2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Class-Guided Network with Rare-Class Amplification for Sea State Estimation Based on Ship Motion Data

Wei Xia[1], Kexin Wang[1], Weiwei Tian[1], Xiufeng Liu[2], Fan Shi[1], and Xu Cheng[1*]

***Abstract*— Accurate, real-time Sea State Estimation (SSE) is crucial for the safety and operational efficiency of Autonomous Surface Vessels (ASVs). However, existing deep learning methods for this task commonly face three major challenges: the inherent class imbalance of marine environments, the ambiguous boundaries between discrete sea state levels, and the difficulty of extracting multi-scale temporal features from vessel motion. To address these challenges, this paper proposes a novel framework named the Class-guided Rare-boosted Multi-Scale Net (CRUISE). The framework is built upon a multi-scale encoder-decoder architecture and integrates two key innovations: a Rare-Boosted Class Embedding (RBCE) module at the network's bottleneck and a class-guided decoding mechanism. The RBCE module first generates a preliminary class prediction and then dynamically enhances the representation of rare sea state classes to create a class-balanced conditional vector. This vector subsequently provides top-down guidance to the decoder, injecting class-aware information by modulating the feature reconstruction process. This synergistic design fundamentally addresses the data imbalance problem at the feature level and effectively sharpens the decision boundaries between easily confused transitional sea states. Extensive experiments on multiple public benchmarks, simulated ship motion, and real-world datasets demonstrate that CRUISE significantly outperforms existing state-of-the-art methods, showing a pronounced advantage in improving the recognition accuracy of rare and high-risk sea states. Furthermore, real-time inference tests on a physical model vessel validate the model's performance on edge computing devices, further confirming its feasibility and robustness for deployment in real-world marine environments.**

I. INTRODUCTION

Autonomous Surface Vessels (ASVs) are increasingly deployed in critical missions spanning scientific exploration, commercial logistics, and maritime security. Their safety and operational efficiency in dynamic ocean environments hinge on precise, real-time environmental perception. Central to this capability is Sea State Estimation (SSE), which provides essential a priori information for high-level planners to optimize motion planning, mitigate risks, and improve energy efficiency. Consequently, an advanced SSE system is a cornerstone for developing robust and environment-adaptive maritime autonomous systems.

Traditional SSE methods, relying on wave buoys or remote sensing, are often ill-suited for modern maritime operations due to high deployment costs and inherent data latency. These limitations hinder real-time tasks like dynamic route planning. A more practical approach treats the vessel itself as a mobile sensor—a large wave buoy. However, existing model-driven methods within this paradigm suffer from substantial computational overhead and a strong dependence on precise hydrodynamic models. This reliance on a priori knowledge restricts their adaptability and robustness in unfamiliar marine conditions.

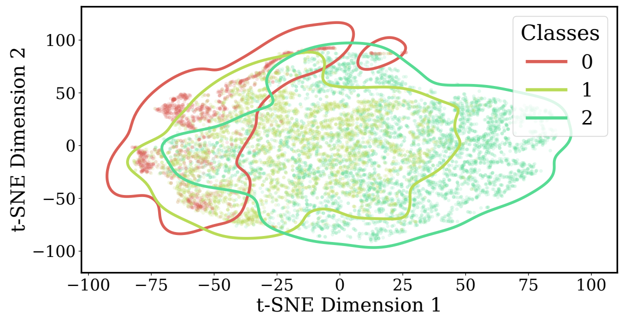

Fig. 1: t-SNE Visualization of the Learned Feature Space.

Data-driven deep learning methods offer a promising alternative, capable of learning the complex mapping between vessel motion and sea state directly from onboard sensor data. This end-to-end paradigm circumvents the need for explicit physical models and shows great potential for real-time inference. Despite this promise, the effective application of deep learning to SSE faces three fundamental challenges:

- **Class Imbalance.** The distribution of sea states is highly skewed, with calm and moderate conditions far more frequent than extreme ones. This imbalance biases models toward majority classes, making rare but safety-critical states difficult to detect, which is particularly concerning for risk-aware decision-making in autonomous vessels.
- **Class Boundary Ambiguity.** Sea state is a continuous physical process discretized into categorical levels, where transitional samples between adjacent classes exhibit high similarity. As shown in Fig. 1, this inherent ambiguity blurs decision boundaries and undermines the system's ability to recognize critical transitions essential for robust autonomy.
- **Multi-Scale Features.** Sea state signals are composed of features at multiple time scales, from low-frequency wave energy trends to high-frequency wind-wave disturbances. Capturing this full spectrum is crucial for accurate classification, yet conventional networks often focus on a single feature scale.

To address the aforementioned challenges, this paper proposes the **C**lass-Guided **R**are-Class Boosting **MU**lt**I**-**S**cal**E** Network (CRUISE), a robust SSE solution for Autonomous

*Corresponding author
[1]School of Computer Science and Technology, Tianjin University of Technology, 300384, Tianjin, China.
[2]Department of Technology, Management and Economics, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark.

Surface Vehicles ASVs. Centered on a multi-scale encoder-decoder architecture, the model integrates two key innovations: Rare-Class Boosting Embedding (RBCE) and Class-Guided Decoding. These mechanisms effectively mitigate data imbalance, sharpen ambiguous class boundaries, and efficiently capture multi-scale features within vessel motion. CRUISE not only achieves high-precision classification but also bridges the gap between sensor data and autonomous decision-making, significantly enhancing the perception capabilities and safety of ASVs in complex maritime environments.

In summary, the main contributions of this paper are as follows:

- This paper proposes CRUISE, a novel model that enhances a multi-scale encoder-decoder architecture with a dynamic, class-aware bottleneck, offering a solution specifically tailored for robust SSE from complex time-series data.
- A unique class-guidance mechanism centered on the RBCE module is introduced. This mechanism provides a structural solution to data imbalance, operating independently of loss function weighting, to adaptively enhance minority class features and sharpen ambiguous decision boundaries.
- Extensive experiments on multiple public benchmarks and two real-world ship motion datasets demonstrate that the proposed model consistently outperforms state-of-the-art methods. It shows significant advantages in both overall performance and, critically, in the recognition of rare sea state classes.

II. Related Work

A. Model-based Methods

Given the reliance of traditional methods on dedicated, often costly and stationary, equipment, they struggle to meet the modern demands for real-time, cost-effective, and flexible SSE. To overcome these limitations, the Wave Buoy Analogy (WBA) method was proposed. Model-based methods, grounded in the WBA, treat sensor-equipped vessels as mobile wave observation platforms [1]. Using the Response Amplitude Operator (RAO), they map vessel motion to wave spectra for buoy-free SSE [2].

In frequency-domain approaches, the RAO functions as a wave–motion transfer operator to reconstruct directional spectra; Zhang and Ren introduced an RAO invertibility index and a fusion strategy based on the restricted isometry property to improve heave and pitch reconstruction [3], though these methods require long observations and are sensitive to RAO errors. Time-domain approaches instead apply state-space modeling with Kalman filtering [4]. These support minute-scale, non-stationary seas but remain constrained by system identification and model assumptions. However, purely model-driven methods are often limited by their significant computational overhead and reliance on precise hydrodynamic models. This strong dependence on prior knowledge makes it difficult for them to adapt to changing marine environments.

B. Data-driven Methods

With the advancement of computational technologies, data-driven methods have achieved significant progress in SSE. Early research primarily relied on classical algorithms such as Artificial Neural Networks (ANNs) [5] and Support Vector Machines (SVMs) [6] to classify or regress ship motion data for estimating wave parameters. Han et al. proposed a nearshore estimation method combining statistical analysis with wavelet analysis [7]. However, such methods are constrained by the limited representational capacity of traditional machine learning, resulting in insufficient accuracy and generalization performance under complex sea conditions.

The emergence of Deep Learning (DL) has greatly enhanced the accuracy and robustness of SSE. In recent years, SeaStateNet, proposed by Cheng et al., integrated LSTM, CNN, and FFT, and was later developed into the more robust SSENet [8]. They also introduced ImbalanceSSE [9], which combines multi-scale features with a prototype classifier to address data imbalance.And ASK-ProtoNet from Xu et al. combined an adaptive selective kernel with prototype learning, achieving superior performance in few-shot scenarios [10]. However, existing methods for handling class imbalance largely rely on loss re-weighting or resampling. While these strategies can partially improve performance, they neither fundamentally correct the distributional bias of the feature space nor effectively address the issue of indistinct class boundaries, leaving notable limitations unresolved.

III. Method

A. SSE Problem Statement

SSE seeks to find the prevailing wave conditions at a specific location and time on the open sea [11]. The SSE problem is modeled as a time-series classification task. The input is a multivariate time series $\mathbf{X} \in \mathbb{R}^{N\times C}$, where N is the time series length, and C represents ship motion parameters. The goal is to learn a model $g : \mathbf{X} \to \mathbf{Y}$, mapping motion data $\mathbf{X}$ to a finite set of sea state labels $\mathbf{Y} = \{y_1, y_2, \dots, y_K\}$. Each sea state y_k is associated with a significant wave height range $H_s \in [h_{\min}(y_k), h_{\max}(y_k))$.

B. Model Overview

The proposed CRUISE, built upon a multi-scale encoder-decoder architecture (Fig. 2), is engineered to overcome three primary challenges in sea state classification: inadequate multi-scale feature extraction, indistinct class boundaries, and imbalanced class distributions. The end-to-end pipeline integrates data preprocessing, feature encoding, and a novel class-guided and rare-class boosted decoding mechanism, culminating in a final classification stage.

The raw multivariate time series were preprocessed prior to training: outliers were corrected to reduce noise; Pearson correlation analysis was used to detect redundant features to guide subsequent processing; Z-score normalization was applied; and a sliding-window strategy (window size: 128, stride: 64) was employed to generate uniformly structured

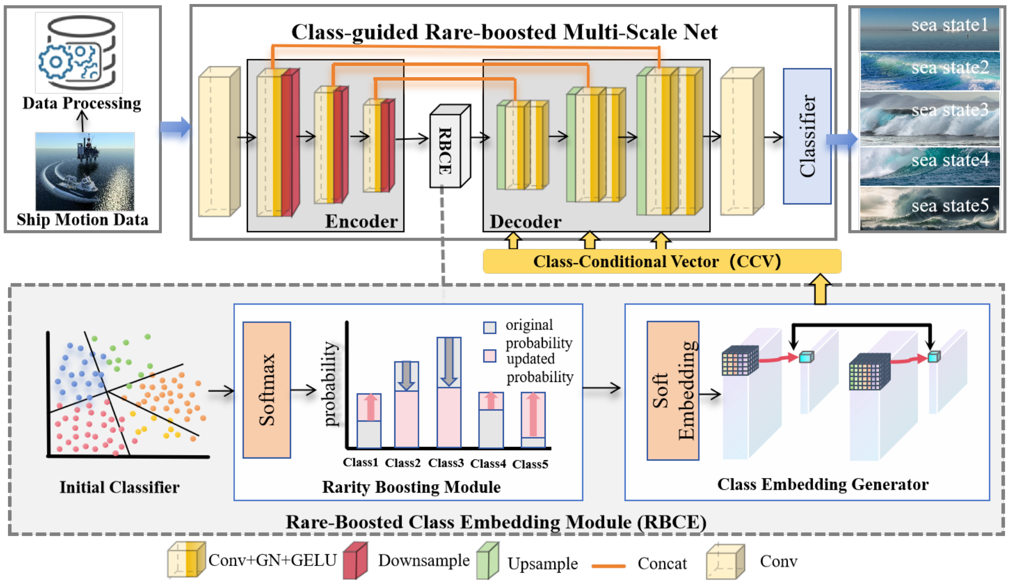

Fig. 2: Architecture of the proposed model.

training samples. Subsequently, a 1D convolutional layer embeds the processed sequences into a high-dimensional feature space for input to the encoder. This encoder hierarchically abstracts features via successive convolution and downsampling operations, adeptly capturing both global dynamics and local fluctuations. At the architectural bottleneck, our RBCE module dynamically recalibrates feature representations based on a class probability distribution from a preliminary classifier. This process strategically amplifies features of rare classes and generates a class conditional vector that guides the subsequent decoding phase. The decoder then synergistically fuses multi-scale features under this class-prior guidance, progressively recovering fine-grained discriminative details and thereby sharpening the boundaries between adjacent classes. This process is designed to progressively recover and refine features into a highly separable latent space, thereby sharpening the boundaries between adjacent classes. A final convolutional layer integrates these refined features before they are passed to the classifier, which maps them to the prediction space to yield the final sea state classification.

Through its synergistic design of multi-scale modeling and targeted class enhancement, CRUISE delivers robust sea state identification in complex marine environments, exhibiting a pronounced superiority in the discrimination of rare classes.

C. Encoder

The encoder aims to extract comprehensive temporal features from the pre-processed sequence $X \in \mathbb{R}^{B \times C_{\text{in}} \times L}$. A 1D convolutional layer first embeds the input into a high-dimensional feature space, preparing it for deep feature abstraction. The sequence is then processed by N stacked encoding modules, each progressively compressing temporal information into a refined latent representation. This hierarchical design captures both long-term global dynamics and short-term local perturbations, enabling robust modeling of complex motion patterns.

Within the n-th module ($n = 1, \ldots, N$), the input features are passed through a standard block (1D convolution, GroupNorm, GELU). The intermediate output $\tilde{F}^{(n)}$ is split into S groups along the channel dimension, $\{\tilde{F}_i^{(n)}\}_{i=1}^{S}$, each processed by a convolution with kernel size k_i. The resulting features are concatenated to form the multi-scale representation:

$$F_{\text{ms}}^{(n)} = \text{Concat}\big(\{\text{Conv}_{k_i}(\tilde{F}_i^{(n)})\}_{i=1}^{S}\big). \tag{1}$$

This representation is then downsampled, yielding $F^{(n)}$, which serves both as input to the next module and as a skip connection $h^{(n)}$ for the decoder.

After N modules, the final feature $F^{(N)}$ constitutes the bottleneck representation F_b. Through this progressive, multi-scale abstraction, the encoder systematically extracts both global semantics and fine-grained local cues essential for discriminative sea state classification.

D. Rare-Boosted Class Embedding Module

In multi-class classification tasks, rare categories often suffer from insufficient representation due to class imbalance. When the decoder operates without category awareness, it performs unconditional reconstruction in a latent space dominated by frequent classes, causing rare class features to be overwhelmed and resulting in blurred decision boundaries. To address this bottleneck, we propose the RBCE module, which explicitly injects class priors between the encoder and decoder, transforming the latent space from class-agnostic to class-sensitive. This operation enhances the visibility and separability of rare classes, directly mitigating the adverse effects of imbalance.

Formally, the encoder first outputs preliminary logits $z_{prelim} \in \mathbb{R}^{B \times C}$, where B denotes the batch size and C is the total number of classes. These logits are converted to initial class probabilities using a temperature-scaled softmax:

$$p_{prelim}(b) = \text{softmax}\left(\frac{z_{prelim}(b)}{\tau}\right), \quad b = 1, \ldots, B, \tag{2}$$

where $\tau > 0$ is the temperature coefficient controlling distribution sharpness. Due to class imbalance, rare classes naturally receive low probabilities in p_{prelim}, resulting in weak representations in the latent space.

To actively alleviate this imbalance, RBCE introduces a class weight vector $\omega \in \mathbb{R}^C$ to boost rare class probabilities:

$$p_{boost}(b,i) = \frac{\omega_i \, p_{prelim}(b,i)}{\sum_{j=1}^{C} \omega_j \, p_{prelim}(b,j)}, \quad i = 1, \ldots, C, \quad (3)$$

where $\omega_i = \alpha > 1$ for $i \in R$ (the set of rare classes) and $\omega_i = 1$ otherwise. R is a predefined set of indices corresponding to rare classes, supplied as a model hyperparameter. This operation not only increases the influence of rare classes but also actively reshapes the geometry of the latent space, creating more distinct and well-separated regions for rare class representations, thereby mitigating bias introduced by imbalanced training data.

The reweighted distribution is then mapped back to logits:

$$z_{boost}(b) = \tau \cdot \log p_{boost}(b), \quad (4)$$

and further projected to a class-conditional vector(CCV) using a learnable function $\Phi(\cdot)$:

$$CCV = \Phi(z_{boost}), \quad CCV \in \mathbb{R}^d, \quad (5)$$

where d is the embedding dimension. The function Φ is implemented as a two-layer MLP with a GELU activation. The vector CCV serves as an explicit prior guiding the decoder's feature synthesis. By continuously aligning the reconstructed features with this class prior, the decoder is forced to emphasize rare classes and maintain balanced, separable representations across all categories. As a result, RBCE not only mitigates class imbalance but also equips the network with structured, category-aware discriminative capability, improving robustness in challenging multi-class scenarios.

E. Class-Guided Decoder

In conventional multi-class classification, the standard paradigm maps the encoder output directly to a classifier, while decoders are primarily used for reconstruction or generation. Sea state classification, however, differs: its continuous spectrum and ambiguous boundaries make a single compressed feature insufficient to encode discriminative information. To address this, we redefine the decoder as a class guided mechanism, transforming feature reconstruction into a conditional discriminative task that constructs independent and separable subspaces for each sea state.

At the n-th decoder layer, the output from the previous layer $x^{(n+1)} \in \mathbb{R}^{B \times C \times L}$, where B is the batch size, C the number of channels, and L the temporal length, is fused with encoder features $h^{(n)} \in \mathbb{R}^{B \times C \times L_n}$ via a fusion operator $\Phi^{(n)}(\cdot,\cdot)$ that includes upsampling and channel concatenation to align cross-scale contextual information:

$$x_{\text{cat}}^{(n)} = \Phi^{(n)}(x^{(n+1)}, h^{(n)}), \quad (6)$$

The RBCE module generates a class-conditional embedding vector $CCV \in \mathbb{R}^d$ (with d the embedding dimension), which encodes category-sensitive priors that guide the decoder. This vector is projected via a layer-specific function $f_\theta^{(n)}(\cdot)$ into scale $\gamma^{(n)}(CCV)$ and shift $\beta^{(n)}(CCV)$ parameters:

$$\begin{aligned} f_\theta^{(n)}(CCV) &= [\gamma^{(n)}(CCV), \beta^{(n)}(CCV)] \\ &= W^{(n)} \, \sigma\big(U^{(n)} CCV + b^{(n)}\big), \end{aligned} \quad (7)$$

where $W^{(n)}$ and $U^{(n)}$ are learnable projection matrices, $b^{(n)}$ is the bias vector, and $\sigma(\cdot)$ is a non-linear activation function such as GELU.

The fused feature cat is then modulated via class-conditional normalization, a technique also known as a Feature-wise Linear Modulation (FiLM) layer:

$$x^{(n)} = \gamma^{(n)}(CCV) \odot Norm(x_{\text{cat}}^{(n)}) + \beta^{(n)}(CCV), \quad (8)$$

where $\odot$ denotes element-wise multiplication, and Norm$(\cdot)$ is a standard normalization operator (e.g., LayerNorm). This process effectively applies a class-dependent projection $\Pi_c : \mathbb{R}^d \to H_c$, mapping the fused feature into a conditional subspace H_c defined by the class prior CCV.

Theoretically, the class-conditional injection is designed to align the decoder's features with the Fisher discriminant criterion, which aims to maximize feature separability:

$$\max \frac{tr(S_B)}{tr(S_W)}, \quad (9)$$

where S_B and S_W denote the between-class and within-class scatter matrices. Maximizing tr$(S_B)/$tr(S_W) enhances the separability of rare classes, making decision boundaries sharper and improving robustness in ambiguous sea states.

Unlike conventional encoder–classifier paradigms, our approach elevates classification to a hierarchical reasoning process of encode–fuse–condition–discriminate, where the decoder is explicitly guided by class priors. This structurally enforces discriminative representations and improves generalization in complex environments.

IV. EXPERIMENTS

A. Experimental Settings

1) Implementation settings: The model was developed using the PyTorch framework (v2.4.1), and all experiments were conducted on a system equipped with a single NVIDIA GeForce RTX 4090 GPU featuring 24 GB of dedicated memory. The channel dimension for all convolutional layers in the network was set to 128. The model was trained for 500 epochs using a learning rate of 1e-4. To ensure the reliability of the results, each experiment was repeated five times.

2) Datasets:

a) Public datasets: For performance evaluation, we selected the publicly available UEA time series classification archive [12]. This archive encompasses a diverse range of domains, including human activity recognition, EEG/ECG classification, and numerical recognition. The datasets within it exhibit significant variations in data dimensionality and class distribution, providing a robust benchmark for effectively evaluating the model's generalization capabilities.

b) Ship motion dataset: The ship motion dataset was generated using the Marine Systems Simulator (MSS) toolbox [13], covering seven sea state levels, with the first three merged due to their similarity. Sample sizes were assigned according to historical occurrence probabilities, yielding more samples for common states (e.g., 3 and 4) and fewer for rare ones such as state 6. Two datasets, Worldwide and North Atlantic, were then constructed. Each sample contains four input features: heave velocity, pitch angle, pitch velocity, and yaw angle [8]. The data was finally partitioned into training, validation, and test sets with a 70%, 10%, and 20% split.

3) Evaluation metrics: For the public datasets, Accuracy was used as most are class-balanced. For the ship motion dataset, which is class imbalanced, Precision (P) Recall (R) and F1-score (F1) were employed.

B. Performance Comparison on Public Dataset

To rigorously evaluate the performance of the proposed approach, we benchmark it against eleven representative time series classification methods drawn from recent literature, namely: EDI [12], [14], DTWI [14], DTWD [14], MLSTM-FCNs (MF) [15], WEASEL + MUSE (WM) [16], Negative Sampling (NS) [17], TapNet (TN) [18], ShapeNet (SN) [19], ImbalanceSSE (ISSE) [9], ASK-ProtoNet (ASK) [10] and KATN [20]. TABLE I Performance comparison results directly reported from ASK-ProtoNet [10] under the same experimental protocol.

As shown in Table I, CRUISE achieves an average accuracy of 76.4% on the UEA archive, outperforming the next-best method, KATN, by 1.1%. It attains the best or joint-best performance on 17 datasets. The model demonstrates particularly pronounced gains on specific datasets, such as LSST, MotorImagery, and Heartbeat, with significant improvements of 4.0%, 3.0%, and 3.0%, respectively. These results underscore its broad application potential in domains like speech and medicine. This advantage is primarily attributed to the synergistic effect of its multi-scale temporal modeling and class-enhancement guidance mechanism.It should be noted, however, that the model's superiority is not universal across all scenarios. For instance, it ranks only fourth on the InsectWingbeat dataset and lags significantly behind the top-performing method on the ERing dataset. This suggests that there is still room for improvement, particularly in handling high-frequency sampled data and in certain specialized application tasks.

C. Comparison with Baseline Methods on Ship Motion Datasets

To assess the effectiveness of the proposed model for SSE, we compare it against ten advanced methods, including classical time series classifiers and recent deep learning-based SSE models: MLP [21], CNN [22], LSTM-FCNs [15], TapNet [18], SeaStateNet [23], SSENET [8], SpectralSeaNet [24], a data-driven Hybrid model [25], ImbalanceSSE [9] and ASK-ProtoNet [10].

As shown in TABLE II, CRUISE achieves the best performance on both the Worldwide and North Atlantic datasets. It surpasses the runner-up ASK-ProtoNet with relative F1-score gains of 2.12% and 3.03% over ASK-ProtoNet, and also attains the highest recall, demonstrating strong capability in recognizing all sea state categories, especially rare ones. The performance gap among models reflects their architectural differences. MLP and CNN fall short due to limited feature learning, while ASK-ProtoNet benefits from multi-scale fusion and prototype updating but remains constrained by static similarity measures, leading to reduced sensitivity to rare or ambiguous classes. In contrast, CRUISE introduces two key innovations: the RBCE module, which recalibrates class distributions to strengthen minority representations, and the class-guided decoder, which preserves class information during hierarchical feature restoration. Together, these mechanisms sharpen decision boundaries and enhance robustness, explaining CRUISE's consistent superiority over ASK-ProtoNet.

D. Comparison with Class-imbalanced Methods on Ship Motion Dataset.

To rigorously assess our model's efficacy in tackling class imbalance, we benchmarked it against six methodologies on two ship motion datasets, including ClassBalancedLoss (CBLoss) [26], WeightedCE, LDAMLoss [27], FocalLoss [28], PFLoss and Balanced softmax loss (BFLoss)[29]. For the purpose of a fair and controlled experiment, a consistent model architecture and hyperparameter configuration were maintained across all methods. Crucially, this ablation study involved removing our proposed Rarity Boosting module components and solely substituting the loss function for each baseline.

The results, presented in Table III, unequivocally demonstrate the marked advantage of our proposed method in managing class imbalance. CRUISE attained the highest F1-scores on both the World Wide and North Atlantic datasets, decisively outperforming all competing approaches. While certain loss functions, such as PFLoss, occasionally yielded higher recall, this was typically achieved by compromising precision. In contrast, CRUISE not only secured the highest precision across both datasets but also achieved the optimal trade-off between precision and recall, underscoring its robust and well-rounded performance.

This superior outcome reveals a fundamental limitation of conventional approaches: methods that depend exclusively on manipulating loss weights or sample contributions are subject to an intrinsic performance cap. The excellence of CRUISE originates from its strategy of resolving the imbalance issue at the more foundational feature representation level. By leveraging its RBCE module and class-guided decoder, the model proactively reinforces the features of minority classes and guides the network to focus on hard-to-distinguish class boundaries prior to the final prediction. This ablation study confirms that such a synergistic architectural design is a more fundamental and potent solution than merely applying an advanced loss function, significantly elevating the model's

TABLE I: Performance comparison with baseline methods on UEA archive (%)

Dataset	EDI	DTWI	DTWD	MF	WM	NS	TN	SN	ASK	ISSE	KATN	Ours
ArticularyWordRecognition	97	98	98.7	97.3	99	98.7	98.7	98.7	99	**99.3**	**99.3**	**99.3**
AtrialFibrillation	26.7	26.7	22	26.7	33.3	13.3	33.3	40	46.7	**66.7**	53.3	**66.7**
BasicMotions	67.6	**100**	97.5	95	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
CharacterTrajectories	96.4	96.9	98.9	98.5	99	99.4	**99.7**	98	99.2	**99.7**	99.3	99.2
Cricket	94.4	98.6	**100**	91.7	**100**	98.6	95.8	98.6	**100**	**100**	**100**	**100**
DuckDuckGeese	27.5	55	60	67.5	57.5	67.5	57.5	**72.5**	70	40	66	70
EigenWorms	54.9	N/A	61.8	50.4	89	87.8	48.9	87.8	58	**93.1**	73.3	87.8
Epilepsy	66.6	97.8	96.4	76.1	**100**	95.7	97.1	98.7	**100**	98.6	91.3	99.3
ERing	13.3	13.3	13.3	13.3	13.3	13.3	13.3	20	30	30	**95.6**	40
EthanolConcentration	29.3	30.4	32.3	37.3	**43**	23.6	32.3	31.2	38.8	33.1	39.9	39.2
FaceDetection	51.9	N/A	52.9	54.5	54.5	52.8	55.6	60.2	57.5	58.9	66.4	**67.5**
FingerMovements	55	52	53	58	49	54	53	58	60	67	62	**68**
HandMovementDirection	27.8	30.6	23.1	36.5	36.5	27	37.8	33.8	40.5	46	**68.8**	52.7
Handwriting	20	31.6	28.6	28.6	60.5	53.3	35.7	45.1	60.5	62.4	32	**63.2**
Heartbeat	61.9	65.8	71.7	66.3	72.7	73.7	75.1	75.6	76.1	78.5	76.20	**81.5**
InsectWingbeat	12.8	N/A	N/A	16.7	N/A	16	20.8	25	16.2	17.9	**51.5**	18.6
JapaneseVowels	92.4	95.9	94.9	97.6	97.3	98.9	96.5	98.4	**99.5**	98.7	96.8	99.2
Libras	83.3	89.4	87	85.6	87.8	86.7	85	85.6	91.7	93.9	83.9	**95.6**
LSST	45.6	57.5	55.1	37.3	59	55.8	56.8	59	64.4	63.5	40.7	**68.6**
MotorImagery	51	N/A	50	51	50	54	59	61	63	64	62	**67**
NATOPS	85	85	88.3	88.9	87	94.4	93.9	88.3	96.1	98.8	88.9	**98.9**
PEMS-SF	70.5	73.4	71.1	69.9	N/A	68.8	75.1	75.1	86.7	87.3	**95.3**	89
PenDigits	97.3	93.9	97.7	97.8	94.8	98.3	98	97.7	**99.2**	**99.2**	98.2	**99.2**
Phoneme	10.4	15.1	15.1	11	19	24.6	17.5	29.8	23.3	31.6	**36.6**	20.6
RacketSports	86.8	84.2	80.3	80.3	**93.4**	86.2	86.8	88.2	88.1	91.5	87.5	89.5
SelfRegulationSCP1	77.1	76.5	77.5	87.4	71	84.6	65.2	78.2	88.7	90.1	90.8	**91.5**
SelfRegulationSCP2	48.3	53.3	53.9	47.2	46	55.6	55	57.8	58.9	61.1	60	**63.3**
SpokenArabicDigits	96.7	95.9	96.3	99	98.2	95.6	98.3	97.5	99.4	**99.6**	98.3	99.1
StandWalkJump	20	33.3	20	6.7	33.3	40	40	53.3	**66.7**	**66.7**	60	**66.7**
UWaveGestureLibrary	88.1	86.8	90.3	89.1	91.6	88.4	89.4	90.6	90.9	91.9	88.1	**92.2**
Average accuracy	58.5	66.8	65.1	62.1	69.1	66.9	65.7	69.9	72	74.3	75.4	**76.5**
Wins/Ties	0	1	1	0	5	1	2	2	6	9	8	**17**

TABLE II: Performance evaluation with baseline methods on two SSE datasets (%).

	World Wide			North Atlantic		
	P	R	F1	P	R	F1
MLP	66.18	58.01	59.43	68.95	58.97	60.87
CNN	72.05	55.85	57.94	71.72	55.72	59.77
LSTM-FCNs	80.16	71.98	74.31	75.66	75.47	75.09
TapNet	72.48	62.14	64.31	73.76	63.25	65.13
SpectralSeaNet	73.46	63.76	67.89	73.58	64.93	66.45
Hybrid	75.13	65.12	68.91	75.57	63.87	69.32
SeaStateNet	77.13	68.97	71.57	76.23	70.12	72.43
SSENET	79.02	70.14	75.75	77.46	70.74	73.03
ImbalanceSSE	78.68	78.22	78.38	77.01	74.46	75.46
ASK-ProtoNet	**84.57**	77.06	79.82	80.05	76.67	77.92
Ours	83.59	**80.06**	**81.51**	**82.19**	**78.73**	**80.28**

TABLE III: Performance comparison with state-of-the-art methods of imbalanced learning on ship motion datasets (%)

	World Wide			North Atlantic		
	P	R	F1	P	R	F1
CBLoss	81.82	79.81	80.15	80.44	79.04	79.49
WeightedCE	81.37	79.84	80.48	80.47	79.84	80.12
LDAMLoss	82.60	79.19	80.5	81.21	79.69	80.23
FocalLoss	81.00	74.97	80.03	78.18	80.17	79.07
PFLoss	78.76	80.47	79.48	78.57	**80.44**	79.41
BSLoss	80.75	78.64	79.47	78.23	78.29	78.28
Ours	83.59	**80.06**	**81.51**	**82.19**	78.73	**80.28**

ability to navigate class-imbalanced scenarios.

E. Ablation and Sensitivity Analysis

To assess the contribution of each module, we define three ablation variants of our model:

- **rRB**: The RB module is removed, while all other modules remain intact.
- **rRBCE**: The RBCE module is removed, with all other components unchanged.
- **rCGD**: The class-guided decoder (CGD) module is removed, while keeping the rest of the architecture unchanged.

As shown in Fig. 3, removing any module results in

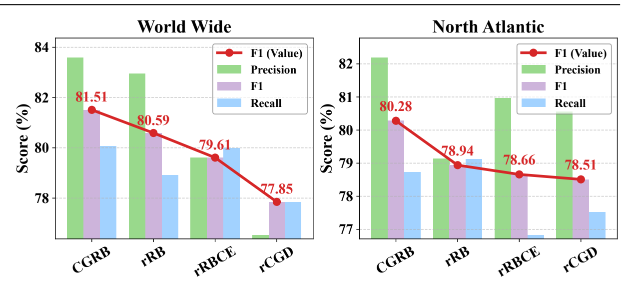

Fig. 3: Ablation study.

performance degradation, confirming the necessity of each component. Removing the RB module (rRB) leads to only a slight drop, indicating its primary role in enhancing minority class recognition. In contrast, removing the RBCE bottleneck (rRBCE) causes more significant degradation, highlighting the importance of the synergy between rare-class enhancement and class-conditional vectors for global discrimination. The most substantial loss occurs when the CGD module is removed, where the F1 score on the Worldwide dataset drops to 77.85, underscoring its central role in feature reconstruction and maintaining class consistency. In summary, RB provides minority class compensation, RBCE strengthens global–class synergy, and CGD serves as the key driver of overall performance.

To assess the influence of rare boosting and class embedding, we evaluated the effects of RB strength and class embedding dimension on precision, recall, and F1 across two datasets.

As shown in the Fig. 4, the model exhibits some sensitivity to the rare-boosting magnitude, but performance remains stable within a moderate range. Properly enhancing minority-class features can effectively improve overall discriminative ability, whereas excessively high or low boosting yields no additional benefit, indicating that the model is robust to this

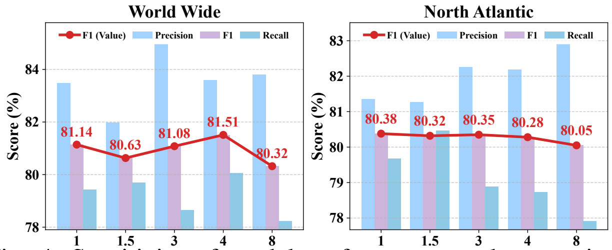

Fig. 4: Sensitivity of model performance to the magnitude of rare-boosting across two SSE datasets.

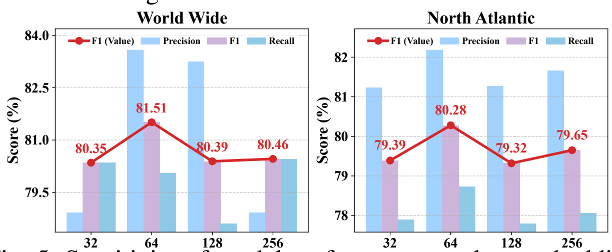

Fig. 5: Sensitivity of model performance to class embedding across two SSE datasets.

parameter.

As shown in the Fig. 5, the dimension of class embeddings also affects model performance. Embeddings that are too small limit the expression of class information, while overly large embeddings may introduce redundancy or overfitting. The results indicate that a moderate embedding dimension optimizes overall performance while maintaining strong feature discriminability.

V. DISCUSSION

A. Comparison With Baseline Methods on Ship Movement Datasets

To rigorously evaluate the model's efficacy and robustness in a real-world operational context, we conducted experiments on the publicly available Ship Movement Dataset [30]. This dataset provides a highly authentic and challenging testbed, as it comprises motion records of 45 diverse vessels moored at the Outer Port of Punta Langosteira, A Coruña, Spain, spanning from 2015 to 2020. Authoritative, concurrently recorded significant wave height data serves as the ground truth for ambient sea conditions, enabling a direct assessment of the model's practical applicability.

The preprocessing pipeline was designed to mirror the challenges of onboard robotic systems dealing with inconsistent sensor data. Missing values, a common issue in field deployments, were handled via imputation. Subsequently, to simulate a higher frequency perception cycle suitable for real-time decision-making, we upsampled the original 1-hour data to a uniform 15-minute interval, ensuring temporal consistency across the dataset. For our comparative analysis, this augmented dataset was then randomly partitioned into training (80%) and testing (20%) sets.

The model was trained using a batch size of 128 and an Adam optimizer with a learning rate of 5e-5. We set the maximum number of training epochs to 100. To prevent overfitting and ensure the model captured generalizable features, an early stopping mechanism was employed: training was halted if the F1 score on the test set did not improve for 10 consecutive epochs after the 50th epoch.

TABLE IV: Performance evaluation on the Punta Langosteira dataset.

Models	Punta Langosteira		
	P	R	F1
MLP	39.44	45.02	41.20
CNN	42.58	47.52	43.98
FCN	81.44	79.04	79.37
LSTM-FCNs	78.20	75.30	74.99
SeaStateNet	62.97	63.13	60.70
SSENET	90.41	92.03	91.09
CSM-SSE	87.98	89.70	88.24
PKCN	90.42	91.34	90.31
Ours	**94.72**	**94.77**	**94.52**

As demonstrated by the experimental results, our proposed method significantly outperforms the strongest baseline, PKCN, improving the F1-score, Precision, and Recall by 4.21, 4.30, and 3.43 percentage points, respectively. This superior performance is attributed to the synergistic interplay of multi-scale temporal modeling, class-guided decoding, and rare-class enhancement. These components collectively bolster recognition accuracy, maintain class balance, and enhance robustness against the complexities of real-world sea conditions. The findings underscore the practical relevance and superior efficacy of our approach in operational maritime scenarios.

B. Real-Time Inference

To bridge the gap between offline evaluation and real-world application, we validated the model's real-time performance on a representative edge computing platform, an NVIDIA Jetson TX2, which is commonly used in mobile robotics. This experiment was designed to assess the model's responsiveness and stability when processing a continuous stream of sensor data from an integrated IMU, simulating an onboard operational scenario. As illustrated in Fig. 6, the continuous motion data was fed directly into our model for inference, with the resulting sea state classifications visualized in real time through a custom GUI. This setup allowed for a direct evaluation of the algorithm's practical utility in a dynamic environment.

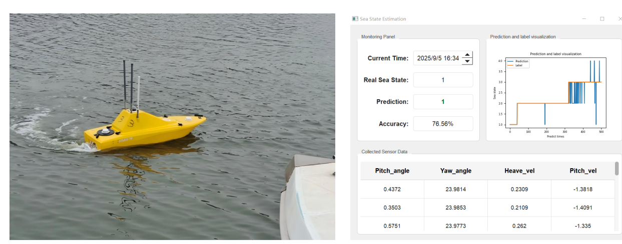

Fig. 6: The specific steps for implementing the deployment of our model.

The system demonstrated strong robustness during the test, achieving a predictive accuracy of 76.64% on the live, continuous data stream. The slight discrepancy from our offline results is primarily attributable to the inherent distributional shift between the evolving test data and the static training set, a common challenge in real-world robotic applications. In conclusion, this validation confirms our algorithm's ability to meet real-time processing demands and provides a feasibility validation for its practical deployment on resource-constrained, onboard systems.

VI. CONCLUSIONS AND FUTURE WORK

To address the three primary challenges in sea state estimation—class imbalance, indistinct decision boundaries, and inadequate multi-scale feature extraction—this paper proposes a novel deep learning framework, CRUISE. Its core RBCE module and class-guided decoder reshape class distributions at the feature level and sharpen discriminative boundaries. Extensive experiments on public and real-world datasets, along with real-time tests on a physical model ship, demonstrate that our method outperforms existing state-of-the-art models in overall performance, showing significant advantages, particularly in identifying rare sea states, thus laying the foundation for onboard applications. Future work will focus on two aspects: first, conducting full-scale ship trials across various vessel types and complex sea conditions to enhance the model's cross-scenario generalization and real-time accuracy; second, advancing model lightweighting research to meet the deployment requirements of embedded platforms, thereby accelerating the implementation of intelligent sea state awareness in the maritime domain.

ACKNOWLEDGMENT

This work was partially supported by the National Natural Science Foundation of China under Grants T2422015, 62306212, and the Beijing-Tianjin-Hebei Natural Science Foundation Cooperation Project under Grant 25JJJJC0009. Xu Cheng also gratefully acknowledges support from the Marie Skłodowska-Curie Actions (MSCA) under Project No. 101111188.

REFERENCES

[1] N. Ahmad, R. A. R. Ghazilla, N. M. Khairi, and V. Kasi, "Reviews on various inertial measurement unit (imu) sensor applications," *International Journal of Signal Processing Systems*, vol. 1, no. 2, pp. 256–262, 2013, https://doi.org/10.12720/ijsps.1.2.256-262.

[2] H. Majidian, L. Wang, and H. Enshaei, "Part. a: A review of the real-time sea-state estimation, using wave buoy analogy," *Ocean Engineering*, vol. 266, p. 111684, 2022.

[3] T. Zhang and Z. Ren, "Restricted isometry property in wave buoy analogy and application to multispectral fusion," *IEEE Transactions on Intelligent Transportation Systems*, 2025.

[4] J. Reis, P. Batista, P. Oliveira, and C. Silvestre, "Discrete-time kalman filter for heave motion estimation," *Ocean Engineering*, vol. 277, p. 114240, 2023.

[5] D.-S. Kwon, S.-J. Kim, C. Jin, and M. Kim, "Parametric estimation of directional wave spectra from moored fpso motion data using optimized artificial neural networks," *Journal of Marine Science and Engineering*, vol. 13, no. 1, p. 69, 2025, https://doi.org/10.3390/jmse13010069.

[6] J. Berbić, E. Ocvirk, D. Carević, and G. Lončar, "Application of neural networks and support vector machine for significant wave height prediction," *Oceanologia*, vol. 59, no. 3, pp. 331–349, 2017, https://doi.org/10.1016/j.oceano.2017.03.007.

[7] P. Han, G. Li, S. Skjong, B. Wu, and H. Zhang, "Data-driven sea state estimation for vessels using multi-domain features from motion responses," in *2021 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2021, pp. 2120–2126.

[8] X. Cheng, G. Li, A. L. Ellefsen, S. Chen, H. P. Hildre, and H. Zhang, "A novel densely connected convolutional neural network for sea-state estimation using ship motion data," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 9, pp. 5984–5993, 2020.

[9] X. Cheng, K. Wang, X. Liu, Q. Yu, F. Shi, Z. Ren, and S. Chen, "A novel class-imbalanced ship motion data-based cross-scale model for sea state estimation," *IEEE Transactions on Intelligent Transportation Systems*, 2023.

[10] Z. Xu, M. Liu, T. Li, X. Liu, X. Cheng, and S. Chen, "Leveragable adaptive multi-scale features and learnable prototypes for imbalanced sea state estimation based on ship motion data," *IEEE Transactions on Intelligent Transportation Systems*, 2025.

[11] T. I. Fossen, "Handbook of marine craft hydrodynamics and motion control," *John Willy & Sons Ltd*, 2011.

[12] A. Bagnall, H. A. Dau, J. Lines, M. Flynn, J. Large, A. Bostrom, P. Southam, and E. Keogh, "The uea multivariate time series classification archive, 2018," *arXiv preprint arXiv:1811.00075*, 2018.

[13] T. Perez and T. I. Fossen, "A matlab toolbox for parametric identification of radiation-force models of ships and offshore structures," *Modeling, Identification and Control*, 2009.

[14] M. Shokoohi-Yekta, J. Wang, and E. Keogh, "On the non-trivial generalization of dynamic time warping to the multi-dimensional case," in *Proceedings of the 2015 SIAM international conference on data mining*. SIAM, 2015, pp. 289–297.

[15] F. Karim, S. Majumdar, H. Darabi, and S. Harford, "Multivariate lstm-fcns for time series classification," *Neural Networks*, vol. 116, pp. 237–245, 2019.

[16] P. Schäfer and U. Leser, "Multivariate time series classification with weasel+ muse," *arXiv preprint arXiv:1711.11343*, 2017.

[17] J.-Y. Franceschi, A. Dieuleveut, and M. Jaggi, "Unsupervised scalable representation learning for multivariate time series," *Advances in neural information processing systems*, vol. 32, 2019.

[18] X. Zhang, Y. Gao, J. Lin, and C.-T. Lu, "Tapnet: Multivariate time series classification with attentional prototypical network," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, 2020, pp. 6845–6852.

[19] G. Li, B. Choi, J. Xu, S. S. Bhowmick, K.-P. Chun, and G. L.-H. Wong, "Shapenet: A shapelet-neural network approach for multivariate time series classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 9, 2021, pp. 8375–8383.

[20] Z. Xiao, H. Xing, R. Qu, H. Li, H. Tong, S. Luo, J. Song, L. Feng, and Q. Wan, "Knowledge aggregation transformer network for multivariate time series classification," *IEEE Transactions on Big Data*, 2025.

[21] H. Miao, Y. Guo, G. Zhong, B. Liu, and G. Wang, "A novel model of estimating sea state bias based on multi-layer neural network and multi-source altimeter data," *European Journal of Remote Sensing*, vol. 51, no. 1, pp. 616–626, 2018.

[22] J. N. Alfsen, "Imu-based sea state estimation using convolutional neural networks for dp vessels," Master's thesis, NTNU, 2020.

[23] X. Cheng, G. Li, R. Skulstad, S. Chen, H. P. Hildre, and H. Zhang, "Modeling and analysis of motion data from dynamically positioned vessels for sea state estimation," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 6644–6650.

[24] X. Cheng, G. Li, R. Skulstad, H. Zhang, and S. Chen, "Spectralseanet: Spectrogram and convolutional network-based sea state estimation," in *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 2020, pp. 5069–5074.

[25] P. Han, G. Li, X. Cheng, S. Skjong, and H. Zhang, "An uncertainty-aware hybrid approach for sea state estimation using ship motion responses," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 2, pp. 891–900, 2021.

[26] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9268–9277.

[27] K. Cao, C. Wei, A. Gaidon, N. Arechiga, and T. Ma, "Learning imbalanced datasets with label-distribution-aware margin loss," in *Advances in Neural Information Processing Systems*, 2019, pp. 1567–1578.

[28] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[29] J. Ren, C. Yu, X. Ma, H. Zhao, S. Yi *et al.*, "Balanced meta-softmax for long-tailed visual recognition," *Advances in neural information processing systems*, vol. 33, pp. 4175–4186, 2020.

[30] A. Alvarellos, A. Figuero, H. Carro, R. Costas, J. Sande, A. Guerra, E. Peña, and J. Rabuñal, "Machine learning based moored ship movement prediction," *Journal of Marine Science and Engineering*, vol. 9, no. 8, p. 800, 2021.