RESEARCH ARTICLE

Data-Driven Operational Assessment Method and Digital Twin System for Unmanned Surface Vehicles

Yuting Bai[1,*], **Jiyuan Hu**[1], **Eziz Tursun**[2] **and Hurxida Yimit**[2]

[1] College of Computer and Artificial Intelligence, Beijing Technology and Business University, Beijing 100048, China
[2] School of Computing and Artificial Intelligence, Xinjiang Hetian College, Hetian 848000, China

Abstract

To address the challenge of effectively leveraging multi-source data for automated operational assessment of Unmanned Surface Vehicles (USVs) and utilizing digital technologies for monitoring and control, this paper proposes a data-driven state assessment method for surface unmanned systems and develops a digital twin system tailored for USVs. First, a dual-channel feature modeling mechanism is constructed by integrating physically interpretable statistical features with temporal convolutional features. Second, a complementary modeling strategy is adopted using CatBoost for static classification and GRU for dynamic modeling, while a Covariance Intersection (CI) fusion strategy is introduced to enhance the classification performance and adaptability of the model. Finally, a digital twin system is designed that incorporates Position Estimation, Attitude Estimation, and State Evaluation, enabling real-time monitoring and multidimensional visualization of USV operational states. Experimental results demonstrate that the proposed method outperforms baseline approaches in terms of accuracy, F1-score, and other key metrics, exhibiting strong generalization capability and promising potential for practical deployment.

Keywords: unmanned surface vehicles, data-driven, state evaluation, digital twin, multi-source temporal modeling, model fusion.

1 Introduction

With the rapid advancement of intelligent shipping and maritime automation, Unmanned Surface Vehicles (USVs) have shown great potential in various applications such as environmental monitoring, maritime patrol, and emergency search and rescue. As task complexity increases, USVs must operate under diverse and dynamic environmental conditions—such as wave disturbances, current velocities, and wind speeds—which significantly impact their operational stability and navigational safety. Therefore, achieving accurate, real-time, and visualized assessment of USV operational states has become a key technological challenge for ensuring system stability and task success.

Traditional assessment approaches often rely on static rules, empirical thresholds, or simple statistical features, which are inadequate for effectively analyzing system behavior under dynamic environments. Particularly in scenarios involving wave disturbances or abrupt state changes, such methods are prone to misjudgments or omissions. In recent years,

Submitted: 21 May 2025
Accepted: 24 December 2025
Published: 08 January 2026

Vol. 2, **No.** 1, 2026.
10.62762/TMI.2025.444910

***Corresponding author:**
✉ Yuting Bai
baiyuting@btbu.edu.cn

Citation
Bai, Y., Hu, J., Tursun, E. & Yimit, H. (2026). Data-Driven Operational Assessment Method and Digital Twin System for Unmanned Surface Vehicles. *ICCK Transactions on Machine Intelligence*, 2(1), 38–52.

data-driven methods have been increasingly applied to state evaluation tasks and have demonstrated strong modeling capabilities. However, purely data-driven deep learning models are susceptible to generalization failure when encountering scenarios outside the distribution of the training data. Several studies indicate that incorporating statistical features with explicit physical interpretations can effectively constrain the solution space of the model, thereby significantly enhancing system robustness under sparse sample conditions [1, 2]. Nevertheless, the majority of existing end-to-end models predominantly ingest raw sensor data directly, neglecting these critical physical priors.

To address these challenges, this study proposes a state evaluation method for surface unmanned systems based on dual-channel feature extraction and temporal model fusion. Building on this, a digital twin system for USVs is developed. The system integrates interactive modules for sensor data acquisition, pose estimation, and state evaluation, supporting real-time monitoring, dynamic modeling, and comprehensive assessment throughout the entire operational process of USVs. Compared to conventional evaluation frameworks that rely solely on algorithmic outputs, the digital twin system enables high-frequency closed-loop interaction between the physical USV and its virtual counterpart, thereby improving the timeliness, controllability, and intelligence level of operational state monitoring.

2 Related Work

2.1 Data-Driven Methods for State Evaluation

In recent years, data-driven state evaluation has emerged as a pivotal research paradigm for intelligent autonomous systems. Foundational methodologies dealing with uncertainty and temporal dynamics have been substantially validated in aerial platforms and industrial robotics. Regarding the handling of epistemic ambiguity in multi-source information, Zhang et al. [3] established a robust assessment framework for industrial robots by leveraging Dempster-Shafer evidence theory and fuzzy logic. In the domain of Unmanned Aerial Systems (UAS), Chen et al. [4] and Qiao et al. [5] developed hierarchical health assessment architectures for flight control and electromechanical subsystems, providing theoretical references for mobile platform monitoring. Furthermore, deep learning approaches have demonstrated superior efficacy in feature engineering: Wei et al. [6] and Xiao et al. [7] validated the capability of Long Short-Term Memory (LSTM) networks and Deep Belief Networks (DBNs) in capturing latent temporal degradation patterns. Similarly, Ren et al. [8] and Zhang et al. [9] proved that Transfer Learning and Random Forests could effectively mitigate model overfitting and optimize feature weight allocation in complex equipment under varying working conditions.

Notwithstanding these advancements, the operational environment of Unmanned Surface Vehicles (USVs) imposes distinct challenges characterized by strong hydrodynamic nonlinearities and stochastic environmental disturbances. Consequently, recent maritime research has necessitated domain-specific adaptations focusing on three dimensions: dynamic modeling under wave excitation, robust multi-sensor fusion, and hybrid architecture evaluation.

In the context of temporal modeling under wave environments, research aims to decouple environmental interference from system dynamics. Qu et al. [10] proposed a hybrid scheme coupling Digital Twins with an improved Unscented Kalman Filter (UKF) to enhance navigation attitude prediction fidelity in dynamic waves. To model the spatiotemporal dependencies inherent in real-ship data, Zhang et al. [11] constructed a deep learning-based combined framework for ship motion attitude. Additionally, Wang et al. [12] leveraged a premonition-driven deep learning model to predict short-term violent roll motions, effectively utilizing hull attitude precursors to capture rapid dynamic changes. Addressing the control challenges under such disturbances, van der Saag et al. [14] validated an Active Disturbance Rejection Control (ADRC) scheme through field experiments, while Cao et al. [13] quantitatively analyzed optimal formation control under position constraints and time-varying yaw limitations, highlighting the generalization limitations of static models in severe sea states.

Regarding multi-sensor fusion, Zhao et al. [1] designed a probabilistic framework integrating end-to-end deep learning with Bayesian inference. Building on this, Huang et al. [15] proposed an "uncertainty-aware" deep distributed reinforcement learning algorithm, achieving robust navigation in complex environments by explicitly modeling the uncertainty in sensor channels. Moreover, Elsanhoury et al. [16] substantiated the reliability of resilient navigation using novel LEO-based fusion positioning in GNSS-denied environments, confirming the critical role of multi-modal redundancy in ensuring system availability.

To address complex coupling effects, the field is evolving towards hybrid architectures. Saptoe et al. [17] demonstrated the potential of deep learning-based visual perception for obstacle avoidance in water quality monitoring tasks. Notably, to resolve the interpretability bottleneck, Xu et al. [2] introduced Physics-Informed Neural Networks (PINN) into USV dynamics prediction. Similarly, the hybrid prediction models proposed by Cen et al. [18] and the data-driven dynamics analysis by Liu et al. [19] have validated the complementary advantages of combining attention mechanisms with CNN-GRU features for high-fidelity state estimation and operational analysis.

However, a critical gap remains in the optimal integration of these methodologies. While hybrid architectures like heterogeneous ensembles have validated the potential of feature combination, they often treat structured and unstructured data as homogeneous inputs, lacking a mechanism to explicitly leverage their complementary nature—specifically, the physical interpretability of statistical features versus the dynamic capturing capability of deep learning. Furthermore, regarding multi-sensor fusion, most existing frameworks typically assume independence among data sources, often failing to address the unknown correlations induced by coupled environmental disturbances. These limitations directly motivate the design of the proposed Dual-Channel architecture (combining CatBoost and GRU) and the adoption of the Covariance Intersection (CI) fusion strategy in this study, aiming to bridge the gap between physical reliability and data-driven precision.

2.2 Advances in the Application of Digital Twin Systems for Unmanned Surface Vehicles

Meanwhile, research on Digital Twin (DT) technology in the field of Unmanned Surface Vehicles (USVs) has been continuously deepening, making it a critical approach to enhancing autonomous operational capability and navigational safety. Hasan et al. [20] proposed a predictive digital twin framework for autonomous surface vessels, focusing on the co-modeling of state prediction and control feedback, and highlighting the role of digital twins in real-time interaction and dynamic response. Raza et al. [21] developed a multi-layered digital twin system architecture for USVs that integrates environmental modeling and behavioral analysis modules to improve operational stability and perception capabilities in complex water environments. Peng et al. [22] systematically reviewed typical applications of digital twins for USVs, covering aspects such as sensor fusion, control strategies, and virtual–physical synchronization, and summarized the current development and technical roadmap of the field. Madusanka et al. [23] designed a USV digital twin system based on a virtual reality platform, exploring modeling methods for interaction and control within simulated environments. Vasconcello et al. [24] combined simulation environments with reinforcement learning to develop a digital twin control system with autonomous learning capabilities, supporting path planning and policy optimization in complex tasks.

While these studies have significantly advanced the fidelity of scene rendering and data synchronization, a gap remains in the deep interpretation of the visualized data. Existing systems typically prioritize the replication of physical kinematics but often lack the capability to directly quantify operational risks. Consequently, operators receive high-fidelity visual feedback without synchronized safety insights. This limitation motivates the system design in this study, which integrates the proposed state evaluation model to extend functionality from passive monitoring to active safety analysis, thereby supporting more informed decision-making.

3 Methodology

3.1 Design of the State Evaluation Method for Unmanned Surface Vehicles

To address the inherent limitation where single-modal features fail to simultaneously capture steady-state distributions and transient evolution, this study adopts a 'dual-channel' hybrid modeling strategy. This design philosophy draws inspiration from recent advancements in multi-modal time-series prediction [1, 18], which demonstrate that processing structured and unstructured data through parallel architectures effectively achieves complementarity among heterogeneous features. Specifically, the proposed framework comprises two parallel branches: a statistical feature extraction channel designed to delineate the distributional boundaries of physical states, and a deep temporal channel focused on capturing dynamic trends under wave disturbances. This dual-pathway approach ensures the completeness of assessment information at the feature level. The

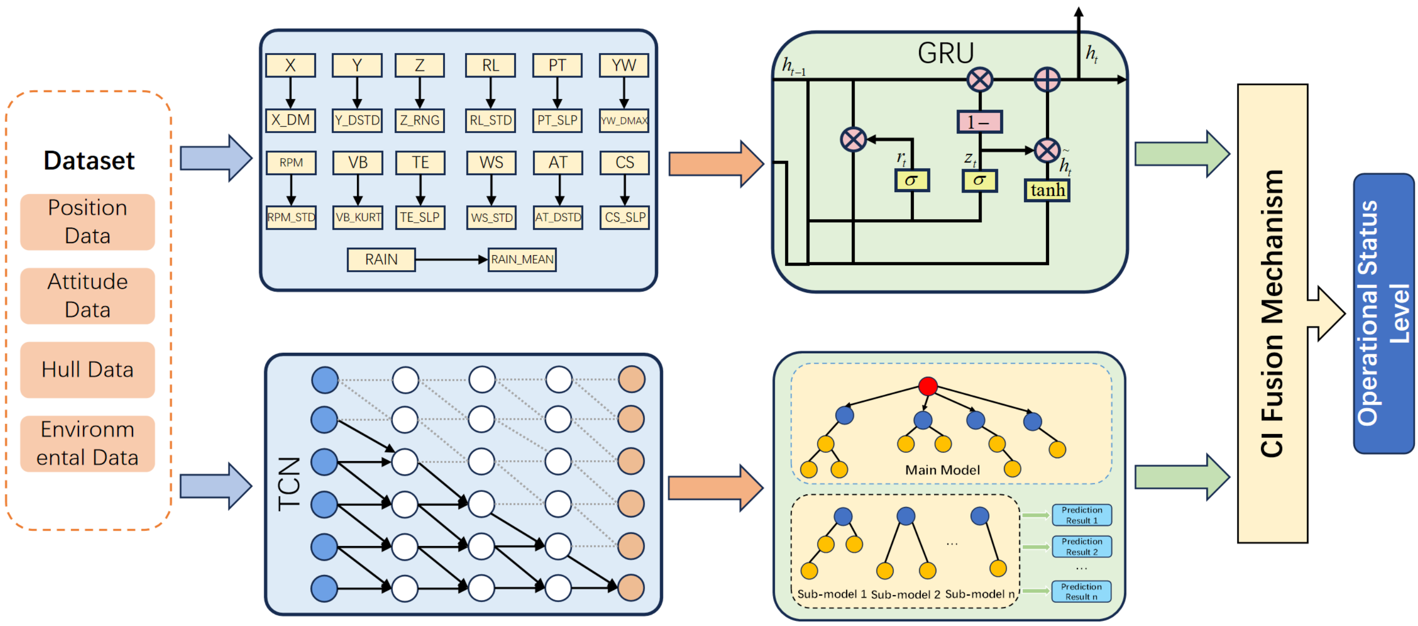

Figure 1. Overall architecture of the proposed model.

overall architecture of the proposed method is illustrated in Figure 1.

To enhance the model's ability to capture both structural characteristics and dynamic patterns from multi-source time-series data, this study proposes a dual-channel feature extraction method that fuses structured and unstructured representations. Based on fixed-size time windows, the raw sensor data is sliced using a sliding window approach and simultaneously fed into two parallel modeling channels. One channel focuses on extracting physically interpretable statistical features, while the other captures temporal evolution patterns that reflect local dynamics. The extracted features from both channels are then concatenated at the feature level to form a unified multi-level fused feature representation, which preserves global trends of system states while enhancing the perception of local fluctuations and transient changes. This design balances feature semantic richness with downstream model compatibility, providing a robust and expressive foundation for the subsequent classification modules.

In the structured channel, the system aims to characterize the physical evolution of the USV during operation. It constructs a set of statistical descriptors based on typical state variables such as position, attitude, velocity, and temperature. These descriptors include first-order difference, linear trend, standard deviation, kurtosis, and skewness. Each metric is associated with a specific physical interpretation: for example, the mean of first-order differences in the X-direction reflects lateral drift tendencies; the standard deviation of the roll angle indicates the intensity of roll motion; and the linear slope of cabin temperature suggests potential thermal instability risks. The corresponding mathematical definitions are given in Equations (1)–(3). These physical-statistical features effectively describe the system's macro-level operational state and environmental responsiveness. They exhibit strong engineering interpretability and are well-suited for integration with traditional operation and maintenance knowledge or expert rule systems, providing practical guidance for applications such as state monitoring, anomaly detection, and risk assessment.

$$X\_DM = \frac{1}{N-1}\sum_{t=2}^{N}(X_t - X_{t-1}) \tag{1}$$

$$RL\_STD = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\mathrm{Roll}_t - \overline{\mathrm{Roll}})^2} \tag{2}$$

$$TE\_SLP = \frac{N\sum t \cdot TE_t - \sum t \cdot \sum TE_t}{N\sum t^2 - (\sum t)^2} \tag{3}$$

where X_t denotes the lateral position at time t, Roll_t denotes the roll angle at time t, TE_t represents the cabin temperature at time t, X_DM denotes the mean of first-order differences in lateral position, RL_STD denotes the standard deviation of roll angle, and TE_SLP denotes the linear trend slope of cabin temperature.

In the unstructured channel, this study employs a Temporal Convolutional Network (TCN) to extract dynamic features from raw multi-dimensional time-series data. As a deep neural network architecture that combines the strengths of convolutional structures with temporal modeling capabilities, TCN enables parallel computation while preserving the sequential order of time-series data for efficient and causal modeling.

Structurally, the network integrates two key mechanisms: causal convolution and dilated convolution. Causal convolution ensures that the output at each time step depends only on the current and past inputs, strictly preserving temporal causality and preventing information leakage from future time points. On top of this, dilated convolution increases the receptive field by inserting gaps within the convolution kernel, allowing the network to model long-range dependencies even with a shallow structure. This design enhances the network's ability to capture long-term trends and periodic patterns in sensor data.

Compared with traditional Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), TCN offers superior training stability and computational efficiency. On the one hand, the use of residual connections and layer normalization improves the trainability of deeper networks; on the other hand, the application of fixed-width sliding window convolutions allows parallel processing of the input sequences, reducing latency and making the model well-suited for deployment in real-time state evaluation systems.

Additionally, TCN naturally supports multi-channel inputs, enabling the unified modeling of joint temporal dynamics across various sensor channels (e.g., speed, vibration, wind speed) within a single network. This avoids the information loss and structural bias that may arise from manually engineered feature combinations. The output of this channel captures high-order nonlinear dynamic patterns within each time window, providing essential deep semantic features to support the downstream state classification model. For a given time series

$$X = \{x_0, x_1, x_2, \ldots, x_{t-1}, x_t\},$$

the dilated convolution operation on each element x_t is defined in Equation (4).

$$C(F, x_t) = \sum_{k=1}^{K} f_k \, x_{t-(k-1)d} \tag{4}$$

where

$$F = \{f_1, f_2, \ldots, f_{K-1}, f_K\}$$

denotes the convolution kernel of size K, and f_k represents the k-th kernel weight. x_t denotes the input sample at time t in the time series, and d is the dilation factor. $C(F, x_t)$ denotes the output after applying dilated convolution to the sequence.

After processing by the structured physical feature channel, the original time-series data is compressed into low-dimensional, semantically clear statistical vectors, which primarily capture the overall trends and evolution patterns of each variable within the sliding window. In parallel, the unstructured channel performs a dimensional expansion of the original multi-dimensional data through the Temporal Convolutional Network (TCN), extracting deep semantic representations that reflect local dynamic changes and nonlinear disturbances in the system. These two channels reconstruct and encode the same time segment from different perspectives—"physical laws" and "temporal dynamics"—and their outputs are concatenated at the feature fusion layer to form a unified multi-level fused feature vector. This representation simultaneously preserves global information about stable structures and provides detailed descriptions of transient behaviors, thereby offering more discriminative inputs for downstream classification models.

To further enhance the model's ability to distinguish fused features, this study adopts a parallel modeling strategy based on model complementarity. Specifically, the fused features are simultaneously input into two types of heterogeneous models for independent learning and inference: one model focuses on static classification of structural attributes, while the other emphasizes temporal modeling under sequential dependencies. This structurally differentiated design introduces diversity into the learning process, enabling the system to capture both static decision boundaries in the feature space and behavioral evolution over the temporal axis. The parallel mechanism effectively addresses the limitation of using a single model to represent multi-attribute features, thereby improving the overall robustness and generalization performance of the state evaluation system.

In the temporal modeling branch, a Gated Recurrent Unit (GRU) network is introduced as the primary modeling tool. Concerning dynamic temporal features, the GRU achieves an optimal balance between real-time performance and computational efficiency. Compared to LSTM networks, the GRU reduces parameter redundancy by simplifying the gating structure. As explicitly formulated by Cho et al. [25], the GRU employs a streamlined gating mechanism to effectively capture long-term dependencies in sequences without the complexity of a separate memory cell. This architectural efficiency enables the model to achieve convergence performance comparable to LSTM but with lower computational overhead, making it highly compatible with the resource-constrained scenarios of USV edge computing terminals.

Structurally, GRU is a lightweight and efficient variant of recurrent neural networks (RNNs), incorporating two key control mechanisms: the update gate and the reset gate, which regulate memory retention and forgetting in an adaptive manner. This architecture effectively alleviates the gradient vanishing and instability issues encountered by traditional RNNs when modeling long sequences, making it particularly suitable for temporal modeling of sensor sequences that exhibit nonstationary behavior during USV operations. GRU not only responds quickly to short-term disturbances but also maintains state continuity for long-term trend modeling. The mathematical formulation of the GRU update process is provided in Equations (5)–(8). By integrating GRU, the model is able to accurately capture the temporal dependencies embedded in the fused features, thereby improving its ability to fit and predict state evolution trajectories.

$$z_t = \sigma(W_z \times [h_{t-1}, x_t]) \tag{5}$$

$$r_t = \sigma(W_r \times [h_{t-1}, x_t]) \tag{6}$$

$$\tilde{h}_t = \tanh(W \times [r_t \times h_{t-1}, x_t]) \tag{7}$$

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times \tilde{h}_t \tag{8}$$

where x_t denotes the current input, h_t is the current hidden state, and h_{t-1} is the hidden state at the previous time step. z_t represents the update gate vector, and r_t represents the reset gate vector. W_z, W_r, and W denote the weight matrices corresponding to the update gate, reset gate, and the main network, respectively. $\sigma(\cdot)$ denotes the Sigmoid activation function, and $\tanh(\cdot)$ denotes the hyperbolic tangent function.

In the structural modeling branch, this study employs Categorical Boosting (CatBoost) as the static classification model to perform discriminative learning on the fused feature vectors. Regarding structured physical features, CatBoost demonstrates distinct advantages over traditional deep learning models. Recent studies indicate that gradient boosting frameworks based on symmetric decision trees typically outperform Multilayer Perceptrons (MLPs) or Transformer architectures when processing tabular data and physical features with significant statistical distribution disparities, while effectively mitigating overfitting under small sample conditions [26, 27]. Furthermore, the ordered boosting mechanism inherent to CatBoost addresses the prediction shift problem, which is critical for accurately capturing static physical laws.

CatBoost is an ensemble learning method built upon the Gradient Boosting framework, with strong capability in modeling complex nonlinear relationships. Compared with conventional gradient boosting decision trees, CatBoost introduces a symmetric tree construction strategy, which improves model convergence speed and prediction consistency. In addition, CatBoost utilizes a unique ordered target encoding technique to handle categorical variables, effectively avoiding target leakage without requiring one-hot encoding or manual embeddings. This design ensures stable training performance and strong generalization ability, particularly under small sample sizes or high-dimensional discrete input conditions.

Within the proposed framework, the primary role of the CatBoost model is to learn static nonlinear classification boundaries from the fused features, thereby compensating for potential blind spots of GRU in handling time-independent components. The model is trained by minimizing a multi-class cross-entropy loss function, which makes it naturally applicable to the multi-class state classification scenario of USV operations (e.g., safety levels 1 to 5). The specific loss function is defined in Equation (9).Considering that certain state levels (such as high-risk categories) tend to have fewer samples, CatBoost's built-in class balancing mechanisms further enhance its ability to recognize minority classes, thereby improving the system's sensitivity and responsiveness to safety-critical states near the decision boundary.

$$L = -\frac{1}{M} \sum_{i=1}^{M} \sum_{k=1}^{K} y_i^{(k)} \log p_i^{(k)} \tag{9}$$

where $y_i^{(k)}$ is an indicator variable that equals 1 if sample i belongs to class k, and 0 otherwise; $p_i^{(k)}$ denotes the predicted probability for class k of sample i; M is the total number of samples, K is the total number of classes, and L denotes the multi-class cross-entropy loss function.

After completing the parallel modeling stage, a Covariance Intersection (CI) fusion strategy is introduced to effectively integrate the complementary strengths of the CatBoost and GRU models in static and temporal modeling. The CI fusion strategy is introduced to address the challenge of 'unknown correlations' in multi-source heterogeneous data. Traditional Bayesian fusion or weighted averaging typically assumes independence among sources; however, USV multi-source sensors are often subject to common environmental noise (e.g., wind and waves), resulting in complex cross-correlations. The Covariance Intersection algorithm proposed by Julier et al. [28] has been proven to be the theoretically optimal solution for fusing information with unknown correlations, providing a consistent and conservative error upper bound for the fused estimate. This guarantees the robustness of the assessment system from a statistical perspective.The core idea of this strategy is to model the uncertainty distribution of each model's prediction and use it to guide the weight allocation in the fusion process, thereby achieving an optimal combination of prediction outputs.

Specifically, the CI fusion strategy takes the covariance matrices of predictions from each model as inputs and constructs a weighting function based on the principle of linear minimum variance estimation. CatBoost and GRU produce independent classification predictions along with their corresponding confidence distributions. The system then adaptively generates fusion weights based on the estimated uncertainty, and integrates the outputs to produce the final state level label. The mathematical formulation of this fusion algorithm is provided in Equations (10) and (11).

$$P_{CI}^{-1} = \omega P_a^{-1} + (1-\omega) P_b^{-1} \tag{10}$$

$$\hat{c}_{CI} = P_{CI} \left(\omega P_a^{-1} \hat{a} + (1-\omega) P_b^{-1} \hat{b} \right) \tag{11}$$

where P_a and P_b denote the covariance matrices of Model A and Model B, respectively; $\omega \in [0,1]$ is the weighting coefficient; P_{CI} is the uncertainty matrix (covariance) after fusion; $\hat{a}$ and $\hat{b}$ represent the output results of Model A and Model B, respectively; and $\hat{c}_{CI}$ denotes the final estimated output after fusion.

This strategy demonstrates strong robustness and adaptability in multi-class imbalanced classification problems. It is particularly suitable for real-world USV applications, where high-risk states are underrepresented, and decision boundaries are often ambiguous due to complex environmental dynamics.

3.2 Digital Twin System Architecture

While contemporary digital twin research for Unmanned Surface Vehicles (USVs) has made significant strides in high-fidelity scene rendering and kinematic data synchronization, there remains a practical need to enhance the interpretation of these visualized data. Existing systems typically prioritize the representation of physical motion (e.g., trajectory and attitude), which, while intuitive, does not directly quantify the underlying operational risks.

Accordingly, this study constructs a digital twin system integrated with real-time state evaluation capabilities. By embedding the proposed dual-channel hybrid model, the system is optimized to extend the functionality from basic motion monitoring to comprehensive safety analysis. This design enables the virtual model to not only replicate the physical behavior of the USV but also to output synchronized safety levels, providing operators with more informative decision support.

To support this integrated function, the system architecture is engineered to ensure efficient data flow between algorithmic inference and visual interaction. As illustrated in Figure 2, the architecture consists of four main layers: the perception layer, network layer, processing layer, and application layer. Based on multi-source sensor data, the system leverages Internet of Things (IoT) communication protocols to achieve state synchronization between the physical and virtual models, thereby supporting dynamic monitoring and visual feedback throughout the real-time operation of USVs.

1. **Perception Layer:** This layer integrates devices such as GPS, IMU, speed sensors, and cameras to continuously collect position, attitude, velocity, and environmental parameters of the USV. These inputs form multi-source temporal data streams that serve as the foundation for both state synchronization within the digital twin system and subsequent physical-statistical modeling and deep feature extraction.

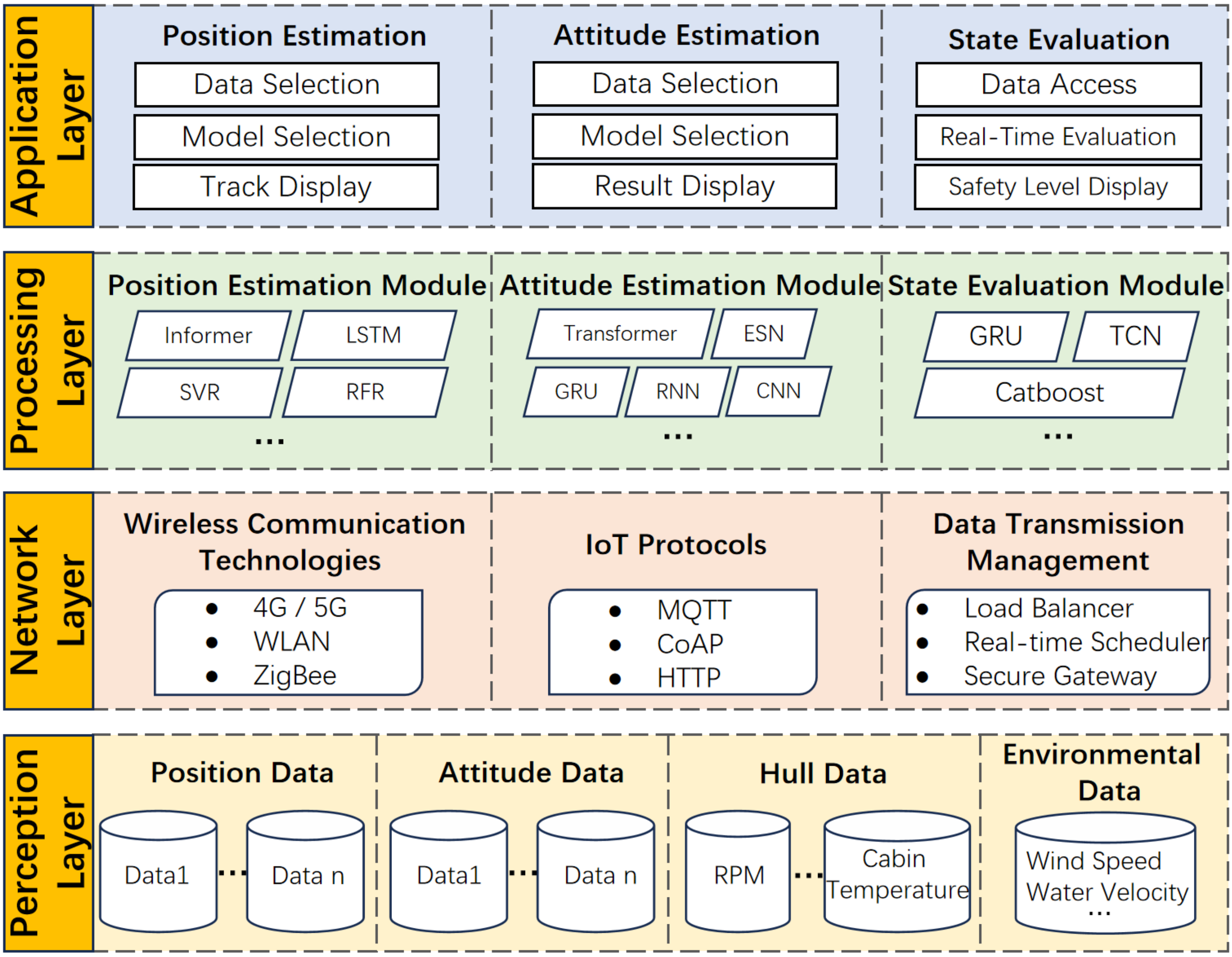

Figure 2. Digital twin system architecture for USVs.

2. **Network Layer:** Responsible for the transmission and scheduling of multimodal data, this layer employs wireless communication technologies such as 4G/5G, WLAN, and ZigBee, along with IoT communication protocols including MQTT, CoAP, and HTTP. These ensure stable, efficient, and secure data flow between the sensing endpoints and the processing unit, thereby supporting the real-time capabilities of the system.

3. **Processing Layer:** Serving as the core analytical component of the system, this layer operates on multi-source temporal data and integrates three key submodules: the Position Estimation module, the Attitude Estimation module, and the State Evaluation module. The former two leverage a combination of methods such as Informer and RNNs to achieve high-precision estimation of USV position and orientation. The State Evaluation module adopts a dual-channel feature extraction mechanism that constructs a hybrid representation combining structured and unstructured features. These are separately input into CatBoost and GRU models to integrate static classification with temporal modeling, ultimately enabling accurate and efficient evaluation of the USV's operational state.

4. **Application Layer:** Built on top of the digital twin platform, this layer provides a graphical interface for real-time visualization of the USV's motion status, estimated pose, and state evaluation outcomes. It allows users to perform remote interactive control based on visual feedback, and supports decision-making adjustments and model updates, thereby realizing a high-frequency closed loop of "Perception–Modeling–Evaluation–Feedback."

Table 1. Results of the ablation study.

Methods	ACC	PREC	REC	F1
TCNOnly-Fusion	0.7804	0.8151	0.7804	0.7809
PhysOnly-Fusion	0.7650	0.7774	0.7650	0.7651
CBOnly-Fusion	0.8010	0.8493	0.8010	0.8048
GRUOnly-Fusion	0.4734	0.5232	0.4734	0.4075
NoCI-Fusion	0.8027	0.8132	0.8027	0.8055
RawInput-Fusion	0.7804	0.8312	0.7804	0.7836
Ours	**0.8611**	**0.8700**	**0.8611**	**0.8629**

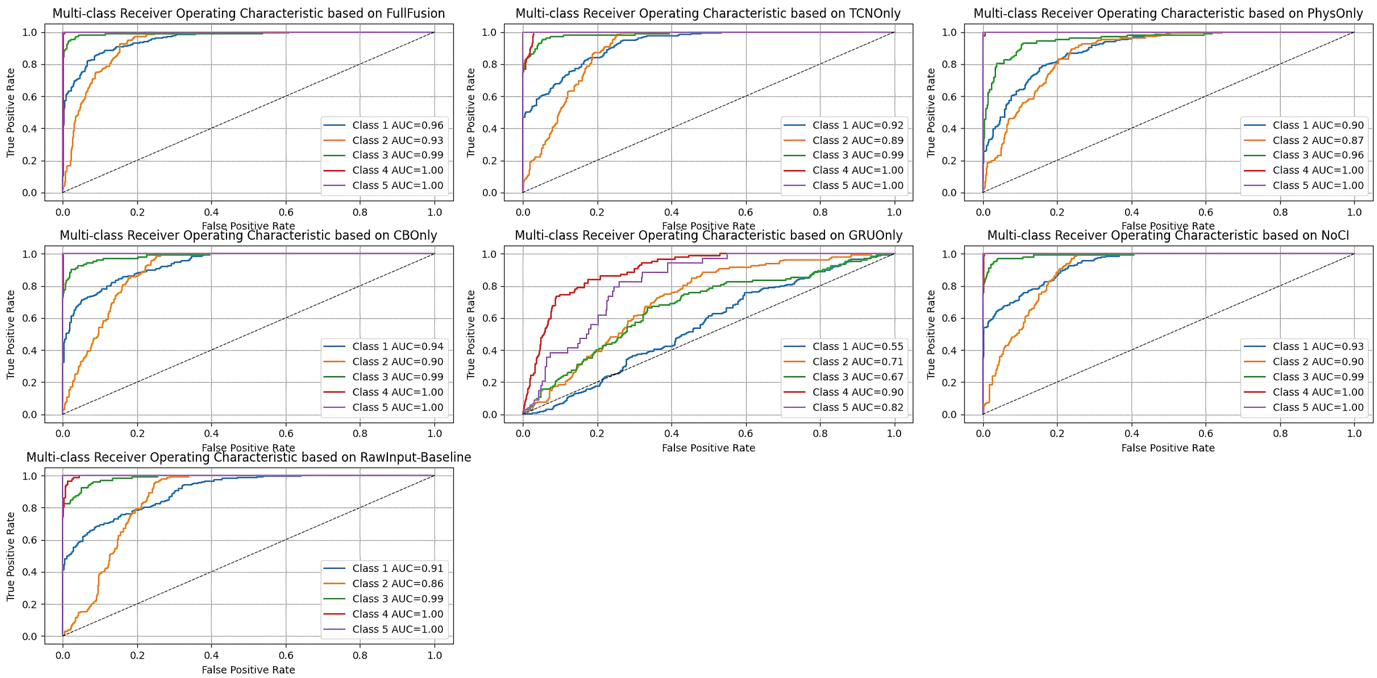

Figure 3. ROC curves of different ablation models.

4 Experiments

To validate the effectiveness of the proposed state evaluation method, a series of ablation experiments were conducted to compare different model architectures and modeling strategies. The evaluation focused on classification accuracy, ROC curves, and other key performance metrics to comprehensively assess the model's evaluation capability.

All experiments were carried out on a computer running a 64-bit Windows 10 operating system, equipped with an Intel i3-12100F processor and 16 GB of RAM. The algorithm was implemented and tested using the Python programming language within the PyCharm integrated development environment. The visual interaction platform of the digital twin system was developed based on the Unity engine, enabling real-time mapping and display of state evaluation results.

4.1 Dataset Description

In this experiment, a total of 2,923 original samples were selected, each containing 13 features: position X, position Y, position Z, pitch angle, roll angle, yaw angle, engine speed, vibration frequency, cabin temperature, wind speed, rainfall, ambient temperature, and water flow velocity. The operational state is classified into five levels, with Level 1 representing the highest safety and Level 5 the lowest.A sliding window approach was adopted to construct time-series samples, using a window size of 10 and a step size of 1, resulting in 2,914 time-series sequences. Among them, 80% (2,331 samples) were used as the training set, and 20% (583 samples) were used as the test set.

4.2 Ablation Study Design and Result Analysis

To systematically evaluate the performance contribution of each module within the state evaluation framework, seven comparative experiments

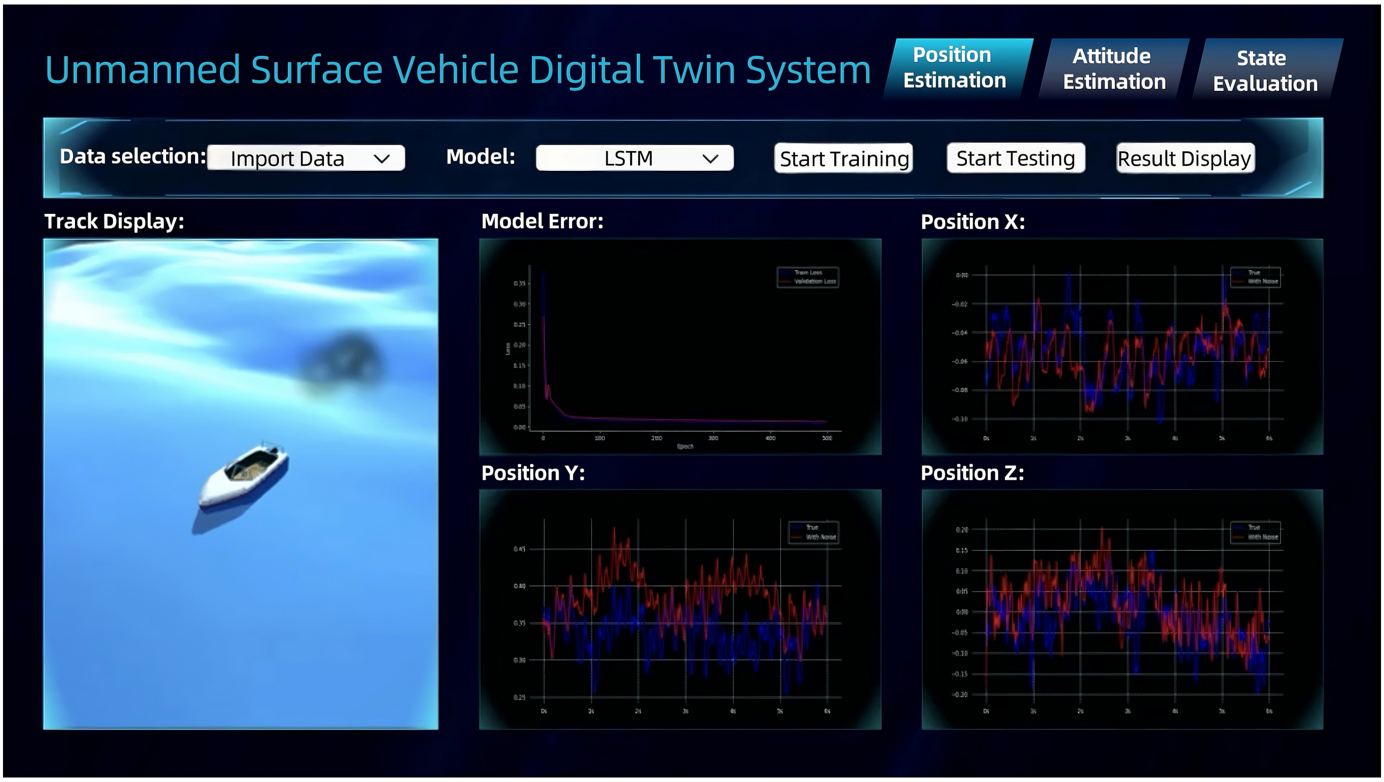

Figure 4. Position estimation function view.

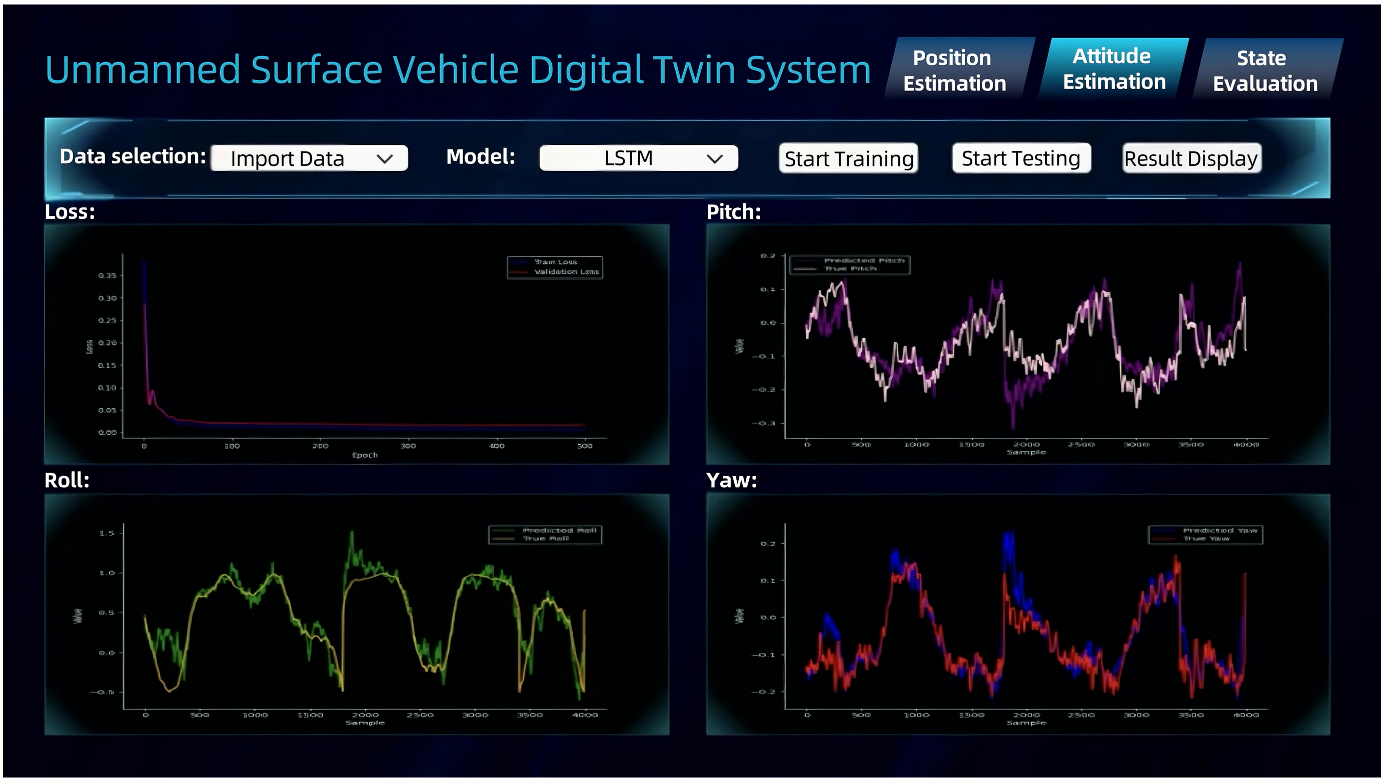

Figure 5. Attitude estimation function view.

were designed. These include: the complete model (denoted as Full-Fusion), models with individual feature extraction channels removed (PhysOnly-Fusion and TCNOnly-Fusion), models with separated classifiers (CBOnly-Fusion and GRUOnly-Fusion), a model without the fusion mechanism (NoCI-Fusion), and a baseline model without feature extraction (RawInput-Fusion).By selectively removing or replacing core components, this study analyzes the impact of structured modeling, temporal modeling, and model fusion on the overall recognition performance. This serves to validate the effectiveness and necessity of each module's design.

To visually illustrate the impact of each component on model performance, the results of all experimental groups are compared in terms of four key metrics: Accuracy, Precision, Recall, and F1-Score. A summary of these results is presented in Table 1.

As shown in Table 1, the complete model (Full-Fusion) achieved the best overall performance, with an accuracy of 0.8611 and an F1-Score of 0.8629, demonstrating the effectiveness of the dual-channel feature extraction, the parallel use of CatBoost and GRU, and the Covariance Intersection (CI) fusion strategy. When either feature extraction channel was removed, model performance dropped significantly: the F1-Score of TCNOnly-Fusion and PhysOnly-Fusion decreased to 0.7809 and 0.7651, respectively, indicating the strong complementarity between structural and dynamic features.In the classifier separation experiments, CBOnly-Fusion maintained a robust F1-score of 0.8048, whereas GRUOnly-Fusion exhibited a substantial degradation to 0.4075. This significant disparity is attributed to the compatibility between model architectures and feature characteristics. Although GRUs excel at capturing long-term temporal dependencies, they lack intrinsic sensitivity to the static physical statistical features (e.g., skewness and kurtosis) constructed in this study. Consequently, the standalone GRU architecture struggles to effectively leverage such non-temporal information for decision support. Furthermore, state transitions of USVs under wave disturbances are typically abrupt. The gating update mechanism of the GRU tends to induce a smoothing effect on such signals, resulting in a response lag to sudden state mutations and subsequent misclassification at dynamic decision boundaries. In contrast, the Full-Fusion model mitigates the limitations of single-temporal modeling by leveraging CatBoost to explicitly capture static feature distributions. Removing the CI fusion mechanism decreased the F1-score from 0.8629 to 0.8055. Unlike the equal-weighting approach in NoCI-Fusion, the CI strategy utilizes prediction covariance matrices to dynamically adjust weights based on uncertainty. This adjustment directly correlates with the observed reduction in false positive and false negative rates for critical states.

Overall, the results confirm that each module plays a critical role in improving the accuracy of state evaluation. The classification ROC curves of different experimental configurations are illustrated in Figure 3.

As illustrated in Figure 3, the Full-Fusion model achieved high AUC values across all classes, with perfect scores of 1.00 for both Class 4 and Class 5. Beyond the numerical AUC metrics, the morphology of the ROC curves further unveils the confidence characteristics of the classifiers. The Full-Fusion curves exhibit a sharp ascent towards the top-left corner, indicating that the model maintains high recall even at exceptionally low false-positive rates, thereby demonstrating superior classification confidence. In contrast, the GRUOnly curves for Class 1 and Class 3 hover near the diagonal line, reflecting substantial predictive uncertainty for these categories and a suboptimal trade-off between sensitivity and specificity. This visual comparison provides intuitive verification of the effectiveness of the CI fusion strategy. Moreover, removing either of the feature extraction channels (as in TCNOnly or PhysOnly) or eliminating the fusion mechanism (NoCI) led to varying degrees of performance degradation. These results further confirm the importance of the coordinated contribution of all modules to the success of the state classification task.

4.3 Digital Twin System Demonstration

The core functional modules of the digital twin platform developed in this study for surface unmanned systems consist of three main components: the Position Estimation Module, the Attitude Estimation Module, and the State Evaluation Module. Together, these modules form a unified digital twin system architecture that integrates data perception, state modeling, and intelligent evaluation.

Figures 4 and 5 illustrate the Position Estimation Module and Attitude Estimation Module within the digital twin platform. By integrating deep learning models with a visual interface, the system enables real-time prediction and dynamic visualization of key state parameters of the Unmanned Surface Vehicles (USVs).

The Position Estimation Module facilitates credibility assessment and path deviation analysis for USV autonomous navigation. Beyond calculating 3D spatial positions in real time, the system visually contrasts the dynamic differences between the 'predicted trajectory' and 'observed trajectory' to directly reflect

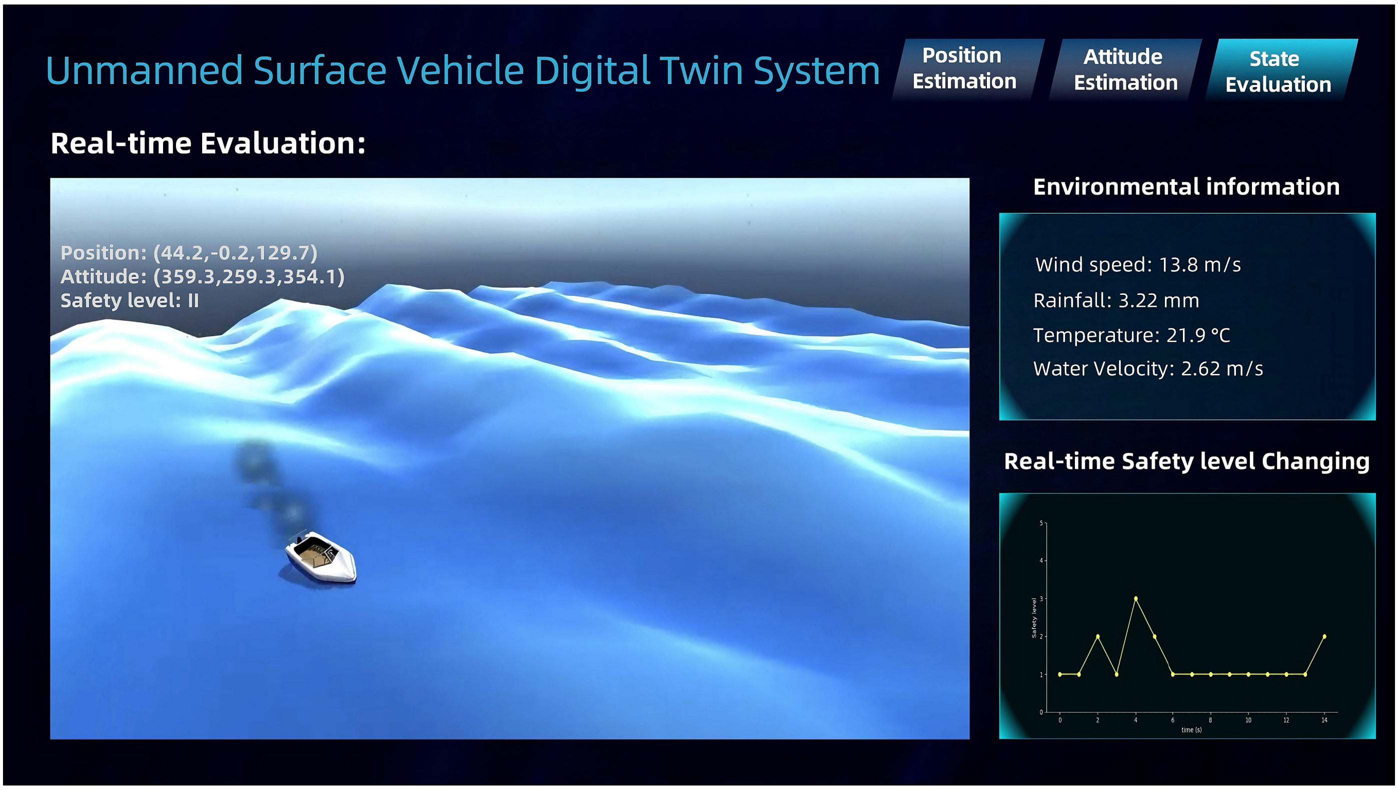

Figure 6. State evaluation function view.

the drift magnitude of the positioning system. This visual feedback mechanism, based on trajectory reconstruction, enables operators to rapidly identify cumulative trends in positioning errors. In conjunction with real-time model error convergence curves, maintenance personnel can diagnose the adaptive state of the deep learning models online (e.g., detecting overfitting risks). Consequently, during long-endurance missions, key decisions regarding manual takeover or course correction can be made based on positioning reliability.

The Attitude Estimation Module focuses on hull stability analysis under complex sea conditions, highlighting the time-varying characteristics of key attitude angles such as roll and pitch. This visualization is critical for navigational safety. By monitoring the oscillation amplitude and recovery rate of attitude angles under wave disturbances, operators can assess the current stability of the USV. Specifically, when encountering sudden crosswinds or large waves, the module assists in detecting precursor signals of instability (such as periodic high-amplitude rolling), enabling timely tactical decisions such as deceleration or course adjustment to mitigate capsizing risks.Through the coordinated operation of the position and attitude modules, the system achieves comprehensive spatial and orientation modeling of the surface unmanned platform, laying a solid foundation for state synchronization and intelligent control within the digital twin framework.

Figure 6 presents the interface of the State Evaluation Module within the digital twin platform. The Module implements real-time perception and intelligent evaluation of the USV's operational state. The system maps multi-dimensional evaluation results into a continuous safety level evolution trajectory, displayed synchronously with environmental parameters like wind and current speeds. Its primary interaction function involves revealing dynamic risk evolution patterns. Rather than relying on instantaneous alarm labels, operators can identify potential performance degradation trends (e.g., a gradual shift from Level I to Level III) by analyzing the slope and fluctuation frequency of safety levels over time. This situational visualization, correlated with environmental inducers, allows personnel to rapidly pinpoint potential risk variations and their occurrence intervals, providing a basis for subsequent intervention decisions.

Overall, this module completes a closed-loop workflow—from sensor data acquisition and feature- level modeling to visualized evaluation output—significantly enhancing state transparency and risk warning capabilities during USV operations. In conjunction with the Position Estimation and Attitude Estimation modules, it forms a comprehensive digital twin architecture that integrates multi-dimensional state perception and dynamic evaluation, providing critical support for intelligent operation and safety management of surface unmanned platforms.

5 Conclusion

This study addresses the problem of operational state perception and intelligent evaluation for Unmanned Surface Vehicles (USVs) operating in complex and dynamic environments. A state evaluation method based on dual-channel feature extraction and temporal model fusion is proposed, and a digital twin system supporting visualization and interactive feedback is developed.At the methodological level, the dual-channel feature modeling mechanism effectively integrates structured statistical features with unstructured temporal features. The parallel use of CatBoost and GRU, combined with a classification-enhancing fusion strategy, enables highly accurate and robust identification of system states. At the system level, the developed digital twin platform achieves a closed-loop design that spans data acquisition, state modeling, and visual presentation, offering reliable support for USV status monitoring, early warning, and operational scheduling.Experimental results demonstrate that the proposed method outperforms baseline approaches across multiple evaluation metrics, showing strong practicality and application potential. Future research may further explore fusion strategies for higher-dimensional sensing data, collaborative multi-task evaluation under complex navigation scenarios, and the extension of digital twin platforms to fleet-level unmanned systems.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported by the Natural Science Foundation of Xinjiang Uygur Autonomous Region under Grant 2024D01A04, and the National Natural Science Foundation of China under Grant 62203020.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Zhao, F., & Wang, S. (2025). A hybrid framework integrating end-to-end deep learning with Bayesian inference for maritime navigation risk prediction. *Journal of Marine Science and Engineering, 13*(10), 1925. [CrossRef]

[2] Xu, P. F., Han, C. B., Cheng, H. X., Cheng, C., & Ge, T. (2022). A physics-informed neural network for the prediction of unmanned surface vehicle dynamics. *Journal of Marine Science and Engineering, 10*(2), 148. [CrossRef]

[3] Zhang, B. C., Wang, J. D., Gao, S., Yin, X. J., & Gao, Z. (2023). Health Status Evaluation of Welding Robots Based on the Evidential Reasoning Rule. *Electronics, 12*(8), 1755. [CrossRef]

[4] Chen, J., Zhao, Y., Wu, C., & Xu, Q. (2020). Data-driven health assessment in flight control system. *Applied Sciences, 10*(23), 8370. [CrossRef]

[5] Qiao, Z., Pan, X., He, Y., Jiangnan, Z., Yu, H., & Geng, C. (2024). A Study Examining the Adverse Effects of Electromagnetic Pulse on System-Level Unmanned Aerial Vehicles and Their Subsequent Damage Assessment and Mitigation Strategies. *Iran. J. Chem. Chem. Eng.*(*IJCCE*), *43*(11), 4185-4199.

[6] Wei, L., Sun, Y., Diao, Q., Xu, H., Tan, X., & Fan, Y. (2024). State of health estimation of lithium-ion batteries based on stacked-LSTM transfer learning with Bayesian optimization and multiple features. *IEEE Sensors Journal*. [CrossRef]

[7] Xiao, X., & Guo, J. (2023). A novel switchgear state assessment framework based on improved fuzzy C-means clustering method with deep belief network. *Frontiers in Energy Research, 11*, 1335184. [CrossRef]

[8] Ren, C., & Xu, Y. (2019). Transfer learning-based power system online dynamic security assessment: Using one model to assess many unlearned faults. *IEEE Transactions on Power Systems, 35*(1), 821-824. [CrossRef]

[9] Zhang, B., Chen, D., Su, W., Liu, T., & Shao, Y. (2024). Aviation fuel pump health state assessment based on evidential reasoning and random forests. *Electronics Letters, 60*(9), e13195. [CrossRef]

[10] Qu, S., Men, X., Liu, M., Cui, J., Wu, H., & Fu, Y. (2025). Navigation Attitude Prediction for Unmanned Surface Vessels in Wave Environments Using Improved Unscented Kalman Filter and Digital Twin Model. *Journal of Marine Science and Engineering, 13*(5), 932. [CrossRef]

[11] Zhang, B., Wang, S., & Ji, S. (2024). A deep learning combined prediction model for prediction of ship

motion attitude in real conditions. *Ships and Offshore Structures*, *19*(11), 1868-1883. [CrossRef]

[12] Wang, Y., Lu, X. R., & Chen, Y. (2024). Premonition-driven deep learning model for short-term ship violent roll motion prediction based on the hull attitude premonition mechanism. *Applied Ocean Research*, *146*, 103970. [CrossRef]

[13] Cao, L., Qin, Y., Pan, Y., & Liang, H. (2024). Prescribed performance-based optimal formation control for USVs with position constraints and yaw angle time-varying partial constraints. *IEEE Transactions on Intelligent Transportation Systems*. [CrossRef]

[14] van der Saag, J., Trevisan, E., Falkena, W., & Alonso-Mora, J. (2025). Active Disturbance Rejection Control for Trajectory Tracking of a Seagoing USV: Design, Simulation, and Field Experiments. *arXiv preprint arXiv:2506.21265*.

[15] Huang, M., Li, X., Li, Z., Zhang, D., & Chen, Y. (2025). Uncertainty-aware deep distributed reinforcement learning for autonomous navigation of unmanned surface vehicles in complex environments. *Ocean Engineering*, *342*, 122899. [CrossRef]

[16] Elsanhoury, M., Koljonen, J., Prol, F. S., Elmusrati, M., & Kuusniemi, H. (2024, December). Resilient Navigation in GNSS-Denied Conditions Using Novel LEO-Based Fusion Positioning. In *2024 IEEE International Conference on Wireless for Space and Extreme Environments (WiSEE)* (pp. 118-123). IEEE. [CrossRef]

[17] Saptoe, J., van Aardt, S., Smith, F., & Hatefi, S. (2025). Unmanned surface vehicle with deep learning-based obstacle avoidance for water quality monitoring. In *MATEC Web of Conferences* (Vol. 417, p. 10002). EDP Sciences. [CrossRef]

[18] Cen, J., Li, J., Liu, X., Chen, J., Li, H., Huang, W., ... & Ke, S. (2024). A hybrid prediction model of vessel trajectory based on attention mechanism and CNN-GRU. *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment*, *238*(4), 809-823. [CrossRef]

[19] Liu, W., Xu, B., & Li, J. (2025). Data-Driven Carbon Emission Dynamics Under Ship In-Port Congestion. *Journal of Marine Science and Engineering*, *13*(4), 812. [CrossRef]

[20] Hasan, A., Widyotriatmo, A., Fagerhaug, E., & Osen, O. (2023). Predictive digital twins for autonomous surface vessels. *Ocean engineering*, *288*, 116046. [CrossRef]

[21] Raza, M., Prokopova, H., Huseynzade, S., Azimi, S., & Lafond, S. (2022). Towards integrated digital-twins: An application framework for autonomous maritime surface vessel development. *Journal of Marine Science and Engineering*, *10*(10), 1469. [CrossRef]

[22] Peng, Z., Yue, Y., Zhu, X., Huang, M., Wong, P., Yao, S., ... & Hou, D. (2023, December). Digital twin applications in unmanned surface vehicles: A survey. In *2023 6th International Conference on Software Engineering and Computer Science (CSECS)* (pp. 1-8). IEEE. [CrossRef]

[23] Madusanka, N. S., Fan, Y., Ahmed, F., Yang, S., & Xiang, X. (2023, October). Development of an Autonomous Pilotage for a Digital Twin-based Unmanned Surface Vessel in Virtual Reality. In *National Technical Seminar on Unmanned System Technology* (pp. 145-166). Singapore: Springer Nature Singapore. [CrossRef]

[24] Vasconcellos, E. C., Sampaio, R. M., Araújo, A. P., Clua, E. W. G., Preux, P., Guerra, R., ... & Sanchez-Pi, N. (2024). Reinforcement-learning robotic sailboats: simulator and preliminary results. *arXiv preprint arXiv:2402.03337*.

[25] Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*.

[26] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018, December). CatBoost: unbiased boosting with categorical features. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems* (pp. 6639-6649).

[27] Gorishniy, Y., Rubachev, I., Khrulkov, V., & Babenko, A. (2021). Revisiting deep learning models for tabular data. *Advances in neural information processing systems*, *34*, 18932-18943.

[28] Julier, S. J., & Uhlmann, J. K. (1997, June). A non-divergent estimation algorithm in the presence of unknown correlations. In *Proceedings of the 1997 American Control Conference (Cat. No. 97CH36041)* (Vol. 4, pp. 2369-2373). IEEE. [CrossRef]

Yuting Bai received the Ph.D. degree in control science and engineering from Beijing Institute of Technology in 2019, the M.S. degree in management science and engineering from Beijing Technology and Business University in 2015, and the B.S. degree in automation from Beijing Technology and Business University in 2012. He is now an associate professor in Beijing Technology and Business University. His research mainly covers the state estimation, information fusion and machine learning. (Email: baiyuting@btbu.edu.cn)

Jiyuan Hu received his Bachelor's degree in Computer Science and Technology from Zhengzhou University of Light Industry in 2023. He is dedicated to the research of surface unmanned systems. (Email: 2330602090@st.btbu.edu.cn)

Eziz Tursun received the B.S. degree in Physics Education from Xinjiang Normal University, China, 1998; the Ph.D. degree in Computer Applied Technology from University of Chinese Academy of Sciences, China, 2017. (Email: eziztursun@qq.com)

Hurxida Yimit In 2011, obtained a Bachelor's degree in Computer Science and Technology from Xinjiang Normal University in China;In 2014, obtained a Master's degree in Software Engineering from the University of Electronic Science and Technology of China. (Email: 448463774@qq.com)