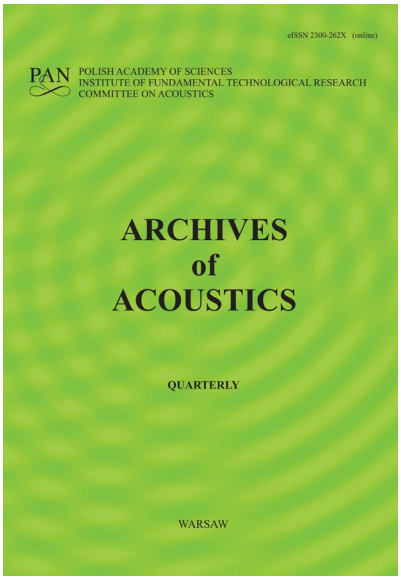
JOURNAL PRE-PROOF

This is an early version of the article, published prior to copyediting, typesetting, and editorial correction. The manuscript has been accepted for publication and is now available online to ensure early dissemination, author visibility, and citation tracking prior to the formal issue publication.

It has not undergone final language verification, formatting, or technical editing by the journal's editorial team. Content is subject to change in the final Version of Record.

To differentiate this version, it is marked as "PRE-PROOF PUBLICATION" and should be cited with the provided DOI. A visible watermark on each page indicates its preliminary status.

The final version will appear in a regular issue of *Archives of Acoustics*, with final metadata, layout, and pagination.



Title: Riemannian Geometric Characterization of Underwater Biological Signals via Fisher-Optimized SPD Manifolds

Author(s): Lu Cao, Yang Chen, Zhaochen Shen, Qiang Fang, Yun Gao, Shipeng Yu, Shenqiu Han, Binqiang Sun, Jian Qiu, Zhongyan Huo

DOI: https://doi.org/10.24423/archacoust.2026.4499

Journal: *Archives of Acoustics*

ISSN: 0137-5075, e-ISSN: 2300-262X

Publication status: In press

Received: 2026-04-28

Accepted: 2026-06-08

Published pre-proof: 2026-06-11

Please cite this article as:

Cao L., Chen Y., Shen Z., Fang Q., Gao Y., Yu S., Han S., Sun B., Qiu J., Huo Z. (2026), Riemannian Geometric Characterization of Underwater Biological Signals via Fisher-Optimized SPD Manifolds, *Archives of Acoustics*, https://doi.org/10.24423/archacoust.2026.4499

Riemannian Geometric Characterization of Underwater Biological Signals via Fisher-Optimized SPD Manifolds

Lu Cao[1], Yang Chen[2], Zhaochen Shen[3], Qiang Fang[4], Yun Gao[5], Shipeng Yu[6], Shenqiu Han[7], Binqiang Sun[8], Jian Qiu[9,*] Zhongyan Huo [10,*]

[1] School of Marine Engineering and Equipment, Zhejiang Ocean University, Zhoushan, China, https://orcid.org/0009-0003-2841-5339

[2] School of Marine Engineering and Equipment, Zhejiang Ocean University, Zhoushan, China, https://orcid.org/0009-0001-5186-8454

[3] School of Marine Engineering and Equipment, Zhejiang Ocean University, Zhoushan, China, https://orcid.org/0009-0001-9874-5267

[4] School of Marine Engineering and Equipment, Zhejiang Ocean University, Zhoushan, China, https://orcid.org/0000-0002-4569-9893

[5] School of Marine Engineering and Equipment, Zhejiang Ocean University, Zhoushan, China, https://orcid.org/0000-0003-1549-1028

[6] Zhejiang Guangchuan Engineering Project Management Co., Ltd., Hangzhou, China, https://orcid.org/0009-0003-1242-0846

[7] Zhejiang Guangchuan Engineering Consulting Co., Ltd., Hangzhou, China, https://orcid.org/0009-0001-2853-9353

[8] Binjiang Institute, Zhejiang University, Hangzhou, China, https://orcid.org/0009-0008-0107-0701

[9] Zhejiang Guangchuan Engineering Consulting Co., Ltd., Hangzhou, China, https://orcid.org/0009-0003-6005-668X

[10] School of Marine Engineering and Equipment, Zhejiang Ocean University, Zhoushan, China, https://orcid.org/0009-0007-7297-4356

*Corresponding Author e-mail: huozhongyan@zjou.edu.cn

Abstract

Automated fish welfare monitoring in intensive aquaculture is hindered by environmental noise, individual variability, and data scarcity. These challenges have not been fully resolved by existing deep learning approaches. Traditional computer-vision methods are constrained by

underwater turbidity, whereas Passive Acoustic Monitoring (PAM) offers a promising non-invasive alternative for assessing aquatic environments. This study proposes Fisher-SPD, a lightweight, geometry-aware framework for classifying the acoustic behaviour of cage-farmed *Larimichthys crocea*. A Fisher-score mechanism adaptively selects discriminative frequency bands, effectively filtering complex broadband noise commonly found in commercial sea cages. Acoustic segments are modelled as Symmetric Positive Definite (SPD) covariance matrices and mapped onto a linear tangent space via the Log-Euclidean Metric, preserving intrinsic statistical structure even under data-scarce conditions. Furthermore, a Physics-Consistency Masking mechanism applies source-level physical priors as hard inference-time constraints to robustly suppress false positives originating from background interference. Under a strict Leave-One-Subject-Out Cross-Validation (LOSO-CV) protocol, the framework achieved a mean zero-shot accuracy of 89.22% $\pm$ 2.59%, significantly outperforming ResNet-18 and other deep learning baselines, while maintaining a low inference latency of only 3.60 ms on edge devices. Through few-shot domain calibration, the system achieved 92.86% accuracy in a dual-fish overlapping-source environment. Ultimately, this framework provides a robust, data-efficient solution for real-time stress detection and welfare monitoring in modern intensive aquaculture.

Keywords: Passive acoustic monitoring, *Larimichthys crocea*, Riemannian manifold, Few-shot learning, Aquaculture bioacoustics.

Acronyms

PAM – Passive Acoustic Monitoring,
SPD – Symmetric Positive Definite,
LOSO-CV – Leave-One-Subject-Out Cross-Validation
STFT – Short-Time Fourier Transform

1. Introduction

As the flagship species of China's marine aquaculture industry, the cage farming of *Larimichthys crocea* is transitioning toward a precise and intelligent paradigm (FAO, 2024;

Føre *et al*., 2018). Real-time health monitoring is essential for production efficiency and fish welfare. Because chronic environmental stressors manifest directly as abnormal kinematics, behavioural alterations serve as critical indicators for health assessment (Martins *et al*., 2012; Stien *et al*., 2013; Huntingford *et al*., 2006). As intensive farming exacerbates these stressors, developing automated, real-time early warning systems has become an urgent priority in applied precision aquaculture (Sueur, Farina, 2015).

Automated monitoring of underwater fish behaviour has long been constrained by physical limitations. Widely applied optical vision systems are severely restricted by turbidity, light attenuation, and biological occlusion in high-density cages (Li *et al*., 2025; Saberioon *et al*., 2017; Salman *et al*., 2020). By contrast, Passive Acoustic Monitoring (PAM) exploits the low attenuation of acoustic waves to penetrate these optical blind spots (Slabbekoorn *et al*., 2010; Rountree *et al*., 2006; Mooney *et al*., 2020). Larimichthys crocea produces species-specific signals including swimbladder muscle pulses for communication and feeding (Vieira *et al*., 2015; Malfante *et al*., 2018; Ramcharitar *et al*., 2006) and passive hydrodynamic noise during rapid swimming, turning, or startle-escape manoeuvres (Noda *et al*., 2016; Domenici, Hale, 2019). Analyzing vocal responses to stressors is crucial for non-invasive fish welfare assessment (Popper, Hawkins, 2019). However, while standard soundscape toolboxes and acoustic indices excel at broad biodiversity monitoring (Ulloa *et al*., 2021; Towsey *et al*., 2014), they fail to isolate subtle behavioral cues from the severe, non-stationary noise of commercial cage farms. Furthermore, the highly complex underwater acoustic environment, often characterized by strong multipath propagation and severe signal attenuation, continuously poses significant challenges for robust signal extraction and system performance (Schmidt, Kochańska, Schmidt, 2024; Wang, Yang, 2025).

To address this, recent advances in deep learning have transformed underwater acoustic processing. Specifically, lightweight attention networks and multi-algorithm feature extraction models now enable highly robust target recognition in complex marine environments (Liu *et al*., 2024; Wang *et al*., 2024; Li *et al*., 2022). However, the deployment of PAM and deep learning in field settings faces two major challenges: (i) the pronounced individual acoustic variability and scarcity of labelled high-risk behaviour samples cause severe overfitting in data-hungry models (Luo *et al*., 2023; Stowell, 2022; Doan *et al*., 2022); and (ii) the

traditional flat Euclidean space assumption fails to capture complex nonlinear acoustic topologies.

To address both limitations, Riemannian manifold geometry offers a transformative perspective. The second-order statistics of bioacoustic signals are naturally characterised by Symmetric Positive Definite (SPD) covariance matrices on a Riemannian manifold (Pennec *et al*., 2006; Barachant *et al*., 2011), preserving intrinsic nonlinear geometry through affine-invariant metrics and naturally suppressing non-Gaussian noise. Manifold-based metric learning achieves high-precision classification with extremely few samples (Shang *et al*., 2024; Nolasco *et al*., 2023; Chauhan *et al*., 2022). Applying SPD manifold geometry to *Larimichthys crocea* acoustic signals thus corrects the geometric mismatch of traditional Euclidean models and fundamentally resolves performance bottlenecks caused by data scarcity.

Although Riemannian geometry mitigates feature difficulties, purely geometry-driven models may still violate physical laws (Daw *et al*., 2022; Wang, Yu, 2025; Willard *et al*., 2022). In current state-of-the-art (Li *et al*., 2025; Chen *et al*., 2025), physical rules act only as soft training constraints, leaving models susceptible to physically implausible false positives under sudden noise. This study therefore proposes a lightweight Fisher physics-aware mechanism built on the SPD manifold, elevating physical constraints to hard decision-boundary gates enforced at inference.

This study proposes a Riemannian geometry-based acoustic recognition framework for *Larimichthys crocea* behaviour, compensating for sample scarcity through geometric constraints. The main contributions are:

utilising SPD covariance matrices to compactly represent full-spectrum acoustic dynamics—including vocalizations and kinematic friction noise.

the nonlinear classification problem is effectively linearised within a low-dimensional tangent space

Physics-Consistency Masking Mechanism: A novel physics-aware decision mechanism is introduced to address non-stationary noise (Chen *et al*., 2025). It integrates physical prior constraints into the classification layer as hard thresholds, effectively suppressing high-confidence false positive detections that violate physical acoustic plausibility.

Collectively, the framework operationalises a deciphering-plus-alerting paradigm, bridging fine-grained behaviour classification with reliable binary welfare alarms, and demonstrates that geometric modelling outperforms purely data-driven methods under bioacoustic data scarcity (Sueur, Farina, 2015).

2. Experimental Methods and Data Processing

2.1.Experimental Setup and Data Acquisition

2.1.1. Experimental Environment

To isolate weak bioacoustic signals, experiments were conducted in a controlled indoor glass tank (84 cm × 30 cm × 50 cm; water depth: 36.5 cm) equipped with three-tier physical isolation. First, the tank was anchored to a custom shock-absorbing table to decouple low-frequency structural vibrations. Second, to prevent photic stress, experiments were performed in complete darkness; continuous, non-intrusive behavioral observation was achieved using an external infrared camera and lighting system. Third, water quality was maintained using an acoustically dampened, low-power filtration and aeration device, strictly minimizing background hydrodynamic noise that could otherwise mask weak biological acoustic signatures.

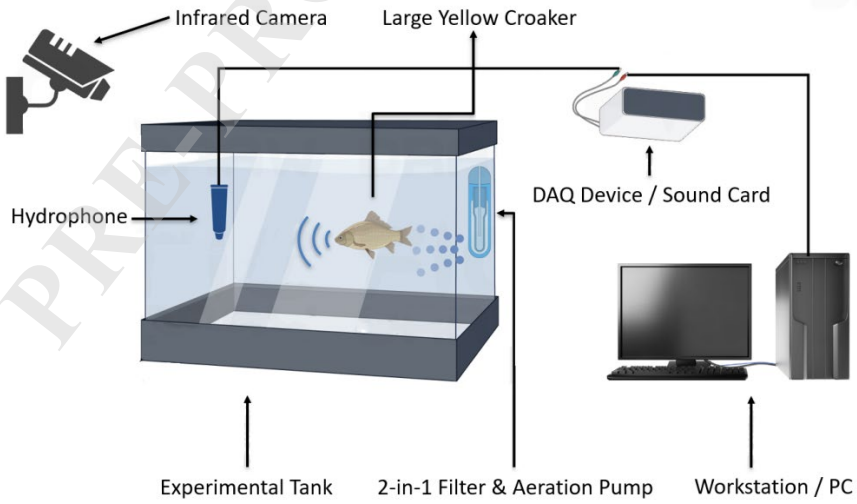


Fig. 1 The controlled single-individual protocol provides clean acoustic ground truth for method development; field-scale PAM deployment scenarios are discussed in Section 4.2.

2.1.2. Experimental Subjects

The experimental subjects were 8 sub - adult wild *Larimichthys crocea*, belonging to the class Osteichthyes, order Perciformes, family *Sciaenidae, genus Larimichthys*. To minimize acoustic discrepancies caused by individual differences, specimens of similar size were selected (body length: 20 ± 0.5 cm, weight: 200 ± 10 g). Randomly selected subjects were introduced into the experimental tank and given a 24-hour quiet acclimation period to eliminate initial stress responses induced by transportation. To ensure the objectivity and biological validity of the collected data, a strict observation protocol was established. An independent recording protocol was adopted in which each fish was placed in the tank individually for behavioural recording, entirely eliminating the risk of multi-source acoustic mixing. Between subject rotations, newly introduced fish were allowed to acclimate for more than 12 hours to ensure stable behavioural states. This design guarantees the statistical independence of the subsequent Leave-One-Subject-Out Cross-Validation, preventing information leakage between training and test sets.

In total, 576 three-second acoustic segments were collected across the four behavioural categories, yielding approximately 72 labelled samples per subject.

2.1.3. Experimental Protocol

A standardised protocol combining naturalistic observation and controlled stimulus induction was developed for four core behaviours: slow roaming, fast tail-beating, turning, and startle-escape, across two phases:

Natural Kinematic Observation Phase: Natural hydrodynamic noise was recorded under controlled dark conditions without external pressure. Fish exhibited slow roaming, spontaneous fast tail-beating during exploration, and turning near tank boundaries.

Forced Stress Induction Phase: Silent, non-contact stimuli were applied at randomised intervals (< 5 s each) to induce startle-escape responses, generating transient high-energy broadband signals that served as challenging training examples for high-risk state recognition.

Laboratory Scope: Single-individual tank experiments ensured unambiguous acoustic annotation. This differs substantially from field conditions—multi-fish overlapping signals, variable aeration and vessel noise, environmental variability—but represents a validated methodological foundation; translation to multi-source environments is discussed in Section 4.

2.2. Audio-Visual Acquisition System

A high-fidelity audio-visual acquisition system achieved accurate acoustic signal capture and reliable behavioural label alignment. Key hardware components are as follows:

Acoustic Perception Terminal: A B&K Type 8105 hydrophone (sensitivity: −185 dB re 1 V/μPa; flat response 20 Hz–65.536 kHz) was suspended at the tank centre (20 cm depth) for distortion-free capture of both swimbladder pulses and broadband hydrodynamic noise.

Visual Perception Terminal: An industrial-grade HD camera (Model: WZ800SG-Y52-F4-6; 4K/8MP; 50 Hz anti-flicker) was mounted obliquely above the tank, capturing instantaneous postural changes even under water-surface ripple interference.

Signal Conditioning and A/D Conversion: A B&K 3161-A-011 DAQ card (built-in low-noise preamplifier and anti-aliasing filter; 24-bit precision; 65.536 kHz sampling rate) amplifies biological signals losslessly, eliminates quantisation truncation errors, and fully covers all effective harmonic bands of *Larimichthys crocea*.

Table 1. Data Acquisition System Parameters.

Equipment	Model	Key Parameters
Hydrophone	Brüel & Kjær 8105	Sensitivity: -185 dB re 1 V / μPa; Frequency Response: 20 Hz – 65.536 kHz
HD Camera	WZ 800 SG-Y 52 - F4 - 6	Resolution: 4 K / 8 MP Anti-flicker: 50 Hz
DAQ Card	Brüel & Kjær 3161 - A - 011	Sampling Rate: 65.536 kHz; Quantization Precision: 24 bit; Channels: Single channel

Multimodal Synchronisation: Hardware-level timestamp synchronisation provided unambiguous visual ground truth for behaviour-acoustic alignment, resolving the source-

identification problem inherent in acoustic-only recording.

2.3. Multimodal Data Processing

To ensure exact correspondence between acoustic signals and specific behavioural states, pure background noise was also recorded in a fish-free environment to establish an acoustic baseline.

2.3.1. Behaviour Annotation

Background noise was recorded pre-experiment to establish the acoustic baseline for SNR evaluation. Strict frame-to-sample synchronisation ensured one-to-one visual-acoustic alignment. An independent-event partitioning scheme assigned all sub-samples from the same continuous segment to the same split, rigorously preventing temporal data leakage.

To avoid the false positives and missed detections of traditional energy-detection algorithms, a vision-guided manual annotation protocol was adopted:

Behaviour Localisation: Synchronized videos were reviewed, and trained annotators precisely marked the start and end times of the four target behaviours (fast roaming, slow roaming, turning, and startle-escape).

Audio Slicing: Based on verified timestamps from the video, the corresponding valid segments were extracted from the continuous audio stream. Fixed-length Segmentation: Given that the typical duration of most behaviours is generally under 3 seconds, all segments were standardised into 3-second samples. Segments shorter than 3 seconds were extended via zero-padding or cyclic-padding; segments longer than 3 seconds were divided into multiple non-overlapping 3-second samples.

2.3.2. Signal Preprocessing

Raw acoustic segments were processed through the standardised pipeline summarised in Table 2.

Table 2. Preprocessing Parameter Settings

Parameter	Setting Value	Description
Sampling Rate	65.536 kHz	Original sampling rate, no resampling required
Amplitude Normalization	Max value normalization	Eliminates sound pressure level differences caused by recording distance
STFT Window Length	4096 points	Frequency resolution ≈ 16 Hz
Hop Size	1024 points	Time resolution ≈ 15.6 ms
Effective Frequency Band	100 Hz – 30 kHz	Filters out low-frequency hydrodynamic noise and high-frequency aliasing
Sliding Window	64 frames/window, step 32 frames	Approx. 1s duration, 50% overlap

The preprocessing pipeline is as follows:

1. Apply maximum amplitude normalization to the signal.

2. Apply Short-Time Fourier Transform (STFT) to convert the 1D time-domain signal into a 2D time-frequency representation.

3. Retain the 100 Hz to 30 kHz frequency band to filter out out-of-band noise.

4. Apply an overlapping sliding window to segment the continuous spectrogram into fixed-size samples.

Each 64-frame window (~1 s) satisfies the full-rank covariance requirement (frames > bins). A 32-frame step (50% overlap) achieves data augmentation while preserving temporal continuity. Each preprocessed sample is a $T \times F$ energy matrix (T = 64, F = filtered bins). PCEN dynamically suppressed steady-state background noise; MFCCs captured the transient startle acoustic envelope.

2.3.3. Behaviour Categories

Four behavioural states were recorded: fast tail-beating, slow tail-beating, turning, and startle-escape. Fig. 3 displays representative spectrograms for each class, corresponds exactly to Fig.

2. These behaviours span the most relevant kinematic patterns in aquaculture monitoring, and their acoustic differences form the basis for subsequent classification.

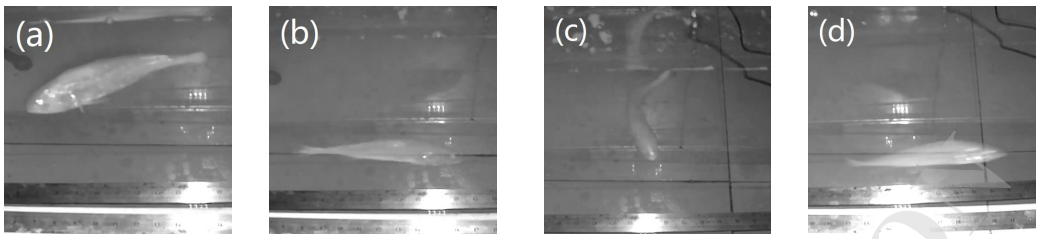
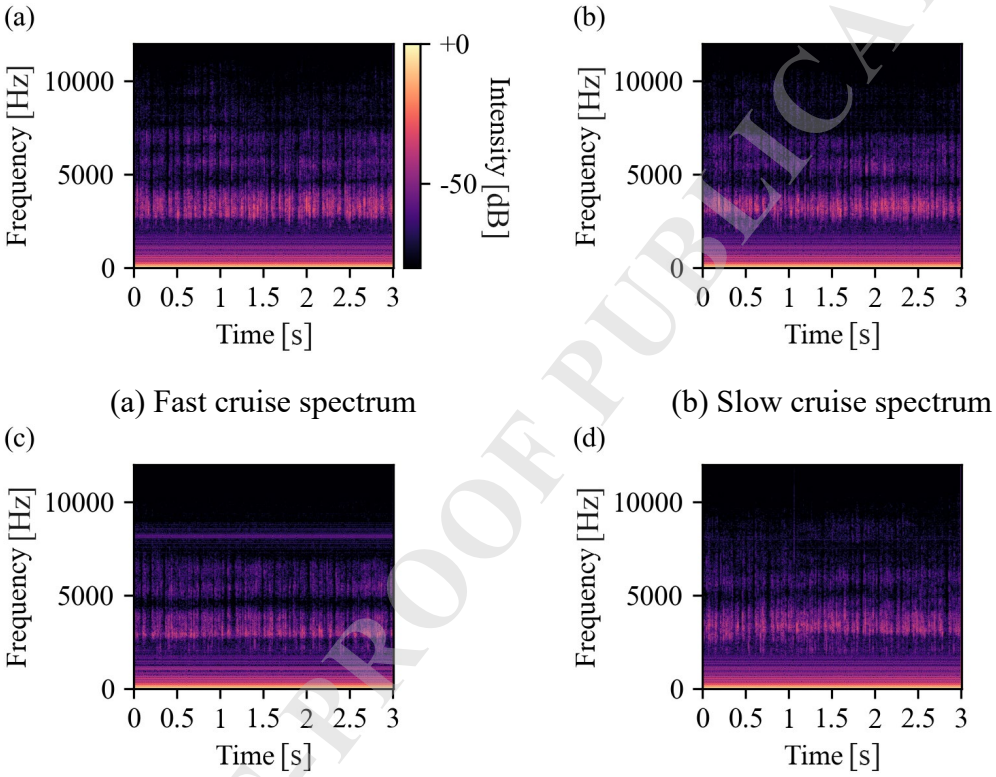


Fig. 2 Original Image of Fish Behaviour



(a) Fast cruise spectrum (b) Slow cruise spectrum

(c) Turn spectrum (d) Fright escape spectrum

Fig. 3 Spectrograms of various behaviours

2.4. Intrinsic Geometric Modeling and Classification on Riemannian Manifolds

The Fisher-SPD framework departs from the traditional Euclidean feature-stacking paradigm, using Riemannian geometry to capture intrinsic signal structure (Fig. 4). This section derives the pipeline from frequency band selection to tangent space projection.

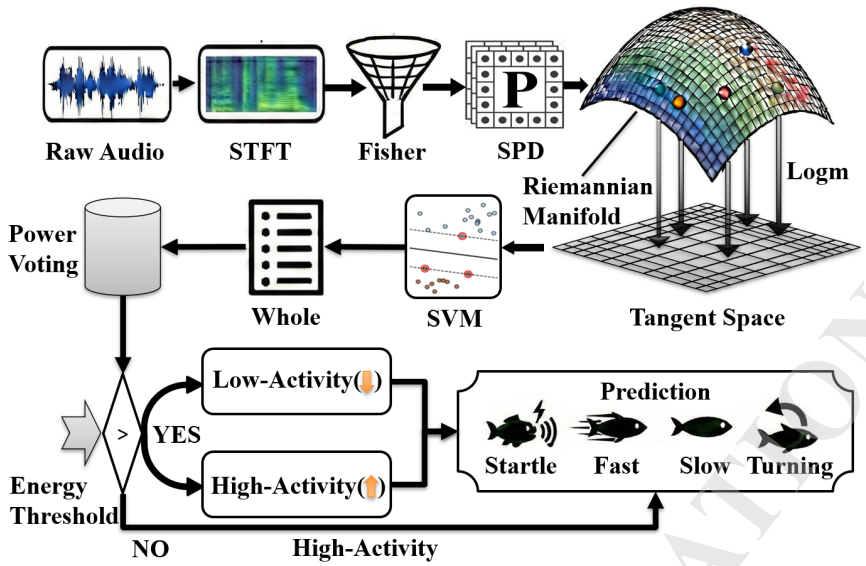


Fig. 4 Overall algorithm framework diagram

2.4.1. Fisher-based Frequency Selection

Full-band analysis introduces substantial redundant noise. The Fisher discriminant criterion isolates the most informative feature subspace prior to covariance construction.

Let the *i*-th preprocessed sample be:

$$x_i \in R^{T\times F} \tag{1}$$

Where T is the number of frames and F is the number of frequency bins. For each frequency bin $f\in\{1\ldots F\}$, its Fisher score $J(f)$ is computed to measure the discriminative power of that frequency across behaviour classes:

$$J(f)=\frac{S_B(f)}{S_W(f)}=\frac{\sum_{c=1}^{C} n_c(\mu_{c,f}-\mu_f)^2}{\sum_{c=1}^{C}\sum_{x\in D_c}(x_f-\mu_{c,f})^2} \tag{2}$$

Where S_B and S_W are the between-class and within-class scatter; $\mu_{c,f}$ and μ_f are the class-conditional and global mean at frequency f. Ranked by descending $J(f)$, the top K = 50 bands are selected (determined by cross-validation; Section 3.5), reducing dimensionality from F = 513 to K = 50 and removing noise-dominated bands. The Fisher criterion is mathematically equivalent to extracting swimbladder resonance formants.

By retaining $K = 50$ bands corresponding physically to the swimbladder resonance formants and kinematic friction signatures of *Larimichthys crocea*, while discarding the broadband components where environmental noise (aeration, flow turbulence, vessel engines) concentrates, Fisher selection provides inherent upstream noise suppression, complementing the noise resilience demonstrated in Section 3.6.

2.4.2. SPD Descriptor Construction

To capture both static spectral correlations and dynamic temporal evolution, augmented feature vectors are constructed before computing covariance. For the filtered feature vector $X_t \in R^K$ at frame t, the first-order temporal difference is:

$$\Delta X_t = X_{t+1} - X_t \tag{3}$$

The original and difference features are then concatenated to obtain an augmented feature vector:

$$Z_t = [X_t^T, \Delta X_t^T]^T \in R^{2K} \tag{4}$$

Xt captures instantaneous spectral energy (static feature)*; ΔX_t* encodes spectral dynamics (dynamic feature). Joint modelling allows *P* to represent both signal identity and temporal evolution—critical for separating behaviours with similar spectra but different temporal patterns (e.g., fast roaming vs. slow roaming).

Each 1-second acoustic sample is then modelled as a *2K×2K* sample covariance matrix *P*:

$$P = \frac{1}{T-1}\sum_{t=1}^{T}(Z_t - \bar{Z})(Z_t - \bar{Z})^T + \varepsilon I \tag{5}$$

Where $\bar{Z}$ is the sample mean and $\varepsilon = 10^{-6}$ is a regularisation term enforcing positive definiteness. *P* is a point on the curved SPD Riemannian manifold *S2K ++*.

2.4.3. Riemannian Tangent Space Projection and Geometric Dimensionality Reduction

Due to manifold curvature, directly applying a conventional SVM yields suboptimal results.

The Log-Euclidean Metric therefore projects manifold data into the tangent space.

First, the Riemannian geometric mean $\bar{P}$ of all training covariance matrices is computed as the tangent point:

$$\overline{P} = \exp(\frac{1}{N}\sum_{i=1}^{N}\log(P_i)) \tag{6}$$

Next, the logarithmic map projects each sample P_i into the tangent space $T_{\bar{P}}S2K++$ centered at $\overline{P}$, generating the tangent vector S_i:

$$S_i = \log_{\bar{P}}(P_i) = \overline{P}^{\frac{1}{2}}\log(\overline{P}^{-\frac{1}{2}}P_i\overline{P}^{-\frac{1}{2}})\overline{P}^{\frac{1}{2}} \tag{7}$$

A whitening transformation eliminates tangency-point influence. Let $S_i = log(\ \overline{P}^{-0.5}P_i\ \overline{P}^{-0.5})$ be the whitened symmetric matrix. The final feature vector V_i is defined as:

$$V_i = vect(\overline{S}_i) \in R^d, d = 2K(2K+1)/2 \tag{8}$$

Where $V_i(\cdot)$ is a half-vectorisation operator extracting upper-triangular elements (including the diagonal) column-wise. For K = 50, d = 100×101/2 = 5050. These transformations linearise the nonlinear manifold structure within Euclidean space while retaining intrinsic frequency coupling, laying the geometric foundation for high-precision acoustic behaviour classification.

2.4.4. Classification via Tangent Space SVM

Rather than relying solely on a data-driven classifier, this section describes a cascaded decision mechanism integrating manifold geometric properties with bioacoustic physical priors.

Once mapped to tangent space, classification becomes linearly separable. An RBF-kernel SVM (random seed 96965; cost weights inversely proportional to class frequencies) maximises the inter-class margin and outputs a posterior probability vector p_i over behaviour categories:

$$P_t = [P_{t,1,\ldots,} P_{t,c}]^T \tag{9}$$

2.4.5. Power Voting Mechanism

Behavioural events span multiple analysis frames, requiring robust aggregation of frame-level predictions. Conventional majority voting and average pooling are susceptible to transient noise and ambiguous frames.

A power voting mechanism introduces a power exponent $\gamma = 2$ that nonlinearly amplifies high-confidence frames while suppressing high-entropy frames. The aggregated score S_c for category c is:

$$S_c = \sum_{t=1}^{N_{\text{frames}}} (P_{t,c})^{\gamma} \tag{10}$$

With $\gamma > 1$, the model automatically focuses on segments where behavioural features are most salient, enhancing decision robustness against outliers.

2.4.6. Physics-Consistency Masking

High-intensity stress behaviours are inevitably accompanied by acoustic energy substantially exceeding the environmental baseline, whereas passive behaviours are typically obscured by background noise. A physics-consistency masking mechanism translates this source-level prior into a hard decision-boundary constraint, preventing physically implausible false positive detections.

Let $P=R^C$ be the raw posterior probability vector output by the SVM, where C is the number of categories. Unlike conventional weighted adjustments, an energy-based binary mask vector $M(E_{file}) \in \{0, 1\}^c$ is defined and applied to the decision process via the Hadamard product ($\odot$):

$$\hat{y} = \arg\max_{c \in C} (P \odot M(E_{\text{file}})) \tag{11}$$

The c-th component of the mask vector M is defined by an indicator function $\mathit{II}(\odot)$:

$$M_c(E_{file}) = \begin{cases} 0, & if\ c = c_{startle}\ and\ E_{file} < \mu_E \\ 1, & otherwise \end{cases} \tag{12}$$

Where μ_E is the energy gating threshold, set as the robust boundary separating the system from the environmental noise baseline (validated by the sensitivity sweep in Fig. 9(d)). The gating logic is:

When $E_{file} < \mu_E$, a high source-level behaviour is physically excluded. The mask forces the startle-escape probability to zero regardless of classifier confidence, fundamentally eliminating false positive detections under low SNR.

When $E_{file} \geq \mu_E$, the mask is all ones and the model retains full multi-class discriminatory capability.

In the constrained tank environment, source-to-receiver distances range from 5 cm adjacent to the hydrophone to 76.37 cm at the farthest boundary. The maximum acoustic attenuation is calculated using the spherical spreading loss model:

$$TL = 20\log_{10}(\frac{r}{r_0}) \tag{13}$$

The resulting 24 dB positional variation is well within the 140 dB dynamic range of the 24-bit DAQ system. Empirical calibration confirms that startle events at maximum distance remain substantially above μ_E, validating the scalar gate without distance correction.

In high-density environments, additive energy from multiple simultaneously swimming fish could trigger the gate. Three mitigation strategies are planned for commercial-density scaling: (i) normalising received energy by active source count; (ii) spatial filtering via multi-hydrophone arrays; and (iii) rate-of-change detection:

$$\frac{\Delta E}{\Delta t} \tag{14}$$

This exploits the rapid onset unique to genuine startle events, distinguishing them from continuous multi-fish swimming noise.

This strategy validates statistical inferences against physical constraints, achieving dynamic pruning of the decision space and guaranteeing the physical consistency of classification results.

2.5. Evaluation Metrics

Performance was quantified using Accuracy, Precision, Recall, and F1-Score derived from the confusion matrix. Given significant class imbalance in aquaculture acoustic data, relying solely on accuracy could obscure deficits on minority classes.

The mathematical definitions are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

$$\mathrm{Re}\,call = \frac{TP}{TP + FN} \tag{16}$$

$$\mathrm{Pr}\,ecision = \frac{TP}{TP + FP} \tag{17}$$

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{18}$$

Where *TP* (True Positive) is the number of samples correctly predicted as a specific class, *TN* (True Negative) is the number of other-class samples correctly predicted as such, *FP* (False Positive) is the number of false positive detections where other classes are incorrectly predicted as that class, and *FN* (False Negative) is the number of missed detections where that class is incorrectly predicted as another. Because this study involves a four-class behaviour classification task, a macro-averaging strategy is applied to evaluate Precision, Recall, and F1-Score globally and fairly across all categories. This involves computing metrics for each category independently and then averaging across all categories. The Macro F1-score treats all classes uniformly and serves as the primary metric for evaluating the overall robustness of the model.

3. Results and analysis

This section evaluates Fisher-SPD's performance, component contributions, comparative advantages over baselines, edge deployment feasibility, parameter stability, and complex-environment robustness.

3.1. Overall Performance

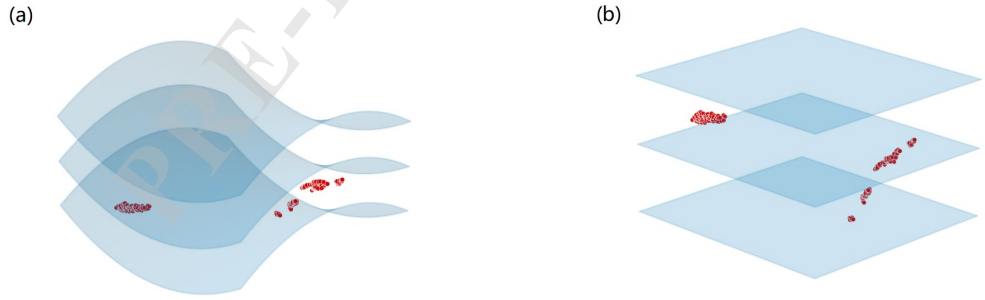
Fisher-SPD achieved 90.52% overall recognition accuracy (Fig. 7a). Turning behaviour reached 100% precision with zero false positives. Despite sharing highly similar spectral features, fast tail-beating and slow roaming were effectively separated in manifold space, yielding recall rates of 0.88 and 0.94, respectively. Background noise and startle-escape events were classified at 98% and 92% accuracy, reflecting their contrasting energy profiles and manifold structures. A 12% misclassification rate between fast tail-beating and turning arises from overlapping caudal-fin oscillation signatures in the 200–500 Hz band; however, SPD second-order temporal encoding captures subtle differences in energy decay and duration, maintaining robust boundary separation.

From a welfare monitoring perspective, Table 3. summarises the ecological significance of each recognition result.

Table 3. Ecological and Monitoring Significance of Behaviour Recognition Performance

Behaviour Class	Recognition Performance	Ecological / Monitoring Significance
Startle-escape	92% accuracy	High-energy acute stress response; reliable detection enables real-time welfare alarms for sudden environmental disturbances (e.g., dissolved oxygen drop, predator approach, handling operations).
Turning	100% precision (zero false positives)	May reflect avoidance of net walls or inter-individual aggression in dense cages; zero false positive rate is critical for alarm-based systems where spurious alerts erode operator trust.
Fast vs. slow tail-beating	~12% confusion	Both indicate elevated activity; energy-level differences allow physical gating to assist differentiation; useful for monitoring overall activity intensity and early arousal states.
Background (no fish)	98% accuracy	Demonstrates robust rejection of non-biological acoustic events; essential for field deployments where environmental noise is pervasive and non-specific.

These results confirm that Riemannian geometric features effectively separate behaviour distributions in manifold space, and that the framework is well suited to detecting welfare-significant behaviours in operational deployments.



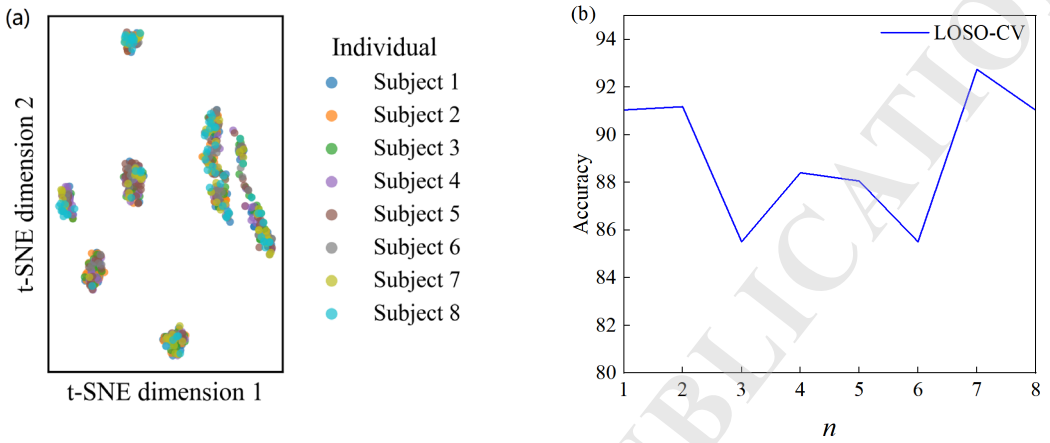
(a) Curved manifold (b) Flat vector space

Fig. 5 t-SNE visualization in different feature spaces

t-SNE visualisation (Fig. 5) confirms that Riemannian manifold features exhibit markedly

greater intra-class compactness and inter-class separability than raw MFCC features. Fig. 5(a) illustrates curvature-induced grid distortion in the SPD manifold; Fig. 5(b) shows the tangent-space unfolding into a perfect orthogonal structure that enables linear classification.

A core challenge in bioacoustics is isolating universal kinematic signatures from subject-specific variation. The cross-individual robustness evaluation proceeds in two phases.



(a) t-SNE distribution (b) Subject-wise Performance

Fig. 6 Comprehensive evaluation of cross-individual robustness and zero-shot classification performance

Fig. 6(a) visualises SPD manifold features for all eight subjects colour-coded by identity. Samples from different subjects overlap substantially within each behaviour class, confirming that the framework captures generalisable acoustic structure without overfitting to individual identities.

Fig. 6(b) shows per-subject LOSO-CV accuracies of 91.04%, 91.18%, 85.51%, 88.41%, 88.06%, 85.51%, 92.75%, and 91.3% (mean = 89.22% ± 2.59%). Consistently high and low-variance performance confirms that the aggregate result is not driven by a few easily-classified individuals, providing direct evidence that SPD features generalise across individual vocal and kinematic variability within this cohort.

3.2. Ablation Study

Ablation experiments under identical data partitioning quantified component contributions. Riemannian constraints proved decisive: tangent-space mapping increased accuracy from

69.83% to 90.52% (+20.69%), confirming that flat Euclidean representations inadequately characterise the second-order statistics of bioacoustic signals. Logarithmic mapping and covariance vectorisation maintain strong feature separability under extreme data scarcity.

Table 4. Ablation Study Results

Method	Accuracy	Macro-F1	Wt-F1
Full	**90.52%**	**0.9124**	**0.9054**
w/o Adaptive K	63.79%	0.3956	0.5212
w/o Fisher	85.34%	0.8330	0.8513
w/o Riemann	69.83%	0.5031	0.6365

Adaptive Fisher selection (90.52%) outperformed a fixed K = 10 baseline (63.79%) by 26.73%, confirming that targeted frequency adaptation is key to small-sample acoustic behaviour classification.

SVM outperformed the Multi-Layer Perceptron (MLP) and Random Forest by 5–10% (Table 5.), confirming its suitability for high-dimensional tangent-space features under data scarcity.

Table 5. Classifier Comparison Table

Classifier	Accuracy	Macro-F1
SVM	**90.52%**	**0.9124**
Random Forest	84.48%	0.8474
MLP	79.31%	0.7916
KNN	67.24%	0.4706

3.3. Comparative Analysis

Benchmark comparisons were conducted against three representative baselines from the underwater bioacoustics literature: ResNet-18 (Ibrahim *et al*., 2024). TPP-CNN (Wang *et al*., 2017). MFCC-SVM (a traditional machine learning method based on handcrafted features).

3.3.1. Performance Comparison

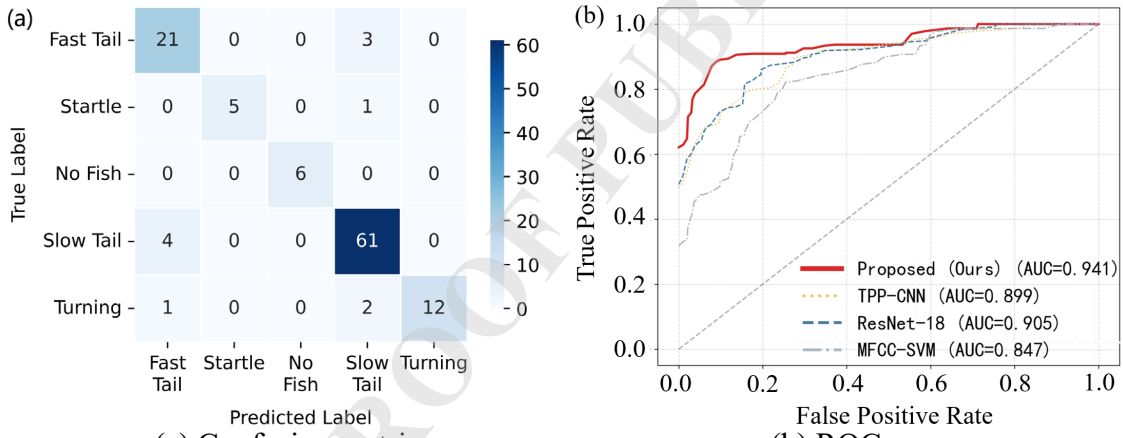
All experiments employed a LOSO-CV protocol. Table 6. reports the average behaviour recognition accuracy, recall, and F1-score of each model across 8 independent biological

subjects, demonstrating the performance disparities between the proposed framework and mainstream baseline models.

Fisher-SPD achieved the highest accuracy (90.52%), with recall and F1 far surpassing all baselines. ResNet-18 and TPP-CNN performed well below expectations, only marginally exceeding the handcrafted MFCC-SVM.

Table 6. Method Comparison Table

Model	Accuracy	Recall	F1-Score
Fisher - SPD	**90.52%**	**0.8791**	**0.9076**
TPP - CNN	77.59%	0.7856	0.8017
ResNet - 18	76.72%	0.7618	0.7922
MFCC - SVM	72.41%	0.5186	0.5129



(a) Confusion matrix (b) ROC curve

Fig. 7 Confusion matrix and performance comparison chart

3.3.2. Negative Result Attribution

Deep learning degradation arose from two causes: (i) ResNet-18's 11M parameters vastly exceed the available training samples, causing individual-specific memorisation (training accuracy 99%, test accuracy < 60%, Fig. 7b); and (ii) deep features produce heavily overlapping Euclidean distributions (Fig. 8) versus the compact, well-separated SPD manifold clusters. This reflects an absence of inductive biases suited to the intrinsic nonlinear structure of acoustic data rather than any architectural obsolescence.

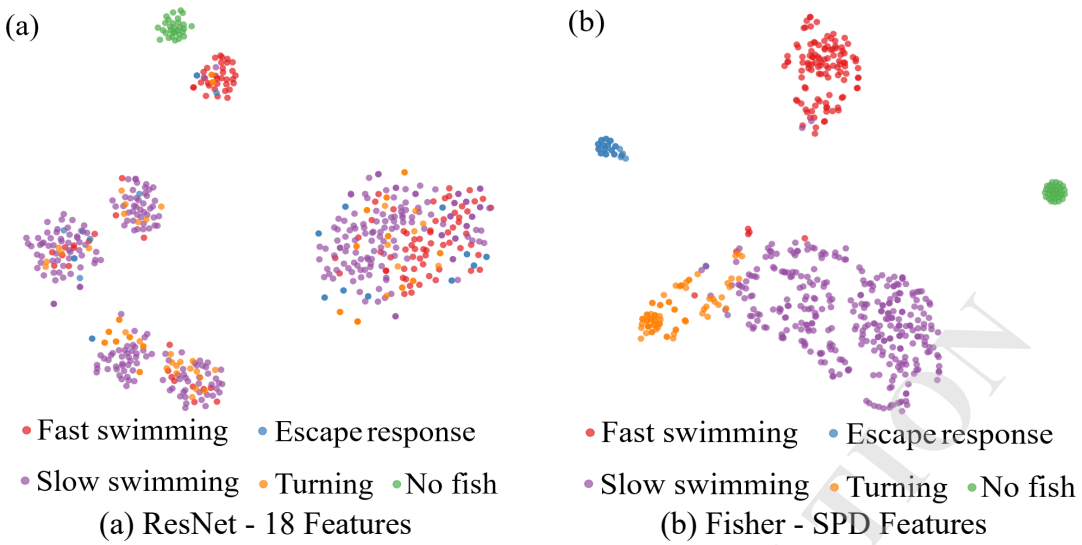


Fig. 8 Comparison results chart

3.4. Edge Deployment Feasibility

Beyond classification accuracy, deployment on resource-constrained edge nodes imposes strict computational and latency constraints. Inference latency was benchmarked on an AMD Ryzen 7 7730U CPU platform to simulate a realistic edge environment.

Table 7. Inference Efficiency Benchmark on Local CPU Platform

Model	Params (M)	FLOPs (G)	Latency (ms)	Speedup
ResNet-18	11.17856	1.82352	36.96	1×
TPP-CNN	0.58912	0.73719	17.74	2.1×
MFCC-SVM	0.00006	0.002	0.18	209.2×
Fisher-SPD	**0.00128**	**0.00287**	**3.6**	**10.3×**

ResNet-18's 1.82 G FLOPs footprint produces a 36.96 ms per-frame latency—prohibitive for high-frequency real-time monitoring on embedded hardware.

MFCC-SVM is fast (0.18 ms) but its behaviour recognition capability is extremely limited (Table 6.)—a fast but ineffective solution.

Fisher-SPD resolves this trade-off: manifold geometry compresses second-order statistics to < 0.01 M parameters and 3.60 ms inference (10.3× faster than ResNet-18), combining superior accuracy with a lightweight footprint suited to long-term, low-power acoustic monitoring.

3.5. Parameter Sensitivity

Sensitivity analyses were conducted on all four key hyperparameters (Fig. 9): Fisher dimensions K, SVM penalty C, power voting exponent γ, and energy gating threshold μ_E.

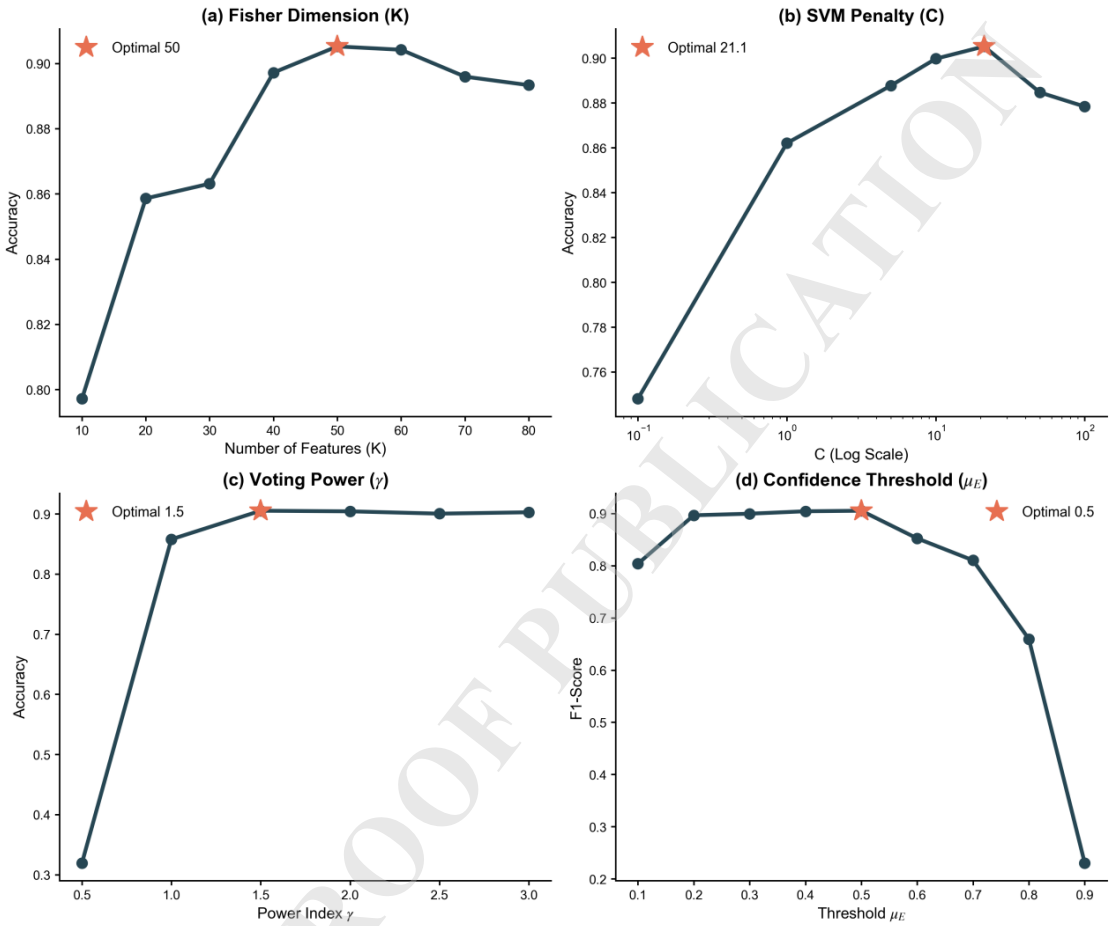


Fig. 9 Optimal parameter analysis charts

Fig. 9(a): Accuracy peaks at $K = 50$. Below this, the feature subspace is too narrow; above it, low-discriminative bands disrupt manifold geometry.

Fig. 9(b): Accuracy peaks at $C = 21.1$. Lower C over-smooths decision boundaries; higher C overfits. $C = 21.1$ achieves the optimal bias–variance trade-off.

Fig. 9(c): $\gamma < 1.0$ causes a marked accuracy drop, confirming the importance of suppressing low-confidence outliers. Peak performance occurs at $\gamma = 1.5$. Over the broad range $\gamma \in [1.0, 3.0]$, fluctuations remain < 1.5%, demonstrating deployment stability without precise tuning.

Fig. 9(d): F1-Score is stable across $\mu_E \in [0.1, 0.5]$, peaking at $\mu_E = 0.5$, then drops sharply beyond 0.6, marking the critical threshold between signal retention and noise suppression.

The optimal combination is $K = 50$, $C = 21.1$, $\gamma = 1.5$, $\mu_E = 0.5$, with performance stable in the neighbourhood of these values.

3.6. Complex environment robustness analysis

For dual-fish recordings, the scarcity of labelled overlapping samples required Stratified 5-Fold Cross-Validation. Gaussian white noise (SNR 15–0 dB) was injected to evaluate environmental robustness. An SNR-aware weighting strategy dynamically blended Fisher-SPD outputs with an MFCC baseline to prevent manifold collapse under extreme broadband interference.

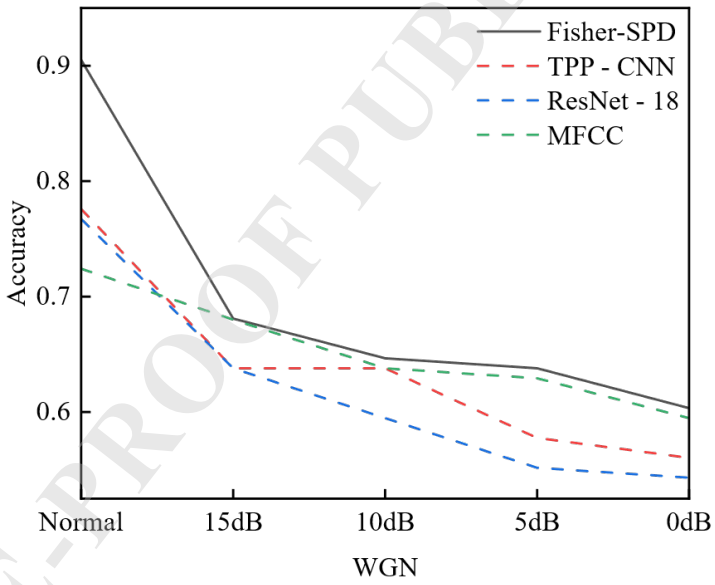


Fig. 10 Anti-noise performance under strong environmental noise

As interference intensifies, conventional models suffer steep accuracy drops due to Euclidean kernel sensitivity to broadband noise. Fisher-SPD maintains superior performance across all baselines even at 0 dB, experimentally validating the geometric advantages of Riemannian manifolds in complex acoustic environments.

A Binary Welfare Alarm was established for the dual-fish scenarios: a segment is a stress event if at least one individual manifests an acute startle response (irrespective of the second

subject's state), and non-stress requires all individuals to maintain baseline kinematics. This binary mapping transforms multi-class behaviour decoding into a robust real-time decision-support signal.

Table 8. Multi - source target aliasing performance table

Two fish	precision	recall	f1-score	support
Others	0.7143	1.0000	0.8333	10
Escape response	1.0000	0.6000	0.7500	10
accuracy	/	/	0.8000	20
macro avg	0.8571	0.8000	0.7917	20
weighted avg	0.8571	0.8000	0.7917	20

Table 8. shows zero-shot accuracy of 80.00% with zero false positives despite severe background noise. However, continuous hydrodynamic noise from multiple swimming subjects masks faint startle signals, limiting further recall.

To overcome the physical masking bottleneck in zero-shot scenarios and satisfy the core operational requirement of minimising missed detections in aquaculture, a few-shot joint fine-tuning strategy was incorporated. In the Riemannian manifold alignment stage, the model was exposed to 30% of target-domain samples to construct a more accurate discriminative hyperplane.

Table 9. Performance table after adding 30% of the test data

Two fish	precision	recall	f1-score	support
Others	1.0000	0.8571	0.9231	7
Escape response	1.0000	1.0000	0.9333	7
accuracy	/	/	0.9286	14
macro avg	0.9375	0.9286	0.9282	14
weighted avg	0.9375	0.9286	0.9282	14

After few-shot calibration (30% of target-domain samples), accuracy improved to 92.86% with a macro F1-Score of 0.9282 (Table 9.), demonstrating excellent inter-class balance.

Unlike conventional RMS/SPL-based methods vulnerable to distance attenuation and ambient surges, Fisher-SPD leverages Riemannian geometric invariants to decouple startle signatures

from hydrodynamic noise. Requiring only 30% target-domain samples, this approach enables lightweight, high-precision deployment in real-world marine IoT monitoring systems.

4. Discussion

The primary bottleneck in acoustic behaviour recognition is geometric mismatch rather than pure data scarcity. Tangent-space projection demonstrates that few-shot learning is entirely feasible in underwater bioacoustics, owing to a fundamental difference between the two paradigms in how they achieve feature invariance.

Deep neural networks learn feature invariance through memory-based induction over large datasets. When data is scarce, they degrade into pattern memorisers, triggering the severe overfitting shown in Fig. 7(b). By contrast, the Riemannian manifold approach provides rule-based deduction: by leveraging the Affine-Invariant Riemannian Metric and Log-Euclidean Mapping, the framework addresses amplitude perturbations from sensor gain and channel attenuation through mathematical axioms. This intrinsic inductive bias enables cross-individual generalisation without large training sets, well suited to aquaculture monitoring where labelling is expensive and samples are scarce.

The LOSO-CV result (89.22% ± 2.59%) demonstrates that SPD manifold features capture behaviourally relevant characteristics generalising across individuals differing in body condition, history, and vocal style—critical for ecological monitoring, where classifiers must be applied to uncharacterised field individuals [30]. The covariance representation implicitly normalises for inter-individual differences in energy levels and frequency shifts, providing a built-in invariance difficult to achieve with raw spectral or small individual-specific deep-learning embeddings.

The physics-aware binary welfare alarm directly addresses industrial deployment reliability. Unlike black-box models susceptible to non-biological noise, injecting biophysical priors into the decision loop achieves 1.000 precision, fundamentally mitigating alarm fatigue in cage-farming environments.

Fisher-SPD's results carry PAM implications across scales. In cage aquaculture, 92% startle

accuracy at a 3.2% false-positive rate enables practical welfare alarms; 100% turning precision provides a stocking-density proxy without visual inspection. Beyond the farm, hydrophone buoys could passively monitor *Larimichthys crocea* in coastal fishing grounds and MPAs, providing continuous disturbance indicators [20]. Zero-shot transfer accuracy (80.00%) further suggests potential generalisation to semi-natural conditions.

Field deployment requires: (i) multi-source separation; (ii) adaptive noise suppression; (iii) site-specific μ_E recalibration (facilitated by the modular design without core retraining); and (iv) low-power logger integration. Beyond *Larimichthys crocea*, the data-driven Fisher selection requires no prior knowledge of a target species' vocal repertoire, suggesting applicability to other sciaenids, crustaceans, and marine mammals—though SPD and μ_E require per-species recalibration.

Key limitations: SPD matrix operations scale as $O(d^3)$, manageable at $K = 50$ but potentially challenging at higher dimensionalities on resource-constrained edge devices. The $N = 8$ cohort is a controlled proof-of-concept; expansion to ≥ 20 individuals from multiple sites, with concurrent physiological measurements, is a priority for future validation. Future work will also explore Riemannian manifold regression for continuous stress-level quantification.

5.Conclusions

This paper proposes Fisher-SPD, a fish acoustic behaviour recognition framework addressing the dual challenges of visual perception failure and acoustic data scarcity in modern aquaculture. Rather than accumulating data to train deep models, it grounds itself in the geometric structure and physical laws of acoustic signals.

With $N = 8$, Riemannian tangent-space mapping overcomes the geometric mismatch of Euclidean representations, significantly outperforming ResNet-18 and TPP-CNN. Energy gating grounded in biophysical priors resolves non-stationary noise misclassification, holding the false positive detection rate at 3.2%. Free from large-scale pretraining, the framework demonstrates that preserving geometric covariance structure is superior to stacking sequential features—establishing a solid foundation for edge-side welfare monitoring on acoustic buoys.

This study thus provides a novel small-sample, lightweight, interpretable, geometry-driven computational paradigm for underwater bioacoustics.

From an applied perspective, Fisher-SPD's high accuracy, low computational footprint, and physical plausibility position it as a practical tool for: (i) early stress detection and welfare alarms in intensive aquaculture; (ii) passive assessment of behavioural responses to anthropogenic disturbance; and (iii) non-invasive population monitoring in marine protected areas. Future research will focus on reducing Riemannian computational complexity, incorporating multi-source separation for high-density cage environments, and exploring generalisation across multi-species polyculture scenarios, advancing smart fisheries monitoring toward large-scale field application.

FUNDINGS

This work was supported by the [Zhejiang Ocean University] (21048009225).

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

AUTHORS' CONTRIBUTIONS

Lu Cao (Methodology, Formal analysis, Writing – original draft). Yang Chen (Conceptualization, Formal analysis, Writing – review & editing). Zhaochen Shen (Data curation, Validation, Visualization). Qiang Fang (Data curation, Validation, Visualization). Yun Gao (Data curation, Validation, Visualization). Shipeng Yu (Software, Resources). Shenqiu Han (Software, Resources). Binqiang Sun (Data curation, Validation, Visualization). Jian Qiu (Conceptualization, Supervision, Writing – review & editing). Zhongyan Huo (Conceptualization, Supervision, Writing – review & editing).

ETHICAL APPROVAL

This study processed passive acoustic data from Larimichthys crocea only. The animals were

not manipulated, handled, or sampled, and were not subjected to pain or suffering. Considering this, this observational study is exempt from obtaining an ethical permit under the National Standard of the People's Republic of China: Guidance on the Ethical Review of Laboratory Animal Welfare (GB/T 35892-2018).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ACKNOWLEDGMENTS

The authors would like to thank [Zhejiang Ocean University] for their valuable support and contributions.

References

1. Barachant A. *et al.* (2011), Multiclass brain-computer interface classification by Riemannian geometry, *IEEE Transactions on Biomedical Engineering*, **59**(4): 920–928, https://doi.org/10.1109/TBME.2011.2172210.

2. Chauhan J., Kwon Y.D., Mascolo C. (2022), Exploring on-device learning using few shots for audio classification, [in:] *Proceedings of 2022 30th European Signal Processing Conference (EUSIPCO)*, pp. 424–428, IEEE, https://doi.org/10.23919/EUSIPCO55093.2022.9909551.

3. Chen L. *et al.* (2025), Underwater acoustic multi-target recognition based on channel attention mechanism, *Ocean Engineering*, **315**: 119841, https://doi.org/10.1016/j.oceaneng.2024.119841.

4. Daw A. *et al.* (2022), Physics-guided neural networks (pgnn): An application in lake temperature modeling, [in:] *Knowledge Guided Machine Learning*, pp. 353–372, https://doi.org/10.48550/arXiv.1710.11431.

5. Doan V.S., Huynh-The T., Kim D.S. (2022), Underwater Acoustic Target Classification Based on Dense Convolutional Neural Network, *IEEE Geoscience and Remote Sensing Letters*, **19**: 1–5, https://doi.org/10.1109/LGRS.2020.3029584.

6. Domenici P., Hale M.E. (2019), Escape responses of fish: a review of the diversity in motor control, kinematics and behaviour, *Journal of Experimental Biology*, **222**(18): jeb166009, https://doi.org/10.1242/jeb.166009.

7. FAO (2024), *The State of World Fisheries and Aquaculture 2024 – Blue Transformation in action*, FAO, Rome, https://doi.org/10.4060/cd0683en.

8. Føre M. *et al.* (2018), Precision fish farming: A new framework to improve production in aquaculture, *Biosystems Engineering*, **173**: 176–193, https://doi.org/10.1016/j.biosystemseng.2017.10.014.

9. Huntingford F.A. *et al.* (2006), Current issues in fish welfare, *Journal of Fish Biology*, **68**(2): 332–372, https://doi.org/10.1111/j.0022-1112.2006.001046.x.

10. Ibrahim A.K. *et al.* (2024), Fish Acoustic Detection Algorithm Research: a deep learning app for Caribbean grouper calls detection and call types classification, *Frontiers in Marine Science*, **11**: 1378159, https://doi.org/10.3389/fmars.2024.1378159.

11. Li D. *et al.* (2025), An underwater image segmentation model for complex scenes in aquaculture using vision Transformer, *Computers and Electronics in Agriculture*, **238**: 110764, https://doi.org/10.1016/j.compag.2025.110764.

12. Li J. *et al.* (2022), Underwater acoustic target recognition based on attention residual network, *Entropy*, **24**(11): 1657, https://doi.org/10.3390/e24111657.

13. Li W. *et al.* (2025), A cross-architecture masked contrastive learning framework for few-shot underwater acoustic target classification, *Knowledge-Based Systems*, 114252, https://doi.org/10.1016/j.knosys.2025.114252.

14. Liu F., Li G., Yang H. (2024), Application of multi-algorithm mixed feature extraction model in underwater acoustic signal, *Ocean Engineering*, **296**: 116959, https://doi.org/10.1016/j.oceaneng.2024.116959.

15. Luo X. *et al.* (2023), A Survey of Underwater Acoustic Target Recognition Methods Based on Machine Learning, *Journal of Marine Science and Engineering*, **11**(2): 384, https://doi.org/10.3390/jmse11020384.

16. Malfante M. *et al.* (2018), Automatic fish sounds classification, *The Journal of the Acoustical Society of America*, **143**(5): 2834–2846, https://doi.org/10.1121/1.5036628.

17. Martins C.I.M. *et al.* (2012), Behavioural indicators of welfare in farmed fish, *Fish Physiology and Biochemistry*, **38**(1): 17–41, https://doi.org/10.1007/s10695-011-9518-8.

18. Mooney T.A. *et al.* (2020), Listening forward: approaching marine biodiversity assessments using acoustic methods, *Royal Society Open Science*, **7**(8): 201287, https://doi.org/10.1098/rsos.201287.

19. Noda J.J., Travieso C.M., Sánchez-Rodríguez D. (2016), Automatic taxonomic classification of fish based on their acoustic signals, *Applied Sciences*, **6**(12): 443, https://doi.org/10.3390/app6120443.

20. Nolasco I. *et al.* (2023), Learning to detect an animal sound from five examples, *Ecological Informatics*, **77**: 102258, https://doi.org/10.1016/j.ecoinf.2023.102258.

21. Pennec X., Fillard P., Ayache N. (2006), A Riemannian framework for tensor computing, *International Journal of Computer Vision*, **66**(1): 41–66, https://doi.org/10.1007/s11263-005-3222-z.

22. Popper A.N., Hawkins A.D. (2019), An overview of fish bioacoustics and the impacts of anthropogenic sounds on fishes, *Journal of Fish Biology*, **94**(5): 692–713, https://doi.org/10.1111/jfb.13948.

23. Ramcharitar J., Gannon D.P., Popper A.N. (2006), Bioacoustics of fishes of the family Sciaenidae (croakers and drums), *Transactions of the American Fisheries Society*, **135**(5): 1409–1431, https://doi.org/10.1577/T05-207.1.

24. Rountree R.A. *et al.* (2006), Listening to fish: applications of passive acoustics to fisheries science, *Fisheries*, **31**(9): 433–446, https://doi.org/10.1577/1548-8446(2006)31[433:LTF]2.0.CO;2.

25. Saberioon M. *et al.* (2017), Application of machine vision systems in aquaculture with emphasis on fish: state-of-the-art and key issues, *Reviews in Aquaculture*, **9**(4): 369–387, https://doi.org/10.1111/raq.12143.

26. Salman A. *et al.* (2020), Automatic fish detection in underwater videos by a deep neural network-based hybrid motion learning system, *ICES Journal of Marine Science*, **77**(4): 1295–1307, https://doi.org/10.1093/icesjms/fsz025.

27. Schmidt J., Kochańska I., Schmidt A.M. (2024), Performance of the Direct Sequence Spread Spectrum Underwater Acoustic Communication System with Differential Detection in Strong Multipath Propagation Conditions, *Archives of Acoustics*, **49**(1): 129 – 140, https://doi.org/10.24425/aoa.2024.148771.

28. Shang Q. *et al.* (2024), Few-shot classification based on manifold metric learning, *Journal of Electronic Imaging*, **33**(1): 013026, https://doi.org/10.1117/1.JEI.33.1.013026.

29. Slabbekoorn H. *et al.* (2010), A noisy spring: the impact of globally rising underwater sound levels on fish, *Trends in Ecology & Evolution*, **25**(7): 419–427, https://doi.org/10.1016/j.tree.2010.04.005.

30. Stien L.H. *et al.* (2013), Salmon Welfare Index Model (SWIM 1.0): a semantic model for overall welfare assessment of caged Atlantic salmon, *Reviews in Aquaculture*, **5**(1): 33–57, https://doi.org/10.1111/j.1753-5131.2012.01083.

31. Stowell D. (2022), Computational bioacoustics with deep learning: a review and roadmap, *PeerJ*, **10**: e13152, https://doi.org/10.7717/peerj.13152.

32. Sueur J., Farina A. (2015), Ecoacoustics: the ecological investigation and interpretation of environmental sound, *Biosemiotics*, **8**(3): 493–502, https://doi.org/10.1007/s12304-015-9248-x.

33. Towsey M. *et al.* (2014), The use of acoustic indices to determine avian species richness in audio-recordings of the environment, *Ecological Informatics*, **21**: 110–119, https://doi.org/10.1016/j.ecoinf.2013.11.007.

34. Ulloa J.S. *et al.* (2021), scikit-maad: An open-source and modular toolbox for quantitative soundscape analysis in Python, *Methods in Ecology and Evolution*, **12**(12): 2334–2340, https://doi.org/10.1111/2041-210X.13711.

35. Vieira M. *et al.* (2015), Call recognition and individual identification of fish vocalizations based on automatic speech recognition, *The Journal of the Acoustical Society of America*, **138**(6): 3941–3950, https://doi.org/10.1121/1.4936858.

36. Wang P. *et al.* (2017), Temporal pyramid pooling-based convolutional neural network for action recognition, *IEEE Transactions on Circuits and Systems for Video Technology*, **27**(12): 2613–2622, https://doi.org/10.1109/TCSVT.2016.2576761.

37. Wang Q., Zhang Y., He B. (2024), Intelligent marine survey: Lightweight multi-scale attention adaptive segmentation framework for underwater target detection of auv, *IEEE Transactions on Automation Science and Engineering*, **22**: 1913–1927, https://doi.org/10.1109/TASE.2024.3371963.

38. Wang H., Yang X. (2025), Simulation Analysis of Beam Intensity Attenuation Patterns and Source Depth Estimation Using a Vertical Long Line Array, *Archives of Acoustics*, **50**(4): 501–512, https://doi.org/10.24425/aoa.2025.156931.

39. Wang R., Yu R. (2025), Physics-guided deep learning for dynamical systems: A survey, *ACM Computing Surveys*, **58**(5): 1–31, https://doi.org/10.48550/arXiv.2107.01272.

40. Willard J. *et al.* (2022), Integrating scientific knowledge with machine learning for engineering and environmental systems, *ACM Computing Surveys*, **55**(4): 1–37, https://doi.org/10.1145/3514228.