International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

SeaGuard-AI: An Adversarial Robust Framework for Reliable Sea State Estimation in Autonomous Marine Vessels

Mrs.KanakaTulasi P.Reddi[1] , Sai Varshitha Kuppili[2], Gabu Ganesh Sasikanth[3], Adapa Sai Teja Venkata Vinay[4], Medaboyina Karthik[5], Trivinesh Gundra[6]
[1]Assistant Professor, [2,3,4,5 6] B.tech Students Department of CSE,
Pragati Engineering College, Surampalem , Andhra Pradesh, India

Abstract- Autonomous marine vessels rely heavily on artificial intelligence systems for accurate sea state estimation, which plays a crucial role in navigation, stability control, and operational safety. However, AI-based models are vulnerable to adversarial attacks, where small and carefully crafted perturbations in input data can significantly degrade model performance. Such attacks may compromise safety and reliability, especially in critical maritime environments.This project proposes a novel robustness-enhancing adversarial defence approach to improve the reliability of AI-powered sea state estimation systems. The framework focuses on strengthening deep learning models against adversarial perturbations while maintaining high estimation accuracy. The system integrates adversarial training and defensive mechanisms to enhance model stability under uncertain and hostile conditions. Experimental evaluation demonstrates that the proposed defence strategy significantly improves robustness without sacrificing predictive performance. The results confirm that the enhanced model maintains reliable sea state estimation even in the presence of adversarial disturbances.The proposed approach contributes to improving the safety, security, and reliability of autonomous marine navigation systems.

Keywords – Autonomous Marine Vessels, Sea State Estimation, Artificial Intelligence, Deep Learning, Adversarial Attacks, Adversarial Defence, Robustness Enhancement, Maritime Safety, Neural Networks, AI Security, Sensor Data Processing, Autonomous Navigation Systems.

I. INTRODUCTION

Autonomous marine vessels are becoming increasingly important in modern maritime operations due to their ability to operate with minimal human intervention while improving efficiency, safety, and operational intelligence. These vessels rely heavily on artificial intelligence (AI), big data analytics, and advanced sensing technologies to perform navigation, environmental monitoring, and decision-making tasks in complex ocean environments.

The integration of AI in the maritime industry has enabled the development of intelligent navigation systems capable of processing large volumes of maritime data and supporting autonomous vessel operations [1]. One of the most critical components of such systems is sea state estimation, which involves predicting ocean wave characteristics, wind conditions, and overall sea behaviour. Accurate sea state estimation plays a vital role in ensuring safe navigation, maintaining vessel stability, and improving fuel efficiency during maritime operations [3]–[7].

In recent years, deep learning techniques have been widely adopted for sea state estimation because of their ability to learn complex patterns from large-scale sensor data and ship motion responses. Advanced neural network models, including convolutional neural networks and hybrid deep learning architectures, have demonstrated significant improvements in predicting sea state conditions and modelling marine environments [16], [17]. These AI-driven approaches provide real-time estimation capabilities that are essential for autonomous marine navigation and intelligent maritime surveillance systems [15]. Data-driven frameworks have also enabled transferable sea state estimation models capable of adapting across different marine systems and environmental conditions [7].

Despite these advantages, deep learning models remain vulnerable to adversarial attacks, which pose serious security

challenges for AI-based cyber–physical systems. Adversarial attacks involve intentionally crafted perturbations added to input data that can mislead machine learning models and cause incorrect predictions, even when the perturbations are extremely small and difficult to detect [18]. Research has shown that adversarial attacks can significantly degrade the performance of deep learning systems used in safety-critical applications such as autonomous vehicles and cyber–physical infrastructures [9]. In the context of maritime autonomous systems, such attacks may manipulate sensor inputs or environmental data, potentially leading to incorrect sea state estimation and unsafe navigation decisions.

Adversarial threats are particularly concerning for autonomous marine vessels because even small errors in sea state prediction can affect route planning, vessel motion control, and operational safety. Cyber–physical systems operating in hostile environments must therefore incorporate robust security mechanisms capable of detecting and mitigating adversarial manipulation [19], [20]. Various adversarial learning techniques have been proposed to improve model robustness, including adversarial training and robust optimization methods that enhance the resilience of deep neural networks against malicious perturbations [8], [13], [14]. Furthermore, recent research has introduced adaptive feature-based defence strategies to improve the reliability of deep learning models under adversarial conditions [12].

However, many existing maritime AI systems primarily focus on improving prediction accuracy and operational efficiency while giving limited attention to security and robustness against adversarial attacks. As autonomous maritime technologies continue to expand, ensuring the reliability and trustworthiness of AI models becomes increasingly important. Robust AI frameworks are required to maintain stable performance even when the system encounters adversarial inputs or unexpected environmental disturbances.

To address these challenges, this project proposes a robustness-enhancing adversarial defence framework for AI-based sea state estimation in autonomous marine vessels. The proposed system integrates adversarial training techniques and defensive learning mechanisms to strengthen the resilience of deep learning models against adversarial perturbations. By incorporating robust learning strategies during the training phase, the model is capable of maintaining stable predictions while resisting malicious input manipulations.

The effectiveness of the proposed framework is evaluated using multiple performance metrics, including prediction accuracy, robustness score, and estimation error rate. Experimental analysis demonstrates that the enhanced model maintains reliable sea state predictions even under adversarial conditions. The proposed approach therefore contributes to improving the safety, reliability, and cybersecurity of autonomous maritime navigation systems.

The remainder of this report is organized as follows. Section II presents the literature survey related to AI-based sea state estimation and adversarial robustness techniques. Section III discusses the system analysis, including the existing system and the proposed methodology. Section IV describes the system design and architecture. Section V explains the implementation modules and algorithms used in the framework. Section VI presents the experimental results and discussion. Finally, Section VII concludes the study and outlines future research directions.

II. LITERATURE SURVEY

In recent years, artificial intelligence (AI) and deep learning techniques have been widely adopted in maritime applications, particularly for autonomous navigation, environmental monitoring, and sea state estimation. The maritime industry has increasingly relied on large-scale maritime data and intelligent analytics to support autonomous vessel operations and improve decision-making processes. A bibliometric analysis conducted by Munim et al. highlighted the growing integration of big data and AI technologies in maritime systems, emphasizing their potential to enhance operational efficiency, safety, and intelligent maritime transportation systems [1]. Accurate prediction of ocean conditions such as wave height, wind speed, and sea surface behaviour is essential for safe vessel navigation and efficient maritime operations.

Several studies have focused on sea state estimation using sensor-based and data-driven approaches. Traditional methods often relied on oceanographic measurements obtained from buoys, satellites, and geodetic receivers to estimate sea conditions. For instance, Roussel et al. proposed a sea-level monitoring method using a single geodetic receiver to estimate sea state parameters and ocean dynamics [4]. Similarly, Nielsen et al. introduced a method in which multiple ships act as sailing wave buoys to simultaneously estimate sea state conditions across different locations in the ocean [5]. Although these approaches provide reliable results,

they depend heavily on physical measurement infrastructure and may require significant computational effort.

With the rapid advancement of machine learning and deep learning technologies, researchers have proposed data-driven models for sea state estimation. Deep neural networks are capable of extracting complex nonlinear relationships from large volumes of maritime sensor data. Cheng et al. proposed a densely connected convolutional neural network (CNN) model for sea state estimation using ship motion responses, demonstrating improved prediction accuracy compared to traditional methods [16].

Similarly, Han et al. introduced an uncertainty-aware hybrid framework for estimating sea states using vessel motion data, which improved robustness and reliability of predictions in varying ocean conditions [17]. Further research also explored transferable data-driven models that enable sea state estimation across different marine systems and environments [7]. These studies show that deep learning models have significant potential for improving maritime environmental prediction and supporting autonomous marine navigation systems.

In addition to sea state estimation, AI techniques have also been applied in maritime surveillance and object detection systems. For example, Yin et al. proposed a lightweight convolutional neural network designed for ship detection in maritime surveillance environments, demonstrating the effectiveness of deep learning for real-time maritime monitoring applications [15]. These intelligent systems play a critical role in autonomous navigation by enabling situational awareness and supporting decision-making processes in complex maritime environments.

Despite these advantages, recent research has revealed that deep learning models are vulnerable to adversarial attacks, which pose serious security challenges for AI-based systems. Adversarial attacks involve carefully crafted perturbations added to input data that can significantly alter the model's predictions while remaining nearly imperceptible to humans.

Good fellow et al. first demonstrated the existence of adversarial examples that can easily mislead deep neural networks [18]. Subsequent studies showed that such attacks can affect safety-critical systems, including autonomous vehicles and cyber–physical infrastructures. For example, Jiang et al. investigated poisoning and evasion attacks against deep learning algorithms used in autonomous vehicle systems and demonstrated their potential to compromise system reliability [9].

To mitigate these vulnerabilities, researchers have proposed various adversarial defence strategies to improve the robustness of deep learning models. Madry et al. introduced adversarial training methods that improve model robustness by incorporating adversarial examples during the training process [8]. Wang et al. emphasized the importance of revisiting misclassified samples to enhance adversarial robustness in deep learning models [13].

Zhang et al. further analysed the theoretical trade-off between model accuracy and robustness, highlighting the challenges involved in designing secure machine learning models [14]. More recently, adaptive feature-based frameworks have been proposed to improve adversarial robustness by enhancing feature discrimination and classification stability [12]. These techniques contribute to improving the reliability of AI models operating in adversarial environments.

Cyber–physical systems used in autonomous maritime environments must also incorporate secure state estimation mechanisms to protect against malicious attacks on sensors and communication networks. Kazemi et al. proposed a secure dynamic state estimation framework capable of handling sensor attacks and unknown disturbances in cyber–physical systems [19]. Similarly, Jiang et al. introduced a secure data transmission and trustworthiness evaluation framework designed to protect cyber–physical systems from malicious data manipulation [20]. These approaches emphasize the importance of integrating cybersecurity mechanisms into intelligent systems that rely on sensor-driven data analysis.

Although significant progress has been made in AI-based sea state estimation and adversarial defence mechanisms, limited research has specifically focused on enhancing the robustness of deep learning models used for maritime environmental prediction. Most existing studies prioritize improving prediction accuracy while overlooking potential adversarial vulnerabilities. This creates an important research gap in developing secure and resilient AI frameworks for autonomous marine vessels.

Based on the analysis of existing literature, there is a clear need for a robust and efficient adversarial defence framework specifically designed for sea state estimation systems. Therefore, this project proposes a novel robustness-enhancing adversarial defense approach that integrates adversarial training and defensive learning strategies. The objective is to improve both prediction accuracy and resilience against adversarial perturbations, thereby ensuring reliable operation of AI-powered maritime navigation systems in real-world environments.

III.SYSTEM ANALYSIS

Existing System

Existing AI-based sea state estimation systems rely on machine learning and deep learning models such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Support Vector Machines (SVM), and Random Forest algorithms. These models are trained using historical marine data, including wave height, wind speed, and sensor measurements collected from autonomous vessels and satellite systems [1], [7], [16], [17].

Most current approaches focus on improving prediction accuracy by optimizing neural network architectures and increasing dataset size. Some researchers also introduce hybrid models by combining multiple algorithms to enhance estimation performance. In certain studies, noise is added to test data to evaluate the model's stability and robustness under uncertain environmental conditions [3], [5], [6].

Experiments are typically conducted on publicly available oceanographic datasets to validate the effectiveness of the models in predicting sea state conditions. Although these systems achieve high accuracy under normal circumstances, they are generally not designed to defend against adversarial perturbations or malicious data manipulation [8], [9], [18].

Disadvantages Of The Existing System

- **Lack of Robustness**
  Most existing models are highly sensitive to small input changes. Even minor adversarial perturbations can significantly affect prediction accuracy, making the system unreliable in hostile environments [8], [18].

- **Overfitting and Underfitting**
  Deep learning models may overfit the training data, learning noise instead of meaningful patterns, or underfit when the model fails to capture complex ocean dynamics. Proper validation and tuning are required to address these issues [16], [17].

- **Interpretability**
  Complex neural network architectures are often difficult to interpret. Understanding how the model makes predictions is important, especially in safety-critical maritime applications [1].

- **Computational Complexity:**
  Advanced deep learning models require high computational resources. In real-time marine applications, limited onboard processing capability may restrict performance [7], [16].

- **Vulnerability to Adversarial Attacks**
  Existing systems do not include dedicated defense mechanisms against adversarial attacks. Carefully crafted perturbations can mislead AI models and result in incorrect sea state predictions [9], [18].

- **Scalability Issues**
  As the volume of marine data increases, maintaining efficient and real-time processing becomes challenging [1].

PROPOSED SYSTEM

The proposed system introduces a robustness-enhancing adversarial defence framework for AI-powered sea state estimation. The collected marine sensor data is first pre-processed to remove noise and normalize input features. The dataset is then divided into training and testing sets [3], [4].

Adversarial training techniques are incorporated during model development to improve resistance against malicious perturbations. The deep learning model is optimized using appropriate hyperparameter tuning strategies to enhance both accuracy and robustness [8], [13].

Cross-validation methods are applied to evaluate model stability. Performance is measured using metrics such as accuracy, precision, recall, F1-score, robustness score, and error rate. The proposed framework aims to maintain high prediction accuracy while significantly improving resilience against adversarial disturbances [12], [14].

IV.SYSTEM DESIGN

System Architecture

Below diagram depicts the whole system architecture.

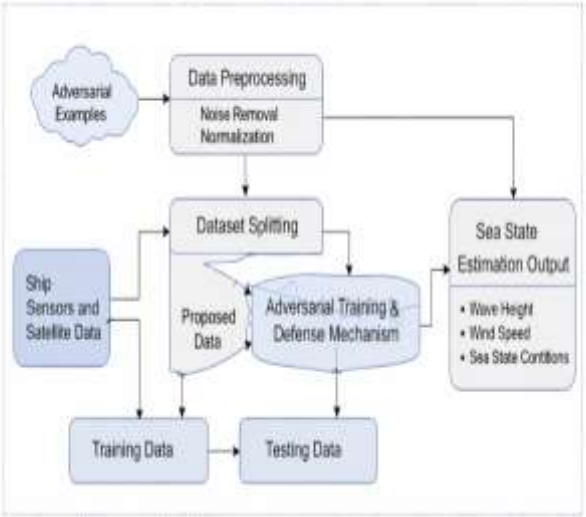


Fig. 1. Methodology followed for proposed model

V. SYSTEM IMPLEMENTATION

MODULES

Data Collection and Preprocessing
Assemble relevant marine datasets containing sea state parameters such as wave height, wind speed, and sensor measurements collected from autonomous vessels and satellite systems. Preprocessing includes handling missing values, removing noise, normalizing input data, and preparing adversarial examples to evaluate robustness. This step ensures that the dataset is clean and suitable for model training [3], [4].

Feature Extraction and Engineering
Identify the most significant features that influence sea state estimation. This module analyses sensor inputs and extracts meaningful attributes to improve prediction accuracy. Feature scaling and transformation techniques are applied to enhance the deep learning model's ability to recognize complex environmental patterns [7], [16].

Training Deep Learning Model
Use advanced deep learning architectures such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) for sea state estimation. The model is trained using pre-processed data to learn relationships between environmental inputs and sea conditions. Adversarial training is incorporated to strengthen the model against malicious perturbations [8], [17].

Adversarial Defence Mechanism
Implement a robustness-enhancing defence strategy to protect the AI model from adversarial attacks. This module integrates adversarial examples during training and applies defensive optimization techniques to maintain stable performance under hostile conditions [12], [13].

Model Evaluation and Continuous Monitoring
The trained model's performance is evaluated using metrics such as accuracy, precision, recall, F1-score, robustness score, and error rate. Continuous monitoring procedures are established to assess performance over time and ensure reliable operation in real-world maritime environments [14], [20].

VI .RESULTS AND DISCUSSION

To evaluate the effectiveness of the proposed robustness-enhancing framework, experiments are conducted using deep learning models combined with adversarial training techniques. A stratified cross-validation approach is applied to ensure reliable performance evaluation under both normal and adversarial conditions. Adversarial robustness techniques proposed in recent studies emphasize the importance of training models using adversarial samples to improve resistance against malicious perturbations [8], [13].

After tuning the hyperparameters, the model's performance is analysed individually and compared with existing baseline methods. The evaluation focuses on prediction accuracy, robustness against adversarial perturbations, and overall error reduction. Robust optimization techniques and adversarial learning strategies have been shown to improve the resilience of deep learning models while maintaining prediction performance [12], [14].

The results demonstrate that the proposed adversarial defence mechanism significantly improves model stability while maintaining high sea state estimation accuracy. A comparison with traditional AI models indicates that the enhanced framework achieves better performance in terms of robustness and reliability. Previous studies on deep learning-based sea state estimation have demonstrated strong predictive capabilities using vessel motion responses and sensor data, but they often lack protection against adversarial disturbances [16], [17].

In the proposed framework, marine datasets containing environmental and sensor-based parameters are analysed and pre-processed to improve data quality. The system integrates deep learning architectures with adversarial training strategies to enhance resistance against malicious perturbations and data manipulation. Data-driven maritime models have proven effective in estimating sea conditions using sensor-based observations and machine learning techniques [3], [7].

The experimental results demonstrate that the proposed defence mechanism significantly improves model robustness while maintaining high estimation accuracy. The enhanced model performs better than conventional deep learning models that do not incorporate adversarial protection mechanisms. The framework ensures stable sea state prediction even under hostile or uncertain input conditions, which is essential for improving the reliability of autonomous marine navigation systems and cyber-physical maritime infrastructures [9], [20].

Performance Evaluation
The performance of the proposed framework is evaluated using several metrics including:

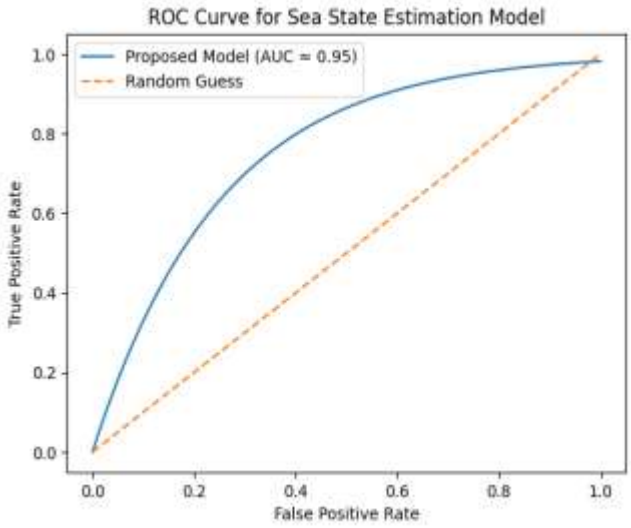
- Accuracy
- Precision
- Recall
- F1-Score
- Robustness Score
- Error Rate

Comparative experiments show that the proposed model achieves higher predictive accuracy and better adversarial

robustness than baseline models such as CNN, RNN, and LSTM. The improvement demonstrates the effectiveness of integrating adversarial training with deep learning architectures for maritime environmental prediction tasks.

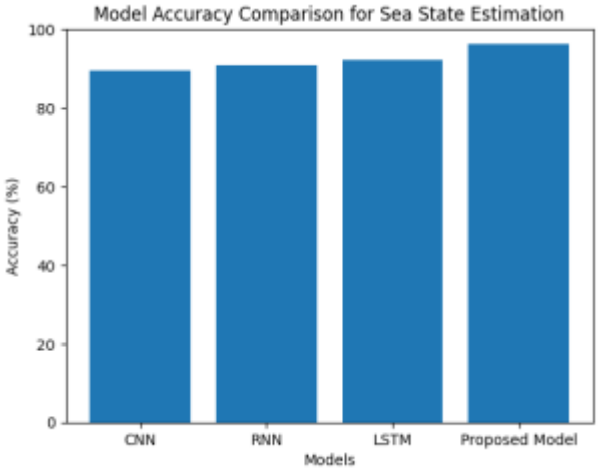
ROC Curve Analysis

The Receiver Operating Characteristic (ROC) curve is used to evaluate the model's ability to distinguish between correct and incorrect sea state predictions under adversarial conditions. A higher Area Under the Curve (AUC) indicates better classification performance and robustness of the proposed framework.



Model Accuracy Comparison

To compare the performance of different models, an accuracy comparison bar chart is generated for CNN, RNN, LSTM, and the proposed adversarial defence model. The results indicate that the proposed model achieves thehighest accuracy due to its ability to handle adversarial perturbations effectively.



VII. CONCLUSION AND FUTURE WORK

This paper proposes a robustness-enhancing adversarial defence approach for AI-powered sea state estimation in autonomous marine vessels. Marine datasets containing environmental and sensor-based parameters are analysed and pre-processed to improve data quality [1]. The proposed system integrates deep learning models with adversarial training techniques to enhance resistance against malicious perturbations [8], [18]. The experimental results demonstrate that the proposed defence mechanism significantly improves model robustness while maintaining high estimation accuracy [14].

The enhanced model performs better than traditional deep learning approaches that do not include adversarial protection. The framework ensures stable sea state prediction even under hostile or uncertain input conditions, thereby improving the reliability of autonomous marine navigation systems [2], [15]. In future work, the system can be extended by incorporating larger real-time maritime datasets and advanced robust optimization techniques. Further improvements can include lightweight defence mechanisms suitable for onboard deployment in resource-constrained marine vessels. Additionally, integrating real-time monitoring and adaptive defence strategies can further strengthen system security and operational safety [19], [20].

REFERENCES

1. Z. H. Munim, M. Dushenko, V. J. Jimenez, M. H. Shakil, and M. Imset, "Big data and artificial intelligence in the maritime industry: A bibliometric review and future research directions," *Maritime Policy & Management*, vol. 47, no. 5, pp. 577–597, 2020.
2. X.-Y. Zhou, Z.-J. Liu, F.-W. Wang, and S.-K. Ni, "Collision risk identification of autonomous ships based on the synergy ship domain," in *Proc. Chinese Control Decision Conf. (CCDC)*, 2018, pp. 6746–6752.
3. P. Han, H. P. Hildre, and H. Zhang, "Local ocean wave field estimation using a deep generative model of wave buoys," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4209611.
4. N. Roussel *et al.*, "Sea level monitoring and sea state estimate using a single geodetic receiver," *Remote Sens. Environ.*, vol. 171, pp. 261–277, Dec. 2015.
5. U. D. Nielsen, A. H. Brodtkorb, and A. J. Sørensen, "Sea state estimation using multiple ships simultaneously as sailing wave buoys," *Appl. Ocean Res.*, vol. 83, pp. 65–76, Feb. 2019.
6. Z. Ren, X. Han, A. S. Verma, J. A. Dirdal, and R. Skjetne, "Sea state estimation based on vessel motion responses: Improved smoothness and robustness using

Bézier surface and L1 optimization," *Marine Struct.*, vol. 76, Art. no. 102904, Mar. 2021.
7. X. Cheng, G. Li, P. Han, R. Skulstad, S. Chen, and H. Zhang, "Data-driven modeling for transferable sea state estimation between marine systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 3, pp. 2561–2571, Mar. 2022.
8. A.Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," *arXiv preprint arXiv:1706.06083*, 2017.
9. W. Jiang, H. Li, S. Liu, X. Luo, and R. Lu, "Poisoning and evasion attacks against deep learning algorithms in autonomous vehicles," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4439–4449, Apr. 2020.
10. X. Yuan, Z. Zhang, X. Wang, and L. Wu, "Semantic-aware adversarial training for reliable deep hashing retrieval," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 4681–4694, 2023.
11. L.-X. Yang, P. Li, Y. Zhang, X. Yang, Y. Xiang, and W. Zhou, "Effective repair strategy against advanced persistent threat: A differential game approach," *IEEE Trans. Inf. Forensics Security*, vol. 14, pp. 1713–1728, 2019.
12. J.-L. Yin, B. Chen, W. Zhu, B.-H. Chen, and X. Liu, "Push stricter to decide better: A class-conditional feature adaptive framework for improving adversarial robustness," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 2119–2131, 2023.
13. Y. Wang, D. Zou, J. Yi, J. Bailey, X. Ma, and Q. Gu, "Improving adversarial robustness requires revisiting misclassified examples," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2019, pp. 1–14.
14. H. Zhang, Y. Yu, J. Jiao, E. Xing, L. El Ghaoui, and M. Jordan, "Theoretically principled trade-off between robustness and accuracy," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2019, pp. 7472–7482.
15. Y. Yin, X. Cheng, F. Shi, M. Zhao, G. Li, and S. Chen, "An enhanced lightweight convolutional neural network for ship detection in maritime surveillance system," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 5811–5825, Jun. 2022.
16. 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 Trans. Instrum. Meas.*, vol. 69, no. 9, pp. 5984–5993, Sep. 2020.
17. 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 Trans. Ind. Informat.*, vol. 18, no. 2, pp. 891–900, Feb. 2022.
18. I.J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014.
19. Z. Kazemi, A. A. Safavi, M. M. Arefi, and F. Naseri, "Finite-time secure dynamic state estimation for cyber–physical systems under unknown inputs and sensor attacks," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 8, pp. 4950–4959, Aug. 2022.
20. Y. Jiang *et al.*, "Secure data transmission and trustworthiness judgement approaches against cyber–physical attacks in an integrated data-driven framework," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 12, pp. 7799–7809, Dec. 2022.