment signals may be consumed.

The fundamental cause of misclassification lies not in precision or classifier capacity, but in spatiotemporal decoupling at the representation level. The baseline method [3] constructs a relative position matrix (RPM) from nearest-neighbor distance profiles, derives scalar autocorrelation indices using Pearson coefficients, and achieves 94.7% accuracy on UCI Swarm Behaviour [9] with CatBoost [10], providing a strong offline reference. However, its autocorrelation operator acts only on instantaneous geometry and is decoupled from velocity: co-directional and counter-directional configurations with the same relative distances produce identical representations. On the low-directional-consistency subset defined in Section 3.3, baseline accuracy falls to 77.3% and swarm recall to 65.0%, demonstrating the representation-level ambiguity that motivates SCDAM.

Beyond this, the baseline produces approximately 5.2 state switches per sequence on the evaluated synthetic boundary sequences, versus one true transition, which would be undesirable if used as a downstream control signal. Section 2.3.1 specifies the fixed lag count used in the implementation and the scope of the computational comparison. In addition, the three scalar features describe average regularity but not the lag-dependent structure of complex configurations.

Representation-level gap. Existing regularity representations include graph-theoretic connectivity measures, statistical-physics order parameters, and distance autocorrelation descriptors. Graph methods are effective for topology and connectivity [1,11,12], but formations with different geometric shapes can share the same connectivity signature. Global order parameters are coordinate-invariant but can hide local heterogeneity when ordered and disordered sub-regions coexist [13–15]. Distance autocorrelation methods retain geometric detail and low computational cost [3,4,16], yet their inputs are instantaneous distances only: geometrically identical co-directional and counter-directional configurations are mathematically indistinguishable. This spatiotemporal decoupling is the representation problem addressed by the SCDAM layer in this work. Classical consensus and flocking formulations provide foundational coordination models [17,18], but they do not by themselves supply a snapshot regularity representation for the classification setting considered here.

Feature-characterization gap. The baseline distance-autocorrelation framework summarizes each snapshot with three scalar statistics [3]. Such summaries capture average regularity but not the lag-dependent shape or multimodal structure of the underlying autocorrelation sequence. Persistent-homology and topological-data-analysis methods can expose richer collective-motion structure [19–22], but their super-linear computational cost limits streaming use for large swarms. The present work uses the established FFT, spectral energy, and spectral entropy tools [23–25], together with an explicitly defined range-to-mean spectral contrast ratio, to build a compact SCDAM-based spectral representation suitable for tabular classification.

Decision-level gap. A high-quality representation can still yield unstable control signals when a classifier posterior fluctuates near a regime boundary. Single-threshold decisions used in existing assessment pipelines [3,9] do not provide state memory or a separate fast path for abrupt disintegration. Hysteresis and dwell operators are standard tools for bounded-noise suppression [26–28], while recent switched and networked consensus studies emphasize the importance of observer-based robustness, H∞ disturbance attenuation, and stability under changing interaction conditions [29,30]. Related switched heterogeneous tracking work further shows that target uncertainty and switching must be treated explicitly in the decision/control loop [31]. These results motivate, but do not directly validate, the decision layer proposed here: a dual-threshold state machine, sliding-window dwell, and emergency bypass that map the classifier posterior to a stable sequence-level alert.

Deep-learning and other end-to-end models provide useful complementary baselines. PointNet [32], DGCNN [33], and temporal encoders [34] represent different coordinate-input choices, while newer Transformer and rotation-aware architectures continue to improve point-cloud processing [35–37]. However, the present study evaluates only three representative architectures under one protocol; it therefore does not claim a universal limitation for all deep-learning models. Instead, the experiments test a narrower question: how do these evaluated coordinate-input representatives behave under rigid-frame changes, and what interpretability and decision-memory properties are available from the proposed physical feature construction? Graph-based multi-agent reinforcement learning has also been studied for large-scale UAV swarm control [38], alongside graph neural networks for decentralized

swarm control [39], and broader evidence shows that neural-versus-boosted-tree rankings on tabular data depend on the data regime [40].

The resulting research gap is specific and testable: a representation should couple geometry with motion; its feature map should retain informative lag structure without violating latency constraints; and its decision layer should separate frame-wise discrimination from temporal stability. The method below addresses these three gaps in that order. The coordinate invariance of pairwise distances and velocity-direction cosines provides a physically defined property of the proposed inputs, while the deep-learning comparisons are interpreted only within the tested model set and stress-test conditions.

On the spectral side, Wiener's generalized harmonic analysis [23] and the FFT algorithm [24,25] provide mature foundations. The contribution here is not a new FFT algorithm or a claim that spectral energy and entropy are novel primitives. Rather, the method extends SCDAM-based autocorrelation from a scalar to a retained-lag autocorrelation representation, uses three transformed autocorrelation profiles (distance, differential, and bias), and summarizes each with energy, range-to-mean spectral contrast ratio, and entropy. To the best of our knowledge, this systematic SCDAM-based construction has not been established in the literature examined here. CatBoost [10] is used as a transparent tabular classifier; XGBoost [41] and LightGBM [42] are relevant alternatives. At the decision side, hysteresis, dwell, and event-triggered mechanisms [26–28,43,44] provide established context for suppressing noise-induced switching while retaining fast response. Observer-based positive consensus and H∞ non-negative consensus studies [29,30], together with switched heterogeneous consensus tracking [31], further highlight why robustness and switching behavior should be specified separately from instantaneous classification accuracy.

To address these gaps, this paper develops a unified three-layer assessment chain: (i) representation, where SCDAM couples pairwise geometry with velocity-direction consistency; (ii) feature characterization, where FFT descriptors expand the information available from the SCDAM autocorrelation; and (iii) decision, where hysteresis, dwell, and emergency bypass convert CatBoost posteriors into stable sequence-level outputs. The layers are sequential rather than interchangeable: SCDAM supplies the physically discriminative signal, spectral features enrich that signal, and the decision logic acts only after classification. Specific contributions are:

1. Representation level—physics–data jointly driven spatiotemporal representation. SCDAM incorporates velocity-direction consistency into pairwise distances, allowing geometrically similar but dynamically different states to be separated while retaining the baseline's nearest-neighbor and Pearson-autocorrelation structure.
2. Feature level—compact FFT-based autocorrelation characterization. On the SCDAM representation, three autocorrelation spectra are summarized by nine compact descriptors: energy, the explicitly defined range-to-mean spectral contrast ratio, and entropy. FFT provides an efficient implementation; the contribution lies in the SCDAM-to-spectrum construction and its physically interpretable feature organization.
3. Decision level—sequence-stable posterior mapping. A dual-threshold hysteresis state machine, sliding-window dwell, and emergency bypass suppress boundary chatter and preserve rapid response to abrupt changes. This layer is evaluated by switching, delay, and DTW metrics, not by independent frame-wise accuracy.

On the UCI Swarm Behaviour dataset [9], the proposed method achieves 99.79% mean classification accuracy across five seeds, exceeding the CatBoost-3D comparison by 6.30 percentage points. On the seed-42 challenge subsets, gains are 21.9, 24.2, and 42.3 pp for dynamic coupling, spatial ambiguity, and boundary judgment. On the evaluated synthetic sequences, switching and DTW distance decrease by 81.9% and 73.8%, respectively; robustness is separately evaluated under the stated Gaussian-noise protocol.

2. Materials and Methods

2.1. Preliminaries and Problem Statement

Let $\mathbb{R}^n$ be the n-dimensional Euclidean space, $A^\top$ be the matrix transpose of the matrix A, $\langle a, b\rangle = a^\top b$ the inner product, $\|a\|_2$ the Euclidean norm, and $\mathbb{I}(\cdot)$ the indicator function. The other symbols are given in Table 1.

2.1.1. Baseline Framework

Consider N UAVs operating in a two-dimensional task plane with $p_k(t) \in \mathbb{R}^2$ being the position of the kth UAV at time t. The pairwise distance matrix is defined as

$$R(t) = \begin{bmatrix} 0 & r_{12} & \cdots & r_{1N} \\ r_{21} & 0 & \cdots & r_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ r_{N1} & r_{N2} & \cdots & 0 \end{bmatrix}, \quad r_{kl}(t) = \|p_k(t) - p_l(t)\|_2 \tag{1}$$

which is symmetric with zero diagonal and invariant under rigid body transformations. For each UAV k, the $N-1$ off-diagonal distances are sorted in ascending order, and the m nearest neighbors are taken to form the RPM:

$$B(t) = \begin{bmatrix} r_{11}^* & \cdots & r_{1N}^* \\ \vdots & \ddots & \vdots \\ r_{m1}^* & \cdots & r_{mN}^* \end{bmatrix} \in \mathbb{R}^{m \times N} \tag{2}$$

where r_{jk}^* is the distance from UAV k to its j-th nearest neighbor, and column $B[k] = [r_{1k}^*, \ldots, r_{mk}^*]^\top$ forms k's local neighborhood descriptor. The row index of $B[k]$ encodes neighbor rank rather than fixed identity. The structural similarity between two UAVs is measured by the Pearson correlation coefficient of RPM columns:

$$\rho_{kl}(t) = \frac{\sum_{i=1}^{m}(b_{ik}-\bar{B}[k])(b_{il}-\bar{B}[l])}{\sqrt{\sum_{i=1}^{m}(b_{ik}-\bar{B}[k])^2 \cdot \sqrt{\sum_{i=1}^{m}(b_{il}-\bar{B}[l])^2}}}, \quad \bar{B}[k] = \frac{1}{m}\sum_{i=1}^{m} b_{ik} \tag{3}$$

with $\bar{B}[k] = \frac{1}{m}\sum_{i=1}^{m} b_{ik}$. The column mean is taken over i = 1, …, m; the same definition applies to column l. By the Cauchy–Schwarz inequality, $\rho_{kl} \in [-1,1]$. The global structural autocorrelation index is

$$\mathrm{corr}(t) = \frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\rho_{kl}\,(t) \tag{4}$$

Three variants are defined by modifying $B[k]$: variant 1 uses raw sorted distances; variant 2 subtracts the swarm means for each neighbor rank; variant 3 uses first-order differences. The three scalars form a feature vector $f \in \mathbb{R}^3$, input to CatBoost [10]. Undefined Pearson coefficients from zero-variance columns are set to zero in the implementation.

Table 1. Symbols and definitions.

Symbol	Definition
N	Number of UAVs in the swarm
m	Neighborhood size parameter
$p_k(t)$, $v_k(t)$	Position and velocity of UAV k
$d_k(t)$	Unit velocity direction vector
r_{kl}	Euclidean distance between UAVs k and l
$R(t)$	Pairwise distance matrix
$B(t)$	Relative position matrix (RPM)
$\omega_{kl}(t)$, $\bar{\omega}(t)$	Pairwise velocity-direction weight and its mean over all unordered UAV pairs
λ	SCDAM coupling sensitivity coefficient
$M(t)$	Spatiotemporal coupled dynamic association matrix
M_{e}	Number of upper triangular elements $N(N-1)/2$
Φ_{dist}, Φ_{diff}, Φ_{bias}	Distance, differential, and bias spectra
$\mathbf{f}_{\mathrm{ext}}$	Enhanced 13-dimensional feature vector
$p(t)$	CatBoost posterior regularity probability
T_{high}, T_{low}, T_{emg}	Hysteresis thresholds and emergency bypass threshold
$S(t)$	Hysteresis state variable
$\hat{y}(t), \hat{y}_{\mathrm{final}}(t)$	Windowed and final regularity decisions

2.1.2. Problem Statement

The baseline [3] constructs a three-dimensional feature vector $f \in \mathbb{R}^3$ from position information only via Equations (1)–(4), limiting fine-grained discrimination in dynamically ambiguous large-swarm snapshots. Given position sequences $\{p_k(t)\}$, velocity sequences $\{v_k(t)\}$, and expert binary labels $y(t)$, the objective is to construct the enhanced mapping in Equation (5).

$$\mathcal{F}: \{p_k(t), v_k(t)\}_{k=1}^{N} \mapsto \hat{y}(t) \in \{0,1\} \tag{5}$$

satisfying five requirements: (R1) Dynamic consistency—configurations with identical geometry and different pairwise direction-consistency patterns satisfy $M(C_1) \neq M(C_2)$; (R2) Coordinate invariance—output invariant under arbitrary rigid body transformations; (R3) Computational feasibility in the PC benchmark—end-to-end latency ≤ 20 ms at $N = 200$; (R4) Temporal stability—at most 2 switches per sequence in boundary regions, abrupt disintegration delay ≤ 20 ms; (R5) Feature compactness—each feature in $\mathbf{f}_{\mathrm{ext}}$ contributes > 0.1% accuracy gain, or maximum pairwise Pearson correlation ≤ 0.85.

2.2. Spatiotemporal Coupled Dynamic Association Matrix

SCDAM is the core representation layer of the framework. The complete processing chain is: raw positions and velocities → SCDAM representation → spectral and scalar feature enhancement → CatBoost posterior → hysteresis/dwell/emergency decision → final temporal output. Thus, FFT does not replace SCDAM and hysteresis does not modify the feature representation or the frame-wise classifier; each layer addresses a different failure mode. Figure 1 summarizes the complete assessment pipeline.

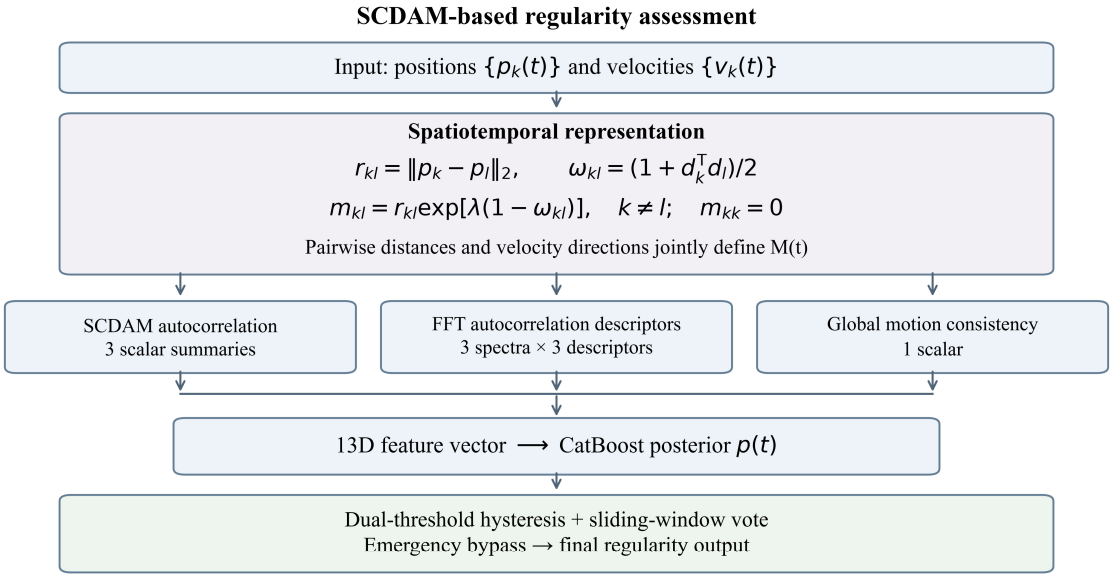


Figure 1. Overall framework of the SCDAM method.

2.2.1. Motion Consistency Weight

For each independent snapshot, the implementation sets $d_k(t) = v_k(t)/\|v_k(t)\|_2$ when $\|v_k(t)\|_2 > 10^{-6}$ in the input velocity units, and $d_k(t) = 0$ otherwise. No previous direction is retained. For two nonzero direction vectors, their inner product equals the angle cosine in Equation (6). If either direction is zero, the implementation assigns $s_{kl}(t) = 0$ and $\omega_{kl}(t) = 0.5$; this is a numerical convention for unavailable direction, not evidence of orthogonal motion.

$$s_{kl}(t) = d_k(t) \cdot d_l(t) = \cos\theta_{kl}(t) \in [-1,1] \tag{6}$$

$$\omega_{kl}(t) = \frac{1 + s_{kl}(t)}{2} \in [0,1] \tag{7}$$

where $\omega_{kl} = 1$ corresponds to fully co-directional, $\omega_{kl} = 0$ to fully counter-directional, and, for two nonzero directions, ω_{kl}=0.5 to orthogonal. The global consistency indicator $\bar{\omega}(t) = (2/[N(N-1)])\sum_{k<l}\omega_{kl}(t)$. Since $(Ad_k) \cdot (Ad_l) = d_k \cdot d_l$ for any rotation $A \in \mathrm{SO}(n)$, $\omega_{kl}(t)$ is invariant under rigid body transformations, guaranteeing the coordinate-system independence of subsequent matrices from the source.

2.2.2. Construction and Properties of SCDAM

Given $R(t)$ and the weight matrix $\Omega(t) = [\omega_{kl}(t)]$, the elements of SCDAM $M(t) \in \mathbb{R}^{N\times N}$ are defined as

$$m_{kl}(t) = r_{kl}(t) \cdot \exp(\lambda(1 - \omega_{kl}(t))), \quad k \neq l, \quad m_{kk}(t) = 0 \tag{8}$$

where $\lambda > 0$ is the coupling sensitivity coefficient. The exponential factor amplifies the effective distance for kinematically uncoordinated UAV pairs: when $\omega_{kl} = 1$, $m_{kl} = r_{kl}$; when $\omega_{kl} = 0$ and $\lambda = 1$, the factor is $e^{\lambda} \approx 2.72$.

The exponential form is strictly positive, smooth, and monotonically increasing in motion inconsistency, so the effective distance remains non-negative and changes continuously with the direction-consistency weight. At moderate λ it can preserve much of the local distance ordering, although pair-specific weights can change neighbor ranks. It also provides a direct interpretation: at $\lambda = 1$, a fully counter-directional pair is assigned e ≈

2.72 times its physical distance. The affine alternative $1 + \lambda(1 - \omega_{kl})$ is also positive for $\lambda \geq 0$ but grows only linearly; the exponential was selected for its multiplicative separation and one-parameter interpretation, not because the affine form becomes non-positive.

SCDAM has four key properties. Backward compatibility: as $\lambda \to 0$, $M(t) \to R(t)$, and the baseline is a special case with $\lambda = 0$. Rigid-body invariance: both r_{kl} and ω_{kl} are invariant under rigid body transformations, so each m_{kl} is invariant, satisfying R2. Symmetry: $M(t)$ is a real symmetric matrix with zero diagonal, inheriting the matrix structure of $R(t)$. Conditional dynamic identifiability: configurations with identical noncoincident positions whose velocity fields yield different pairwise direction-consistency values have at least one pair (k, l) such that $\omega_{kl}^{(1)} \neq \omega_{kl}^{(2)}$, hence $M(C_1) \neq M(C_2)$, satisfying R1. These four properties ensure that SCDAM achieves a clean compromise between "retaining the original mathematical structure" and "injecting physical priors".

2.2.3. Driving Baseline Variants with SCDAM

SCDAM replaces $R(t)$ as input to the three autocorrelation variants in Section 2.1. For each UAV, the effective pairwise distances are sorted and the m smallest values form its enhanced RPM column; Equations (3) and (4) are then applied to obtain the three SCDAM autocorrelation values. Counter-directional pairs receive larger effective separations than co-directional pairs at the same physical distance. This changes the local profiles used by the classifier and, empirically, improves classification on the evaluated low-directional-consistency subset; a lower autocorrelation value is not asserted for every individual configuration.

2.2.4. Coupling Sensitivity Coefficient

A grid search over λ values {0.2, 0.4, ..., 2.0} shows validation accuracy of 99.72–99.89% and test accuracy of 99.58–99.83% across the evaluated grid. We select λ = 1.0 as the default, corresponding to amplification factors $e \approx 2.72$ for counter-directional pairs and $e^{0.5} \approx 1.65$ for orthogonal pairs. Its validation accuracy is 0.08 pp below the grid maximum; the observed plateau supports this interpretable central setting, while deployment-specific calibration remains necessary.

2.3. FFT-Accelerated Spectral Feature Extraction

The three scalar outputs of SCDAM capture average structural regularity. This subsection extends those outputs to retained-lag autocorrelation profiles based on the Wiener–Khinchin theorem, with FFT used to meet the computational constraint. FFT, spectral energy, and entropy are established tools; the range-to-mean spectral contrast ratio is explicitly defined below; the methodological contribution is their systematic application to the SCDAM representation through distance, differential, and bias spectra, yielding an efficient and physically interpretable FFT-computed extension of the baseline. Here "spectra" denotes the retained-lag autocorrelation profiles, rather than power spectral densities; the descriptors below are computed on these lag-indexed profiles.

2.3.1. Computational Bottleneck and FFT Solution

The upper-triangular entries of $M(t)$ form a lexicographically ordered sequence q of length $M_e = N(N-1)/2$. Let L denote the length of the sequence supplied to the autocorrelation routine. For a non-degenerate input, the implementation retains $K = \min(50, L-1)$ normalized autocorrelation values at lags $\tau = 0,\ldots,K-1$. The distance, first-difference, and absolute-deviation branches have lengths M_e, M_e-1, and M_e, respectively. Thus, at $N = 200$ the distance branch retains 50 lag values; the earlier expression $\lfloor M_e/4 \rfloor = 4975$ is not the lag count used to produce the reported feature vectors. For an input variance below 10^{-15}, the implementation returns a zero autocorrelation sequence. Here K counts the retained values including lag zero; the same count is used in Equations (11)–(13). In Equation (9), $\tilde{q} = q - \bar{q}$ is the centered input and $\bar{q}$ is its sample mean. Equation (9) gives the unnormalized FFT identity; the normalization is specified in Equation (10).

$$R(\tau) = \mathcal{F}^{-1}\{|\mathcal{F}\{\tilde{q}\}|^2\}[\tau] \tag{9}$$

Direct computation of K retained lags requires $O(LK)$ operations. This is $O(L^2)$ only when K grows proportionally with L; for the implemented fixed cap $K = 50$, direct computation is $O(L)$. The zero-padded FFT computation requires $O(L \log L)$ operations before truncation. The implementation centers the input, zero-pads to a power of two no smaller than $2L$, computes the power spectrum, applies the inverse FFT, and retains the first K lag values ($\tau = 0,\ldots,K-1$). On the evaluated PC Python environment at $N = 200$, end-to-end latency is 11.26 ms, with FFT autocorrelation-profile extraction accounting for 80.2%. Figure 2 reports the measured component runtimes and scalability. The measured log–log slope over $N \in \{50,\ldots,500\}$ is 1.97; this finite-range observation

does not establish an asymptotic order or an advantage over a matched fixed-K direct method. Since L is proportional to N^2, the stated FFT operation count is O(N^2 log N). For numerical stability, the implementation adds 10^{-15} to the denominator in Equation (10).

2.3.2. Three Spectral Components and the 13-Dimensional Feature Vector

Three transformed autocorrelation profiles extend the baseline scalar summaries. The distance spectrum $\Phi_{\text{dist}}(\tau)$ uses the normalized autocorrelation of q in Equation (10); the differential spectrum $\Phi_{\text{diff}}(\tau)$ operates on the first-order difference $u_j = q_{j+1} - q_j$; the bias spectrum $\Phi_{\text{bias}}(\tau)$ operates on the absolute deviation $\delta_i = |q_i - \bar{q}|$:

$$\Phi_x(\tau) = \frac{\mathcal{F}^{-1}\{|\mathcal{F}\{x - \bar{x}\}|^2\}[\tau]}{(L_x - \tau)\sigma_x^2}, x \in \{q, u, \delta\} \tag{10}$$

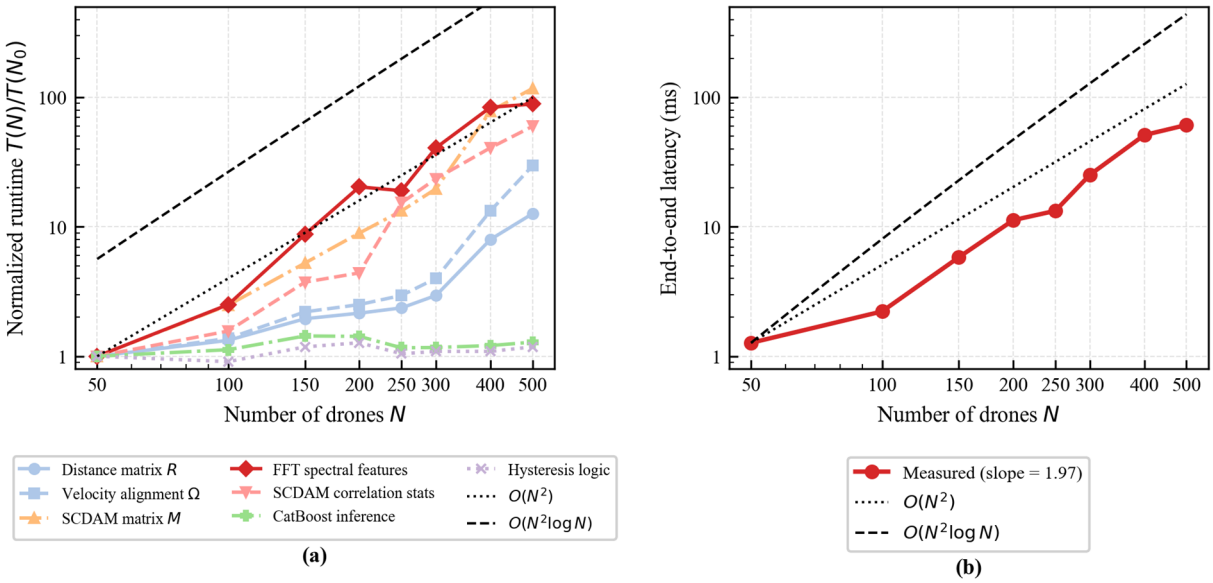


Figure 2. End-to-end latency and scalability. (**a**) Normalized runtime of each pipeline component; (**b**) absolute end-to-end latency. The measured log–log slope is 1.97 over the tested range; it is a finite-range observation, not a proof of asymptotic order or a matched fixed-lag comparison.

For each spectrum Φ, three compact descriptors are extracted: spectral energy $E(\Phi)$, range-to-mean spectral contrast ratio $P(\Phi)$, and spectral information entropy $H(\Phi)$:

$$E(\Phi) = \frac{1}{K}\sum_{\tau=0}^{K-1} \Phi(\tau)^2$$

$$P(\Phi) = \frac{\max_\tau \Phi(\tau) - \min_\tau \Phi(\tau)}{\text{mean}_\tau |\Phi(\tau)| + 10^{-8}}$$

$$\tilde{\Phi}(\tau) = \frac{|\Phi(\tau)| + 10^{-15}}{\sum_{s=0}^{K-1}(|\Phi(s)| + 10^{-15})}$$

$$H(\Phi) = -\sum_{\tau=0}^{K-1} \tilde{\Phi}(\tau)\log_2 \tilde{\Phi}(\tau)$$

$$\mathbf{f}_{\text{ext}} = [E_d, P_d, H_d, E_f, P_f, H_f, E_b, P_b, H_b, \bar{\omega}, \widetilde{\text{corr}}(1), \widetilde{\text{corr}}(2), \widetilde{\text{corr}}(3)]^{\text{T}} \in \mathbb{R}^{13} \tag{11}$$

Here, subscripts d, f, and b denote the distance, differential, and bias spectra, respectively. The nine spectral descriptors, three SCDAM autocorrelation values, and one global velocity-consistency scalar give 9 + 3 + 1 = 13 dimensions. A candidate 14th cross-spectral correlation feature showed no positive ablation gain and was

excluded under R5. The entropy uses the standard convention $0 \log_2 0 = 0$. Table 2 summarizes the resulting feature vector.

Table 2. Compact summary of the 13-dimensional feature vector.

Feature Group	Features	Count	Physical Meaning
SCDAM autocorrelation	$\widetilde{\mathrm{corr}}(1), \widetilde{\mathrm{corr}}(2), \widetilde{\mathrm{corr}}(3)$	3	Enhanced spatial regularity for raw, bias-centered, and first-difference variants
Global motion consistency	$\bar{\omega}$	1	Mean pairwise velocity-direction consistency
Distance spectrum	E_d, P_d, H_d	3	Energy, range-to-mean spectral contrast ratio, and entropy of distance autocorrelation spectrum
Differential spectrum	E_f, P_f, H_f	3	The same descriptors for the first-difference spectrum
Bias spectrum	E_b, P_b, H_b	3	The same descriptors for the absolute-deviation spectrum
Total		13	Compact feature vector used by CatBoost

2.3.3. Relationship to the Baseline Scalar Autocorrelation

The baseline and spectral branches both use second-order correlation information, but they index it differently. Let $q = (q_1, \ldots, q_L)$ denote the sequence of upper-triangular SCDAM entries, with variance σ_q^2. Equation (12) uses centered terms $\tilde{q}_i = q_i - \bar{q}$ and defines their normalized lag autocorrelation; Equation (13) defines the mean over retained lags. For input variance below 10^{-15}, the implementation returns a zero autocorrelation sequence; the normalized formula applies to nondegenerate inputs. The numerical stabilizer specified for Equation (10) is also used in the implementation of Equation (12).

$$\Phi_{\mathrm{dist}}(\tau) = \frac{1}{(L-\tau)\sigma^2{}_q} \sum_{i=1}^{L-\tau} \tilde{q}_i \tilde{q}_{i+\tau}, \tau = 0, 1, \ldots, K-1 \tag{12}$$

Taking the arithmetic mean over all lags:

$$\bar{\Phi}_{\mathrm{dist}} = \frac{1}{K} \sum_{\tau=0}^{K-1} \Phi_{\mathrm{dist}}(\tau) \tag{13}$$

The first SCDAM Pearson feature, by contrast, averages coefficients between sorted neighborhood columns. Flattened-sequence lag indexing and neighborhood-column indexing are not the same operation, so no exact or asymptotic algebraic equality is claimed:

$$\bar{\Phi}_{\mathrm{dist}} \not\equiv \widetilde{\mathrm{corr}}(1) \tag{14}$$

The two branches nevertheless describe complementary aspects of second-order structure: the three SCDAM Pearson summaries aggregate neighborhood similarity, whereas the spectral descriptors retain lag-dependent behavior of the flattened SCDAM sequence. The 13D vector therefore combines, rather than mathematically subsumes, the baseline-style autocorrelation and spectral information.

2.4. Dual-Threshold Hysteresis Decision and Emergency Bypass

The 13-dimensional features $\mathbf{f}_{\mathrm{ext}}$ are input to CatBoost to output the posterior regularity probability $p(t)$. In safety-critical systems, the decision layer must not only pursue high static accuracy, but also simultaneously balance two mutually contradictory requirements: robust suppression of measurement noise near boundaries, and fast response to abrupt disintegration events. The former requires wide dead-bands and long integration windows, while the latter requires narrow thresholds and short windows. This subsection resolves this contradiction at the design level by decoupling noise suppression, gradual smoothing, and abrupt response into three independent decision paths. A single threshold can already achieve 99.78% static accuracy, but produces approximately 3.0 switches per sequence in boundary regions, exceeding the R4 constraint.

2.4.1. Hysteresis State Machine and Sliding Window Dwell

On samples not handled by the emergency bypass in Section 2.4.2, the hysteresis state $S(t) \in \{0,1\}$ is defined using an activation threshold T_{high} and a deactivation threshold T_{low}, where $(0 < T_{\mathrm{low}} < T_{\mathrm{high}} < 1)$, as follows:

$$S(t) = \begin{cases} 1, & p(t) > T_{\text{high}}, \\ 0, & p(t) < T_{\text{low}}, \\ S(t - \Delta t), & \text{otherwise.} \end{cases} \quad (15)$$

The previous state is retained when $p(t)$ lies in the dead band $[T_{\text{low}}, T_{\text{high}}]$. We set $T_{\text{high}} = 0.75$ and $T_{\text{low}} = 0.55$, giving $\Delta T = 0.20$. For a bounded perturbation of the classifier posterior, a single excursion smaller than half the dead-band width cannot traverse both thresholds; this is a conditional property of the decision variable, not a sensor-level guarantee. The buffer retains the latest at most N_{win} state updates, including repeated values; let $b_1, \ldots, b_{n(t)}$ be its entries and $n(t)$ its current length. Its mean is $L(t) = \frac{1}{n(t)} \sum_{j=1}^{n(t)} b_j$, with $\hat{y}(t) = \mathrm{I}(L(t) > \eta)$. With $N_{\text{win}} = 10$ and $\eta = 0.7$, a sustained change in S requires three irregular-state samples for a 1→0 output and eight regular-state samples for a 0→1 output when a full buffer initially contains the opposite state and no bypass intervenes. At $\Delta t = 20$ ms, these nominal window contributions are 60 and 160 ms, respectively; measured end-to-end delays are reported in Section 3.4.

2.4.2. Emergency Bypass

The dwell-window delay can be unsuitable for abrupt disintegration events. Introduce emergency bypass threshold $T_{\text{emg}} < T_{\text{low}}$:

$$\hat{y}_{\text{final}}(t) = \begin{cases} 0, & p(t) < T_{\text{emg}}, \\ \hat{y}(t), & \text{otherwise.} \end{cases} \quad (16)$$

when $p(t) < T_{\text{emg}}$, an irregular alert bypasses the dwell window and is emitted in the current sample, so the sampling-based delay is at most $\Delta t = 20$ ms. We set $T_{\text{emg}} = 0.20$. The bypass is checked first and leaves the hysteresis state and buffer unchanged; Equation (15) and the window update run only on other samples. In the static grid used for the sensitivity analysis in Section 3.5, varying T_{emg} over [0.05, 0.35] did not change overall accuracy because the bypass is a sequence-level mechanism; its dynamic effect is therefore evaluated separately by the no-bypass ablation in Section 3.4. The dead band, dwell window, and emergency path address noise excursions, gradual transitions, and abrupt low-posterior events, respectively. The three decision layers are summarized in Table 3.

Table 3. Latency–stability trade-off of the three-layer decision architecture.

Operating Scenario	Dominant Decision Layer	Response Delay
Sensor noise near boundary	Dead-band isolation	No state change
Progressive regularity transition	Sliding window dwell	Window- and direction-dependent (Section 3.4)
Abrupt disintegration ($p < T_{\text{emg}}$)	Emergency bypass	≤ 20 ms

2.5. Dataset and Experimental Setup

Experiments use the UCI Swarm Behaviour dataset [9], organized by Abpeikar and Kasmarik [45] and annotated using the human-perception protocol of Walker et al. [46].

The source file contains 24,016 records; after removing one incomplete record, 24,015 samples remain, with 12,007 swarm and 12,008 non-swarm labels. Each sample contains $N = 200$ UAVs, two-dimensional positions, velocity components, and a binary label. The velocity field is the simulator's instantaneous output rather than a finite difference of position. The low lag-1 label correlation ($r_1 = 0.047$) is consistent with treating rows as independent snapshots for the random 70%/15%/15% split.

Core parameters are $m = 8$, $\lambda = 1.0$, direction-normalization tolerance = 10^{-6}, $T_{\text{high}} = 0.75$, $T_{\text{low}} = 0.55$, $T_{\text{emg}} = 0.20$, $N_{\text{win}} = 10$, and $\eta = 0.7$; CatBoost uses depth 8, up to 300 iterations, and learning rate 0.03. The primary static comparison is repeated over five seeds {42, 123, 456, 789, 2024}; Section 3.1 reports seed 42 as the representative split and also reports the five-seed summary in the text. Other experiments use the sample counts, seeds, and Monte Carlo repetitions stated in their respective subsections and captions.

3. Results

3.1. Overall Classification Performance

Table 4 reports the test set accuracy of baseline variants and progressive feature enhancement (seed = 42 as representative).

Table 4. Test set static classification accuracy (%).

Method	Swarm Recall	Non-Swarm Recall	Overall Accuracy
Variant 1 (baseline) [3]	17.5	90.2	54.2
Variant 2 (baseline) [3]	71.4	96.3	84.0
Variant 3 (baseline) [3]	61.6	99.7	80.8
CatBoost-3D (baseline) [3]	87.5	99.6	93.6
+ Global motion consistency (4D)	98.1	98.2	98.1
+ Spectral features (13D)	99.9	99.6	99.8
Proposed (13D + hysteresis)	99.9	99.6	99.8

Variant 1/2 thresholds are searched on the validation set; variant 3 uses the dual-interval rule from [3]; the CatBoost-3D row uses the original ensemble scheme of [3]. The 4D row combines the baseline distance features with the global direction-consistency scalar; the 13D row uses Equation (11). Table 4 reports independent frame-wise test accuracy only, without carrying temporal state between samples. Single variants reach 54.2–84.0%, CatBoost-3D reaches 93.6%, adding global motion consistency to the three distance features gives 98.1% (+4.5 pp), and the 13D feature set reaches 99.8% (+1.7 pp). The identical rounded accuracy of the 13D and 13D+hysteresis rows is expected because hysteresis is evaluated on sequences in Section 3.4. Across five seeds, frame-wise accuracy is 99.79% (95% CI [99.68%, 99.90%]) versus 93.49% for CatBoost-3D.

3.2. Ablation Study

Table 5 quantifies the contribution of each component.

Table 5 supports two findings: reverting to the distance-only 3D baseline lowers accuracy by 6.2 percentage points, while removing all spectral features causes a 1.0-point drop. This full-versus-baseline difference does not isolate SCDAM alone. Removing an individual spectral group changes accuracy by 0.1–0.2 points.

Table 5. Ablation results—test set overall accuracy (%).

Configuration	Overall Accuracy	Δ_{acc}
Full method (13D + hysteresis)	99.8	—
Distance-only 3D baseline	93.6	−6.2
Remove all spectral features	98.8	−1.0
Remove the Φ_{dist} group	99.7	−0.1
Remove the Φ_{diff} group	99.6	−0.2
Remove Φ_{bias} group	99.6	−0.2
Remove $\bar{\omega}$	98.2	−1.6
Add cross-spectral C_{12} (14D)	99.7	−0.1

The three individual spectral-group drops sum to 0.5 percentage points, half the 1.0-point drop from removing all spectral features, suggesting a joint benefit without establishing a statistical interaction; removing $\bar{\omega}$ alone causes a 1.6 pp drop (regular-state recall 99.9% → 96.9%), confirming its global velocity alignment signal cannot be fully replaced by SCDAM local weights; C_{12} brings no positive gain ($\Delta_{acc} = -0.1$ pp), excluded per R5. In Figure 3, $\rho_i(M)$ denotes $\widetilde{\text{corr}}(i)$ for $i \in \{1, 2, 3\}$. Both reported importance metrics rank $\bar{\omega}$ first (32.8% / 41.8%), with $\rho_2(M)$ and $\rho_3(M)$ alternating in second and third positions forming the second tier.

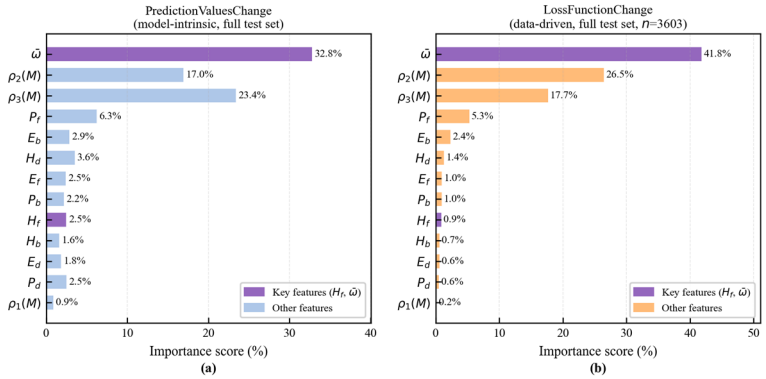


Figure 3. CatBoost-13D feature importance comparison by two methods: (**a**) Prediction Values Change (model-intrinsic); (**b**) Loss Function Change (data-driven, full test set, $n = 3603$).

3.3. Challenge Scenario Analysis

The three challenge subsets are diagnostic filters applied to the 3603 held-out snapshots in the seed-42 test split. Both classifiers are evaluated on the same samples within each subset, without retraining for a subset. The rules use the global mean direction-consistency score $\bar{\omega}$ (Section 2.2), the second baseline autocorrelation variant $\rho_2(R)$ (Section 2.1), or the CatBoost-3D swarm posterior p_{3D}. Here $y = 1$ denotes the dataset swarm label, not an independently verified collision-free formation. The rules and their motivations are:

(i) Dynamic coupling: select $\bar{\omega} < 0.60$ ($n = 946$; 611 swarm and 335 non-swarm samples). This low-directional-consistency subset probes information absent from the distance-only baseline: instantaneous geometry alone does not encode whether velocities are aligned. Low $\bar{\omega}$ can occur with opposed or dispersed headings; this filter is a proxy for weak directional coordination, not proof of physical trajectory interleaving.

(ii) Spatial ambiguity: select $y = 1$ and $\rho_2(R) < 0.08$ ($n = 919$, all swarm samples). These positive examples have a weak value of the baseline spatial descriptor and test missed swarm detections when that descriptor provides insufficient evidence. The label is used only to define this diagnostic subset, not as a classifier input. Because it contains only positives, the accuracy in this row of Table 6 equals swarm recall and does not measure non-swarm rejection.

(iii) Boundary judgment: select $0.30 \leq p_{3D} \leq 0.70$ ($n = 78$; 36 swarm and 42 non-swarm samples). This interval includes the baseline decision threshold of 0.585 and isolates ambiguous baseline scores. It tests frame-wise discrimination near the decision boundary; it is not a test of temporal switching, which is evaluated separately in Section 3.4.

The cutoffs define operational diagnostic subsets, not universal physical transition thresholds. The subsets overlap and no selection rule uses the proposed model's prediction or correctness. Table 6 uses independent frame-wise decisions: the baseline threshold is 0.585, while the saved 13D snapshot rule uses $T_{high} = 0.500$ and $T_{low} = 0.350$, with dead-band samples assigned by the midpoint (0.500 + 0.350)/2 = 0.425. Its $T_{emg} = 0.150$ lies below T_{low} and does not change these independent snapshot decisions. This rule carries no temporal state. By contrast, the single-threshold comparator in Section 3.4 uses the validation-selected 13D threshold 0.450; the two values refer to different evaluation rules. The gains of 21.9, 24.2, and 42.3 pp therefore describe these conditional tests, not a ranking of scenario difficulty or population-wide improvements. Figure 4 shows that, on the dynamic-coupling subset, baseline swarm recall is 397/611 = 65.0%, whereas the 13D classifier achieves 611/611 = 100.0%; its eight false-positive non-swarm predictions give 99.2% overall accuracy.

Table 6. Frame-wise challenge-subset accuracy (%); the spatial-ambiguity row contains only positive samples, so its accuracy equals swarm recall.

Scenario	n	CatBoost-3D [3]	Proposed	Gain
Dynamic coupling	946	77.3	99.2	+21.9
Spatial ambiguity	919	75.7	99.9	+24.2
Boundary judgment	78	56.4	98.7	+42.3

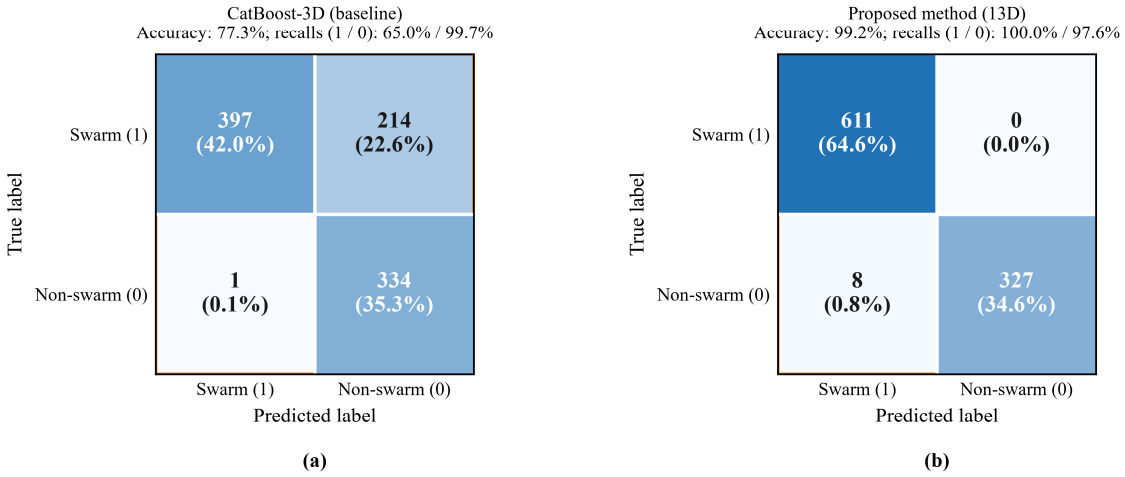


Figure 4. Confusion matrices on the dynamic-coupling subset ($\bar{\omega} < 0.60$). (**a**) CatBoost-3D (baseline); (**b**) proposed method (13D). Rows denote true labels and columns predicted labels; cell percentages are relative to all 946 samples.

3.4. Temporal Stability

Temporal stability is evaluated on 100 parameterized synthetic probability sequences (50 frames each), with transition levels calibrated to the classifier thresholds on the validation split ($\text{thr}_{3D} = 0.585$ and $\text{thr}_{13D} = 0.450$). Each sequence contains a stable region ($\sigma \approx 0.05$) and a transition region ($\sigma_{3D} = 0.13$ and $\sigma_{13D} = 0.07$). Here the state starts at 0 with an empty buffer; during warm-up its mean uses the current buffer length. Table 7 reports aggregate metrics, and Figure 5 shows their distributions.

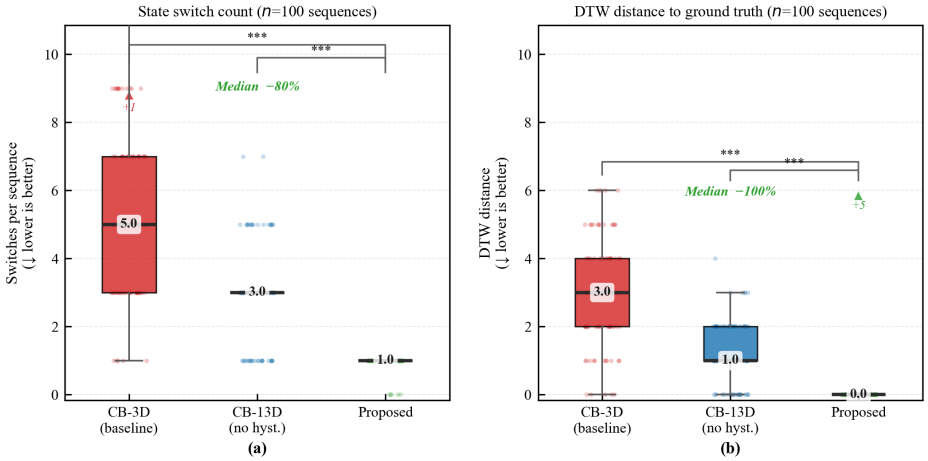


Figure 5. Temporal stability boxplot comparison ($n = 100$ sequences). (**a**) State switch count per sequence; (**b**) DTW distance to ground truth. *** $p < 0.001$ (Wilcoxon signed-rank test, two-sided). Brackets compare each baseline with Proposed.

Table 7. Temporal stability metrics (100 sequences × 50 frames).

Method	Switches/Sequence (mean ± std)	DTW Distance (mean ± std)
CatBoost-3D (single threshold)	5.24 ± 2.21	2.98 ± 1.49
CatBoost-13D (single threshold)	2.96 ± 1.39	1.25 ± 0.90
Proposed (full)	0.95 ± 0.22	0.78 ± 3.40

The proposed method reduces average switching by 81.9% (5.24 → 0.95) and DTW distance by 73.8% (2.98 → 0.78). The median switch count of 1.0 equals the single true transition in each sequence, and the switch-count standard deviation is 0.22. Two-sided Wilcoxon signed-rank tests (normal approximation with tie correction; zero differences excluded; no continuity correction) give $p = 1.43 \times 10^{-17}$ for switch count and $p = 1.29 \times 10^{-11}$ for DTW distance. The progression 5.24 → 2.96 → 0.95 summarizes switching under the specified synthetic-probability model; these results characterize the evaluated synthetic sequences rather than guarantee closed-loop stability. The DTW dispersion for the full method is right-skewed: only 5 of 100 sequences have nonzero DTW distance (median 0; maximum 16), which explains the reported mean of 0.78 and standard deviation of 3.40. Table 7 reports descriptive standard deviations using divisor n; percentage reductions use unrounded means.

Because the UCI records are independent snapshots (lag-1 label autocorrelation $r_1 = 0.047$) and do not directly contain temporal behavior, sequence-level decision stability is evaluated using an offline dual-anchor interpolation simulation. Swarm anchors satisfying $p_{3D} \in [0.60, 0.85]$ and $p_{13D} \geq 0.80$, together with high-confidence non-swarm anchors satisfying $p_{13D} < 0.10$, are selected from the dataset. Thirty sequences of 80 frames are synthesized between paired anchors through a transition kernel $\alpha(t)$, using linear position interpolation and spherical velocity-direction interpolation with a linear fallback for collinear directions. These are synthetic observations, not dynamically integrated flight trajectories. For these swarm-starting sequences, the state starts at 1 and the buffer is prefilled with ten ones. Three operating conditions are considered:

A. stable–disturbance–stable: $\alpha(t) = 0$, with a GPS failure pulse of $\sigma = 2$ m introduced over the middle third of the sequence, testing dead-band suppression of transient disturbances;
B. gradual transition: $\alpha(t)$ follows a cosine ramp from 0 to 1 over the central half of the sequence, testing sliding-window smoothing of gradual changes;
C. abrupt disintegration: $\alpha(t)$ steps from 0 to 1 at the change point, with a 5% probability of sensor-flashback samples to emulate intermittent communication disturbance and test emergency-bypass response. Table 8 summarizes the metrics, and Figure 6 shows the corresponding posterior probabilities and decisions.

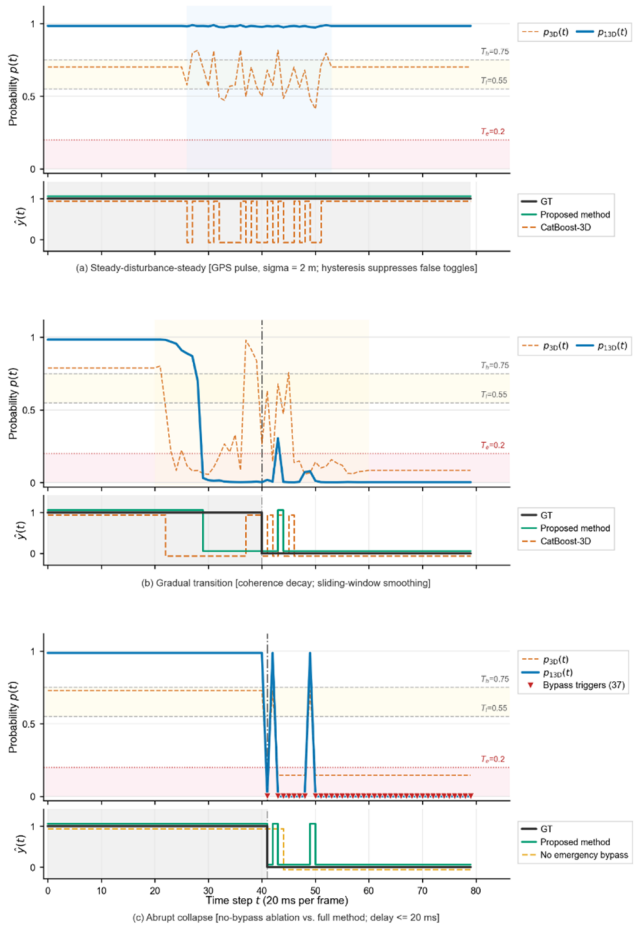


Figure 6. Probability–time waveform comparison of continuous trajectory simulation under three typical operating conditions. (**a**) Stable–disturbance–stable; (**b**) Gradual transition; (**c**) Abrupt disintegration.

Table 8. Continuous trajectory simulation end-to-end stability (N_{seq} = 30; T_{seq} = 80 frames; N = 200).

Condition	Metric	Comparator	Proposed
A: stable–disturb.–stable	False switches (mean±std)	8.07 ± 5.43	0.00 ± 0.00
A	Miss rate (clean segment)	0.00%	0.00%
B: gradual transition	DTW distance (mean±std)	4.60 ± 3.32	0.60 ± 1.08
B	Detection delay (frames)	2.27 ± 2.72	2.20 ± 3.49
C: abrupt disintegration	Detection delay (ms)	42.7 ± 6.8 (no bypass)	0.0 ± 0.0
C	Miss rate	5.33% (no bypass)	4.49%

Note: for Conditions A and B, the comparator is CatBoost-3D with a single threshold; for Condition C, it is the proposed 13D decision chain without the emergency bypass. In this simulation, the bypass changes detection delay from 42.7 ± 6.8 ms to 0.0 ± 0.0 ms under the discrete-time metric; by design, a qualifying $p(t) < T_{\text{emg}}$ event is handled within the current 20 ms sampling interval.

3.5. Robustness and Hyperparameter Sensitivity

For a 500-sample subset, independent Gaussian noise $\mathcal{N}(0, \sigma^2)$ is added to each UAV (20 Monte Carlo runs per setting), with thresholds fixed on the clean validation set, not re-tuned. Table 9 summarizes the accuracy of both methods under increasing noise levels.

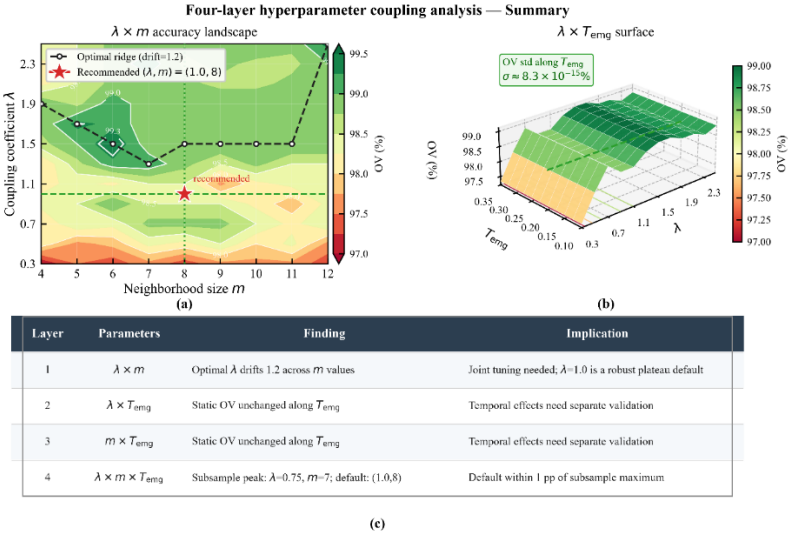
Table 9. GPS noise robustness (%).

Method	Clean	$\sigma = 0.1$ m	$\sigma = 0.3$ m	$\sigma = 0.5$ m	$\sigma = 1.0$ m
CatBoost-3D [3]	94.00	93.99	93.93	93.94	93.96
Proposed (13D)	99.20	99.20	99.20	99.26	99.34

Across the tested position-noise levels, the proposed method's mean accuracy remains between 99.20% and 99.34%. In the position-only experiment, the velocity-direction term $\bar{\omega}$ is unchanged by construction; this is not

a claim of immunity when velocity estimates are also corrupted. Under the saved combined perturbation result (GPS σ = 0.3 m and velocity noise σ_v = 0.5 m/s), accuracy is 99.32%. Other sensor-error distributions and stronger velocity perturbations require separate validation.

The 4000-sample grid summarized in Figure 7 evaluates λ, neighborhood size m, and T_{emg}. The accuracy-optimal λ shifts by 1.2 across m values, so λ and m should be tuned jointly. The subsample joint-grid maximum is at (λ, m) = (0.75, 7); the recommended (1.0, 8) lies within 1 pp and retains the baseline neighborhood size. Static overall accuracy is numerically unchanged along the tested T_{emg} axis, which shows that T_{emg} is decoupled from λ and m for this static metric only, not that it is dynamically irrelevant. The threshold pair (T_{high}, T_{low}) jointly defines the dead band, and (N_{win}, η) jointly defines the vote requirement; both pairs therefore require sequence-level validation rather than being assumed independent.



Layer	Parameters	Finding	Implication
1	$\lambda \times m$	Optimal λ drifts 1.2 across m values	Joint tuning needed; λ=1.0 is a robust plateau default
2	$\lambda \times T_{\text{emg}}$	Static OV unchanged along T_{emg}	Temporal effects need separate validation
3	$m \times T_{\text{emg}}$	Static OV unchanged along T_{emg}	Temporal effects need separate validation
4	$\lambda \times m \times T_{\text{emg}}$	Subsample peak: λ=0.75, m=7; default: (1.0,8)	Default within 1 pp of subsample maximum

(c)

Figure 7. Four-layer hyperparameter coupling analysis summary. (**a**) $\lambda \times m$ accuracy landscape; (**b**) $\lambda \times T_{\text{emg}}$ surface; (**c**) Summary table of coupling findings and implications.

3.6. Comparison with Representative Deep-Learning Baselines

To answer the practical comparison question, this paper evaluates three representative coordinate-input models: PointNet [32] (symmetric-function), DGCNN [33] (dynamic-graph), and an in-house PointNet+LSTM diagnostic baseline (snapshot–temporal fusion). Each frame is treated as a point cloud, and PointNet+LSTM is used only as a pseudo-sequence diagnostic because UCI samples are independent snapshots. All models use the same split and PC CPU environment. These are representative baselines, not an exhaustive evaluation of the deep-learning literature. Reference [34] concerns hierarchical temporal group-activity modeling and is cited as background, not as the source of this implemented baseline.

Recent architectures such as Point Transformer V3 (PTv3) [35] and rotation-aware variants such as RotInv-PCT [36] extend the design space, but they were not reimplemented under the present protocol. The following discussion is therefore limited to the coordinate-input representatives evaluated here and to mechanisms documented in the cited literature. Rotation-invariance surveys describe augmentation, equivariant layers, and pre-aligned reference frames as common design choices [37]; their relative accuracy, cost, and invariance guarantees depend on implementation.

(i) data augmentation with random rotations, which cannot guarantee strict invariance under continuous transformation groups [36];
(ii) specialized equivariant layers, which introduce non-trivial expressivity constraints and computational overhead;
(iii) pre-aligned local reference frames, which require additional preprocessing and can suffer from frame estimation noise. These are common design choices; specific architectures require separate evaluation.

Accordingly, the present results support a conditional conclusion: under the tested models and rigid-frame stress protocol, physically invariant features were less sensitive to coordinate changes than the evaluated raw-coordinate networks. This observation should not be read as a paradigm-wide ranking, and direct PTv3/RotInv-PCT experiments remain outside the current study.

Under ideal fixed coordinate conditions, deep learning and the proposed method both approach the ceiling, with accuracy differences not constituting a meaningful distinction. However, UAV swarms inevitably encounter

changing coordinate distributions when executing missions across different airspaces and base station coordinate systems. For this purpose, a coordinate frame invariance stress test is conducted: deep learning models are trained only on the original training set, and at test time, the same frame undergoes global translation $(x, y) \mapsto (x + 800, y - 500)$, global rotation of 60°, and their combination, with normalization parameters fixed to training set statistics. Table 10 reports results under both conditions.

Table 10. Comprehensive comparison with deep learning baselines (N = 200, UCI Swarm Behaviour).

Method	Clean Stress-Run acc. (%)	Trans. + rot. acc. (%)	Δ(pp)	Dynamic-Subset Recall (%)	Trans. + rot. Recall (%)	Params (MB)
PointNet [32]	100.00	82.10	−17.90	100.0	100.0	0.32
DGCNN [33]	100.00	95.14	−4.86	100.0	99.0	0.98
PointNet+LSTM (in-house)	100.00	—	—	100.0	—	3.30
Proposed (13D)	99.78	99.78	0.00	100.0	100.0	1.28

Table 10 reports frame-wise classification without temporal hysteresis; dynamic-subset recall uses the same $\bar{\omega} < 0.60$ rule as Section 3.3. Within this finite comparison, the matched stress-test runs show drops of 17.90 pp for PointNet and 4.86 pp for DGCNN under the combined translation and rotation, while the proposed features are unchanged to the reported precision (0.00 pp). These values describe only the evaluated models and transformation protocol.

The observed gap is consistent with a distribution-shift explanation: raw-coordinate models can depend on the coordinate distribution seen during training, while pairwise distances and velocity direction cosines are rigid-body invariants by construction. Other deep models may reduce this sensitivity through augmentation or equivariant design; because PTv3 and RotInv-PCT were not directly reproduced here, their performance and trade-offs cannot be inferred from Table 10.

4. Discussion

The three contributions form a sequential chain rather than a simple aggregation: SCDAM first supplies dynamic discrimination, the nine spectral descriptors enrich the SCDAM autocorrelation, and hysteresis maps the resulting posterior to a stable temporal decision. Ablation supports this hierarchy: reverting to the distance-only 3D baseline lowers accuracy by 6.2 pp, reflecting the combined extensions, removing all spectral descriptors causes a 1.0 pp drop, and removing the global velocity scalar causes a 1.6 pp drop. Hysteresis does not improve independent frame accuracy; its contribution appears in the temporal metrics of Section 3.4, where the full decision chain reduces switching and DTW distance.

The coordinate stress test provides evidence for the proposed invariant feature construction under the tested rigid translations and rotations. It does not establish superiority over all deep-learning architectures, nor does it replace direct comparisons with PTv3, RotInv-PCT, or other invariant/equivariant models. The practical implication supported by the present data is narrower: interpretable invariant features can offer a reproducible alternative for the evaluated assessment task, with an explicit temporal decision layer.

Within the present experiments, the framework offers the following observable properties:

(i) Coordinate robustness under the tested rigid-body transformations without retraining;
(ii) Feature-level interpretability, because each dimension is traceable to a defined physical quantity; and explicit decision-layer behavior, because dead-band, dwell, and emergency-bypass parameters can be inspected and measured. These properties are empirical and analytical results under the stated assumptions, not universal guarantees for every sensor stack or model.

Several limitations should be acknowledged. First, all experiments are offline: the UCI Swarm Behaviour dataset and the physically parameterized interpolation sequences do not include real GPS/IMU fusion bias, communication delays, actuator dynamics, or closed-loop flight. The Gaussian position-noise experiment tests one specified Gaussian-noise model and does not establish immunity to all onboard disturbances. Second, the 11.26 ms latency was measured in a PC Python environment; it indicates computational potential but does not prove onboard real-time performance on ARM flight controllers, which requires platform-specific profiling. Third, SCDAM is less sensitive to near-orthogonal than counter-directional motion, motivating trajectory prediction in future work. Finally, only PointNet, DGCNN, and PointNet+LSTM were directly evaluated; PTv3 and RotInv-PCT were discussed from the literature but not reproduced under the same protocol, so paradigm-level claims should be interpreted cautiously.

5. Conclusions

This paper presents a unified UAV-swarm regularity assessment framework in which SCDAM provides the core spatiotemporal representation, FFT-based descriptors complete its autocorrelation information, and dual-

threshold hysteresis stabilizes sequence-level decisions. On the UCI Swarm Behaviour dataset and physically parameterized offline trajectory/interpolation simulations, the framework improves static accuracy, challenge-scenario discrimination, temporal stability, and robustness under the tested coordinate and noise conditions. The 11.26 ms PC measurement and rigid-body stress test indicate computational and coordinate-robustness potential within this evaluation scope; they do not constitute onboard or real-flight validation. Future work will address airborne sensing and actuation, embedded implementation, 3D coordination, and closed-loop control.

Author Contributions

Conceptualization, P.Z. and R.Z.; methodology, P.Z. and R.Z.; software, R.Z.; validation, R.Z., V.G. and Y.F.; formal analysis, P.Z. and R.Z.; investigation, R.Z., V.G. and Y.F.; resources, H.L. and J.H.; data curation, R.Z.; writing—original draft preparation, P.Z. and R.Z.; writing—review and editing, V.G., Y.F. and J.H.; visualization, R.Z.; supervision, J.H.; project administration, J.H.; funding acquisition, J.H. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the Sichuan Provincial Science and Technology Program, grant number 2025YFRG0008.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The UCI Swarm Behaviour dataset used in this study is publicly available through the UCI Machine Learning Repository at https://archive.ics.uci.edu/dataset/524/swarm+behaviour (accessed on 24 September 2026).

Acknowledgments

The authors thank the editors and anonymous reviewers for their constructive comments.

Conflicts of Interest

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Use of AI and AI-Assisted Technologies

During manuscript revision, OpenAI Codex was used to assist with language editing, reference verification and formatting, and consistency checks against existing code and saved experimental results. The authors take full responsibility for the accuracy, integrity, and final content of the manuscript.

Abbreviations

UAV	Unmanned Aerial Vehicle
SCDAM	Spatiotemporal Coupled Dynamic Association Matrix
RPM	Relative Position Matrix
FFT	Fast Fourier Transform
TDA	Topological Data Analysis
GNN	Graph Neural Network
LSTM	Long Short-Term Memory
DTW	Dynamic Time Warping
GBDT	Gradient-Boosted Decision Tree
GPS	Global Positioning System
IMU	Inertial Measurement Unit

References

1. Bu, Y.; Yan, Y.; Yang, Y. Advancement Challenges in UAV Swarm Formation Control: A Comprehensive Review. *Drones* **2024**, *8*, 320. https://doi.org/10.3390/drones8070320.
2. Marek, D.; Biernacki, P.; Szyguła, J.; et al. Collision Avoidance Mechanism for Swarms of Drones. *Sensors* **2025**, *25*, 1141. https://doi.org/10.3390/s25041141.
3. Girgidov, R.A.; Fomichev, V.V. Assessing the Regularity of the Spatial Structure of a Group of Drones Based on Autocorrelation Analysis. *Mosc. Univ. Comput. Math. Cybern.* **2025**, *49*, 205–215. https://doi.org/10.3103/s0278641925700128.
4. Bogdanov, Y.V.; Boiko, A.M.; Girgidov, R.A. Application of variance analysis to the assessment of the structure of a swarm formed by unmanned aircrafts. In Proceedings of the All-Russian Scientific and Practical Conference of Young Scientists "Innovative Ideas in Mechanical Engineering", St. Petersburg, Russia, 24–26 May 2022; pp. 258–263.
5. Ramos, G.S.; Pinto, M.F.; Haddad, D.B. Advancing UAV swarm autonomy with ARCog-NET for task allocation, path planning, and formation control. *Robotica* **2025**, *43*, 2565–2609. https://doi.org/10.1017/s0263574725101914.
6. Moltajaei Farid, A.; Mei Kuan, L.; Kamal, M.A.S.; et al. Effective UAV patrolling for swarm of intruders with heterogeneous behavior. *Robotica* **2023**, *41*, 1673–1688. https://doi.org/10.1017/s0263574723000061.
7. Tan, L.; Shi, J.; Gao, J.; et al. Multi-UAV path planning based on IB-ABC with restricted planned arrival sequence. *Robotica* **2023**, *41*, 1244–1257. https://doi.org/10.1017/s0263574722001680.
8. Zhan, X.; Chen, Y.; Chen, X.; et al. Balanced Multi-UAV path planning for persistent monitoring. *Robotica* **2025**, *43*, 332–349. https://doi.org/10.1017/s0263574724001899.
9. UCI Machine Learning Repository. Swarm Behaviour [Dataset], 2020. Available online: https://archive.ics.uci.edu/dataset/524/swarm+behaviour (accessed on 24 September 2026).
10. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; et al. CatBoost: Unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2018; Volume 31. Available online: https://papers.nips.cc/paper_files/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html (accessed on 24 September 2026).
11. Zavlanos, M.M.; Egerstedt, M.B.; Pappas, G.J. Graph-theoretic connectivity control of mobile robot networks. *Proc. IEEE* **2011**, *99*, 1525–1540. https://doi.org/10.1109/jproc.2011.2157884.
12. Oh, K.K.; Park, M.C.; Ahn, H.S. A survey of multi-agent formation control. *Automatica* **2015**, *53*, 424–440. https://doi.org/10.1016/j.automatica.2014.10.022.
13. Reynolds, C.W. Flocks, herds and schools: A distributed behavioral model. In Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '87), Anaheim, CA, USA, 27–31 July 1987; pp. 25–34. https://doi.org/10.1145/37401.37406.
14. Vicsek, T.; Czirók, A.; Ben-Jacob, E.; et al. Novel Type of Phase Transition in a System of Self-Driven Particles. *Phys. Rev. Lett.* **1995**, *75*, 1226–1229. https://doi.org/10.1103/physrevlett.75.1226.
15. Vicsek, T.; Zafeiris, A. Collective motion. *Phys. Rep.* **2012**, *517*, 71–140. https://doi.org/10.1016/j.physrep.2012.03.004.
16. Girgidov, R.A. Swarm Simulation Modeling Using the Hadamard Product. *Mosc. Univ. Comput. Math. Cybern.* **2023**, *47*, 122–134. https://doi.org/10.3103/s0278641923030032.
17. Olfati-Saber, R.; Murray, R.M. Consensus Problems in Networks of Agents with Switching Topology and Time-Delays. *IEEE Trans. Autom. Control* **2004**, *49*, 1520–1533. https://doi.org/10.1109/tac.2004.834113.
18. Olfati-Saber, R. Flocking for Multi-Agent Dynamic Systems: Algorithms and Theory. *IEEE Trans. Autom. Control* **2006**, *51*, 401–420. https://doi.org/10.1109/tac.2005.864190.
19. Topaz, C.M.; Ziegelmeier, L.; Halverson, T. Topological Data Analysis of Biological Aggregation Models. *PLoS ONE* **2015**, *10*, e0126383. https://doi.org/10.1371/journal.pone.0126383.
20. Corcoran, P.; Jones, C.B. Modelling Topological Features of Swarm Behaviour in Space and Time With Persistence Landscapes. *IEEE Access* **2017**, *5*, 18534–18544. https://doi.org/10.1109/access.2017.2749319.
21. Bhaskar, D.; Manhart, A.; Milzman, J.; et al. Analyzing collective motion with machine learning and topology. *Chaos Interdiscip. J. Nonlinear Sci.* **2019**, *29*, 123125. https://doi.org/10.1063/1.5125493.
22. Sathiyakumar, T.; Sur, S.; Mondal, S.; et al. Detecting transitions between collective motion regimes using functional hypothesis test of the time-varying persistence homology. *arXiv* **2024**, arXiv:2403.03428.
23. Wiener, N. Generalized harmonic analysis. *Acta Math.* **1930**, *55*, 117–258. https://doi.org/10.1007/bf02546511.
24. Cooley, J.W.; Tukey, J.W. An algorithm for the machine calculation of complex Fourier series. *Math. Comput.* **1965**, *19*, 297–301. https://doi.org/10.1090/s0025-5718-1965-0178586-1.
25. Oppenheim, A.V.; Schafer, R.W. *Discrete-Time Signal Processing*, 3rd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2009.

26. Levant, A. Higher-order sliding modes, differentiation and output-feedback control. *Int. J. Control* **2003**, *76*, 924–941. https://doi.org/10.1080/0020717031000099029.
27. Utkin, V.I. *Sliding Modes in Control and Optimization*; Springer: Berlin/Heidelberg, Germany, 1992.
28. Isermann, R. *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*; Springer: Berlin/Heidelberg, Germany, 2006.
29. Sun, Y.; Su, H.; Wang, X. Semiglobal Observer-Based Positive Scaled Edge-Consensus of Networked Discrete-Time Systems Under Actuator Saturation. *IEEE Trans. Syst. Man Cybern. Syst.* **2021**, *51*, 4543–4554. https://doi.org/10.1109/tsmc.2019.2944976.
30. Sun, Y.; Su, H.; Zeng, Z. *H*∞ Control for Observer-Based Non-Negative Edge Consensus of Discrete-Time Networked Systems. *IEEE Trans. Cybern.* **2022**, *52*, 2351–2360. https://doi.org/10.1109/tcyb.2020.3003279.
31. Sun, Y.; Yang, X.; Shi, P.; Su, H. Consensus Tracking of Switched Heterogeneous Nonlinear Systems With Uncertain Target. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2024**, *71*, 5560–5570. https://doi.org/10.1109/tcsi.2024.3376531.
32. Qi, C.R.; Su, H.; Mo, K.; et al. PointNet: Deep learning on point sets for 3D classification and segmentation. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 77–85. https://doi.org/10.1109/cvpr.2017.16.
33. Wang, Y.; Sun, Y.; Liu, Z.; et al. Dynamic Graph CNN for Learning on Point Clouds. *ACM Trans. Graph.* **2019**, *38*, 1–12. https://doi.org/10.1145/3326362.
34. Ibrahim, M.S.; Muralidharan, S.; Deng, Z.; et al. A Hierarchical Deep Temporal Model for Group Activity Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 1971–1980. https://doi.org/10.1109/cvpr.2016.217.
35. Wu, X.; Jiang, L.; Wang, P.S.; et al. Point Transformer V3: Simpler, Faster, Stronger. In Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–22 June 2024; pp. 4840–4851. https://doi.org/10.1109/cvpr52733.2024.00463.
36. He, C.; Zhao, Z.; Zhang, X.; et al. RotInv-PCT: Rotation-Invariant Point Cloud Transformer via feature separation and aggregation. *Neural Netw.* **2025**, *185*, 107223. https://doi.org/10.1016/j.neunet.2025.107223.
37. Fei, J.; Deng, Z. Rotation invariance and equivariance in 3D deep learning: A survey. *Artif. Intell. Rev.* **2024**, *57*, 168. https://doi.org/10.1007/s10462-024-10741-2.
38. Zhao, B.; Huo, M.; Li, Z.; et al. Graph-based multi-agent reinforcement learning for large-scale UAVs swarm system control. *Aerosp. Sci. Technol.* **2024**, *150*, 109166. https://doi.org/10.1016/j.ast.2024.109166.
39. Tolstaya, E.; Gama, F.; Paulos, J.; et al. Learning Decentralized Controllers for Robot Swarms with Graph Neural Networks. In Proceedings of the Conference on Robot Learning, Osaka, Japan, 30 October–1 November 2019; PMLR, 2020; Volume 100, pp. 671–682. Available online: https://proceedings.mlr.press/v100/tolstaya20a.html (accessed on 24 September 2026).
40. McElfresh, D.; Khandagale, S.; Valverde, J.; et al. When Do Neural Nets Outperform Boosted Trees on Tabular Data? In Proceedings of the Advances in Neural Information Processing Systems, New Orleans, LA, USA, 10–16 December 2023; Volume 36, pp. 76336–76369. https://doi.org/10.52202/075280-3337.
41. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. https://doi.org/10.1145/2939672.2939785.
42. Ke, G.; Meng, Q.; Finley, T.; et al. LightGBM: A highly efficient gradient boosting decision tree. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30. Available online: https://papers.nips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html (accessed on 24 September 2026).
43. Zhou, X.; Zhu, J.; Zhou, H.; et al. EGO-Swarm: A fully autonomous and decentralized quadrotor swarm system in cluttered environments. In Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 30 May–5 June 2021; pp. 4101–4107. https://doi.org/10.1109/icra48506.2021.9561902.
44. Li, W.; Chen, X.; Cheng, L.; et al. Dual-channel event-triggered consensus of multi-agent systems under DoS attacks over switching topologies. *Complex Eng. Syst.* **2025**, *5*, 5. https://doi.org/10.20517/ces.2024.104.
45. Abpeikar, S.; Kasmarik, K. Motion behaviour recognition dataset collected from human perception of collective motion behaviour. *Data Brief* **2023**, *47*, 108976. https://doi.org/10.1016/j.dib.2023.108976.
46. Walker, P.; Lewis, M.; Sycara, K. Characterizing human perception of emergent swarm behaviors. In Proceedings of the 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary, 9–12 October 2016; pp. 002436–002441. https://doi.org/10.1109/smc.2016.7844604.