1 Introduction

As the global climate continues to change, the dynamics of Arctic sea ice become increasingly complex and unpredictable. [1]This complexity poses significant challenges for navigation, ecological conservation, and resource management. Accurate forecasting of sea ice extent and behavior is critical for minimizing risks, optimizing navigation routes, and managing natural resources. [2]Traditional predictive models often struggle with transparency, making it difficult for stakeholders to understand the reasoning behind the predictions. This lack of interpretability hinders informed decision-making, which is vital for addressing the impacts of climate change.

This study seeks to tackle the interpretability challenges in sea ice prediction by leveraging interpretable machine learning techniques. [3]The goal is to develop models that not only provide accurate predictions but also offer clear, transparent insights into the factors driving these predictions. By improving the interpretability of these models, this research aims to empower scientists, policymakers, and other stakeholders with actionable insights, enabling more informed decision-making.

The paper outlines the methodology used to achieve model interpretability in sea ice prediction. It begins by defining the problem and highlighting the challenges associated with the lack of interpretability in existing models. The research objectives include exploring the feasibility of interpretable machine learning models and evaluating their effectiveness in achieving both accuracy and transparency. The methodology involves applying various interpretable machine learning algorithms and evaluation techniques to identify the factors influencing sea ice extent. Finally, the study discusses the expected outcomes and potential implications for advancing machine learning applications in climate research, navigation, and resource management.

Additionally, the study introduces an innovative approach by integrating local interpretable model-agnostic explanations (LIME) to assess feature importance in non-stationary datasets[4]. This approach aims to understand how key factors affecting sea ice cover change over time and across different regions, providing more informed and responsive decision-making. Traditional models have struggled with this complexity, particularly due to the lack of interpretability. By leveraging LIME, which provides insights into the local behavior of any classifier, this study addresses this gap and enhances the interpretability of models predicting Arctic sea ice extent.

1.1 Problem definition

The primary challenge this project addresses is the lack of interpretability in current sea ice prediction models. Traditional models, while powerful in their predictive capabilities, are often shrouded in complexity and opacity, making it difficult for decision-makers to understand and trust the predictions they generate. This hindrance in transparency is particularly problematic in critical sectors such as climate research, navigation, and resource management, where accurate forecasts of sea ice dynamics are of paramount importance.The environmental impact of machine learning models is an increasingly important consideration, especially in the context of climate research and other fields where sustainability is a priority. Traditionally, model selection has focused primarily on performance metrics such as accuracy, precision, and recall. However, as this study illustrates, it is also crucial to consider the environmental cost of deploying these models. High-performance models that generate significant carbon emissions could undermine the sustainability goals they are intended to support, particularly in areas like climate change mitigation, where reducing overall carbon footprints is paramount.

By implementing interpretable machine learning techniques, the project's objective is to construct models that elucidate and provide clear, insightful explanations for their prediction processes. The project underscores the necessity of interpretable models in climate research to validate and refine models, ensuring they better reflect real-world phenomena. In navigation, this interpretability can lead to safer and more efficient routes by understanding the factors influencing sea ice conditions. For resource management, particularly in polar regions, interpretability ensures sustainable practices are informed by solid, understandable data, facilitating timely interventions when required.

2 Data collection

[5]The data for this project were sourced from the dataset "Sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS passive microwave data, version 1," provided by Cavalieri et al.(1996). This dataset has brightness temperature data and aims to offer a continuous time series of sea ice concentrations across various passive microwave instruments[5].

To refine the dataset for this project, several steps were undertaken. Firstly, the data were merged from multiple sources to create a comprehensive dataset covering the desired sea area. Next, a series of data cleaning steps were performed, including checking for missing values, identifying and removing duplicates, and conducting descriptive statistics analysis. Specifically, measures such as skewness and kurtosis were computed to assess the distributional properties of the data.

These refinement processes were crucial to ensure the quality and integrity of the dataset for subse- quent analysis and model development. By addressing issues such as missing values and duplicates, the refined dataset provided a reliable foundation for conducting accurate and insightful sea ice prediction experiments.

2.1 Exploratory Data Analysis (EDA)

The EDA process begins with visualizing the distributions of key features and the target variable. In our dataset, we examined the distributions of Sea Surface Temperature (SST), Sosaline (salt concentration in seawater), and Sea Ice Extent. The plots below illustrate these distributions:

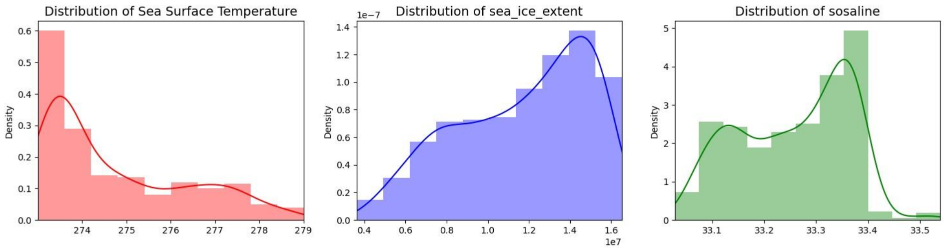

Figure 1: Distribution of some features

The distribution of Sea Surface Temperature (SST) reveals variability in temperature values, providing insights into the temperature range observed in the dataset. The distribution of Sosaline demonstrates the variation in salt concentration levels within the seawater, indicating the diversity of oceanic conditions captured in the dataset.

The distribution of Sea Ice Extent showcases the extent of sea ice coverage, reflecting the seasonal and geographical patterns of sea ice formation and melting. These visualizations offer valuable insights into the central tendencies and variability of the respective features, guiding further analysis and model development.

Next, we explored the relationships between different variables by constructing correlation and covariance matrices:

The correlation heatmap illustrates the correlation coefficients between pairs of variables in the dataset. Positive values indicate a positive linear relationship, while negative values indicate a negative linear relationship. Strong correlations are depicted by values closer to 1 or -1.

The covariance heatmap displays the covariance values between variables, indicating how two variables vary together. Positive covariance suggests that the variables increase or decrease together, while negative covariance implies an inverse relationship. Larger values signify stronger associations between variables.

These visualizations provide insights into the interdependencies among different features, helping to identify potential patterns and relationships within the dataset.

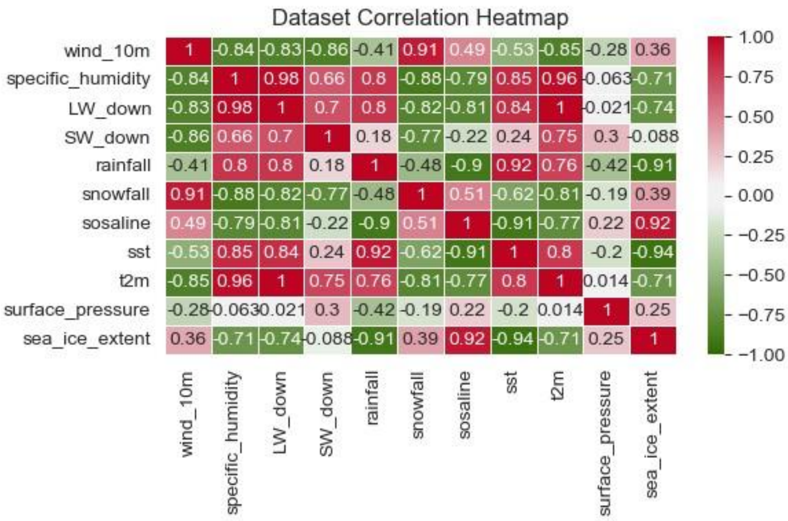

Figure 2: Dataset Correlation Heatmap

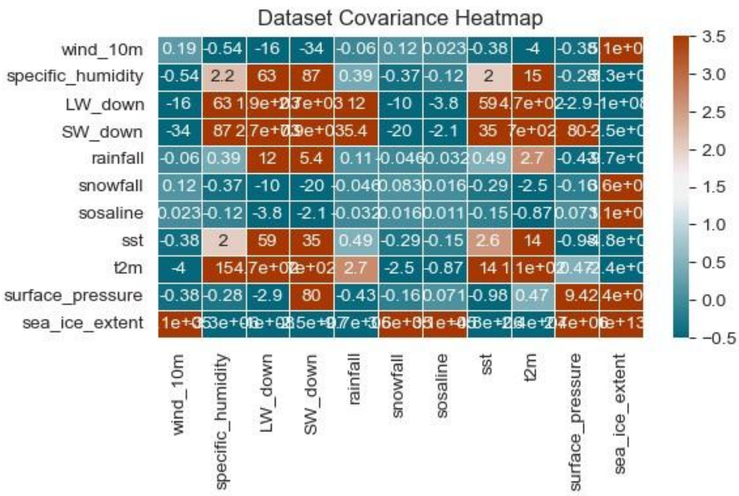

Figure 3: Dataset Covariance Heatmap

3 Implementation

In this project, a variety of machine learning models including Random Forest[6], XGBoost[7], and some other machine learning models will be employed to forecast the extent of Arctic sea ice. The objective is to strike a balance between accuracy and interpretability by carefully selecting the best-performing models from this set.

3.1 Model Selection

Models like Random Forest, XGBoost and etc will be trained and evaluated to forecast the extent of Arctic sea ice[6].The best-performing models will be selected based on their performance on a holdout validation dataset.

3.1.1 Evaluation

To assess the outcomes of our experiments, we will utilize several evaluation metrics including the percentage Root Mean Squared Error (RMSE) loss and R2 scores across all models. These metrics provide insights into the accuracy and goodness-of-fit of the models in predicting the extent of Arctic sea ice.

Specifically, we will compare the percentage RMSE loss and R-squared scores obtained from multiple baseline models trained on the dataset. By analyzing these metrics, we can identify the most optimal model for our forecasting task.

Furthermore, in addition to evaluating model performance, we will also consider the environmental impact of the models by calculating their carbon cost. To accomplish this, we will integrate Code- Carbon into our workflow. CodeCarbon offers estimates of carbon emissions associated with code execution on various hardware resources. By embedding CodeCarbon into our code, we can track the carbon emissions incurred during the training and evaluation of each model.

By incorporating both model performance metrics and carbon cost considerations into our evaluation framework, we aim to make informed decisions regarding modelselection, balancing accuracy with environmental sustainability. This comprehensive evaluation approach ensures that our chosen model not only meets performance requirements but also aligns with our sustainability goals.

3.1.2 Interpretability Analysis

The results of feature importance obtained through the explainable modeling approach will be compared with other feature evaluation methods such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations). This compari- son will provide insights into the importance of different features in the sea ice prediction task and enhance the interpretability of the models[7].

4 Result

The performance of the models trained on the dataset for forecasting the extent of Arctic sea ice varied across different algorithms.

Table 1: Model Evaluation Results

Model	Train R^2 score	Test R^2 score
Linear Regression	0.9805	0.9805
Logistic Regression	0.1276	0.0
Random Forest	0.9974	0.9849
XGBoost	0.999998	0.9695
Neural Network	-0.4125	-0.2555

From the results, we observe that:

Table 2: Mean Squared Error (MSE)

Model	Train MSE	Test MSE
Linear Regression	200,353,936,017.40	185,097,020,558.62
Logistic Regression	-	-
Random Forest	26,604,677,278.11	143,481,548,197.18
XGBoost	15,026,431.22	289,044,916,047.93
Neural Network	14,500,377,895,030.23	11,912,995,179,258.65

Linear Regression, Random Forest, and XGBoost models demonstrate strong performance on both the training and testing datasets, with high R2 scores and relatively low Mean Squared Error (MSE) values. Logistic Regression and Neural Network models exhibit poor performance, as indicated by their low R2 scores and high MSE values. Among the models evaluated, Random Forest achieves the highest R2 score on the testing dataset, indicating its effectiveness in predicting the extent of Arctic sea ice.

Additionally, the scatter plot visualization for the Random Forest model demonstrates a strong correlation between true values and predicted values, further supporting its efficacy in forecasting sea ice extent.These results provide valuable insights into the performance of different machine learning models in predicting Arctic sea ice extent, facilitating informed decision-making in climate research and resource management.

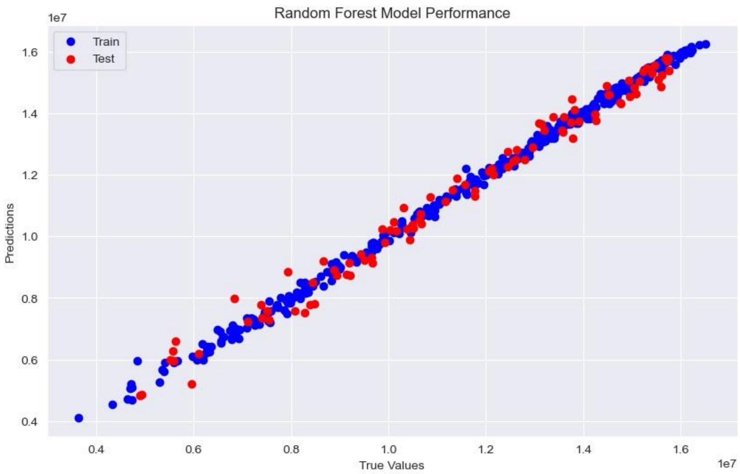

Figure 4: Random Forest Model Performance

4.1 Explainable Machine Learning

The resulting plot provides valuable insights into the features that significantly influence the prediction of the target variable.

Notably, the Sea Surface Temperature (SST) emerges as the most influential feature.

The ICE plot vividly illustrates that as SST increases, there is a discernible decrease in the extent of ice cover. This correlation underscores the crucial role SST plays in shaping sea ice extent dynamics, offering a clear visualization of its impact on the ecosystem. Later on, we employed XAI methods such as Shap and LIME to interpret the model.

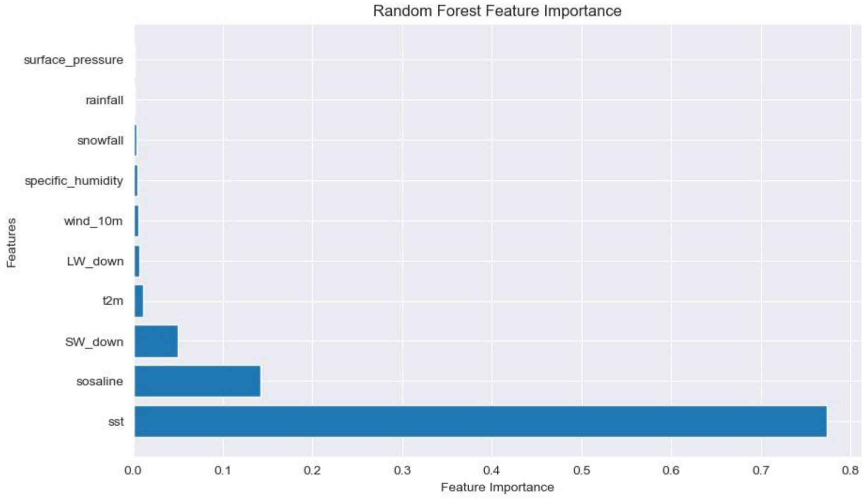

Figure 5: Random Forest feature importance

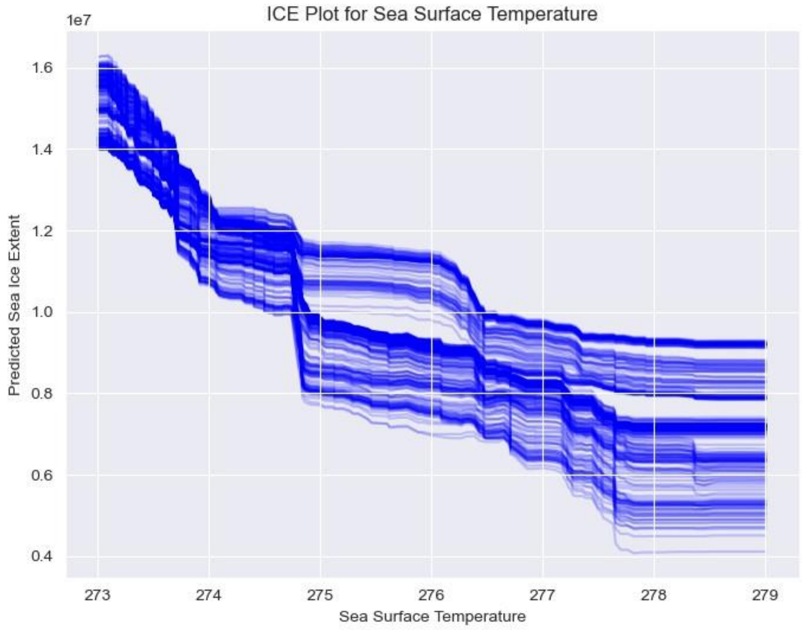

Figure 6: ICE Plot for Sea Surface Temperature

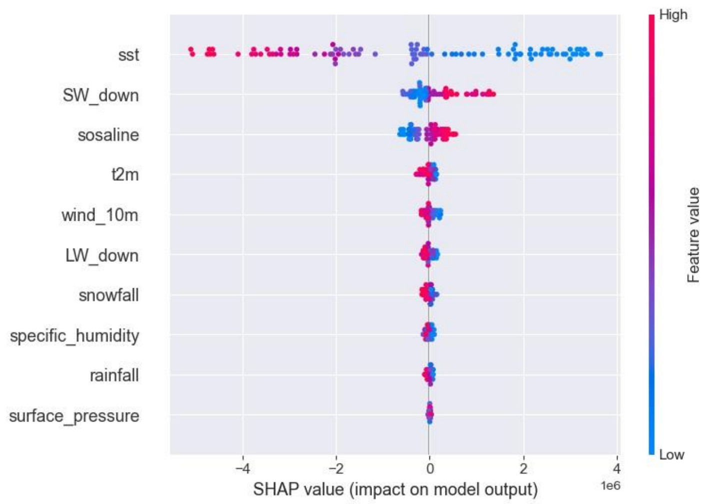

Figure 7: SHAP Value

SHAP and Random Forest both identified Sea Surface Temperature (SST) as the most influential feature in predicting sea ice extent, indicating a strong consensus on the importance of this variable. This alignment between SHAP and Random Forest underscores the robustness of SST as a critical determinant in the model's overall predictions, suggesting that changes in SST are consistently and significantly correlated with variations in sea ice coverage. The global feature importance provided by these models offers valuable insights into the key drivers of sea ice dynamics across the entire dataset, making it a useful tool for understanding broader patterns and trends.

However, LIME presented a different perspective by offering a more granular, localized view of feature importance. Unlike SHAP and Random Forest, which provide a holistic ranking of features based on their average contribution across the dataset, LIME focuses on individual predictions, calculating feature importance for each specific data point. This approach highlights the inherent complexity and variability of real-world data, where the significance of certain features can fluctuate depending on the specific context or conditions of each instance. For example, while SST might generally be the most critical factor in predicting sea ice extent, other features such as salinity, wind speed, or regional geographic characteristics might play a more dominant role in certain localized scenarios or under specific conditions.

LIME's ability to provide this local feature importance is particularly valuable in situations where understanding the nuances of specific predictions is crucial. For instance, in cases where the model is used to forecast sea ice conditions for particular regions or during specific time frames, LIME can help identify which factors are most influential under those particular circumstances. This localized insight can be instrumental for decision-makers who need to understand the specific drivers behind a particular prediction, enabling them to tailor their responses more effectively.

Moreover, the difference in feature importance rankings between SHAP, Random Forest, and LIME reflects the broader debate in the field of machine learning about the trade-offs between global and local interpretability. While global interpretability offers a broad overview of model behavior, which is useful for general understanding and policy-making, local interpretability provides detailed insights into individual predictions, which is essential for context-specific decision-making. In practice, combining both approaches allows for a more comprehensive understanding of the model's behavior, enabling users to gain insights at both the macro and micro levels.

In summary, while SHAP and Random Forest offer a consistent view of global feature importance, with SST emerging as the most critical factor, LIME provides a more nuanced perspective by revealing how feature importance can vary across different data points. This distinction is crucial for applications where localized insights are needed to understand and act on specific predictions. By using these complementary approaches, stakeholders can achieve a more balanced and informed understanding of the factors influencing sea ice predictions, thereby enhancing the model's utility in both broad and targeted contexts.

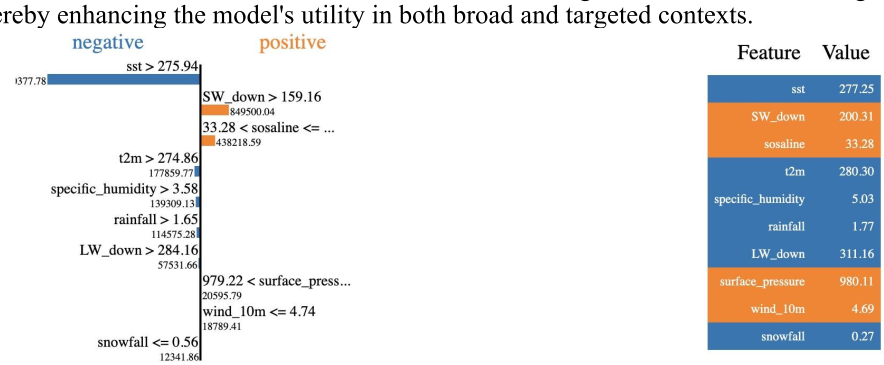

Figure 8: LIME

4.2 Carbon Emission

Table 3: Carbon Emissions and Machine Models

Timestamp	Project Name	Emissions (kg CO2)	Machine Model
2024-04-24T15:09:18	Xgboost	2.04×10^{-7}	Apple M1 Pro
2024-04-24T15:09:20	Random Forest	7.89×10^{-8}	Apple M1 Pro

2024-04-24T15:09:20	SHAP	6.32×10^{-9}	Apple M1 Pro
2024-04-24T15:09:20	LIME	3.29×10^{-8}	Apple M1 Pro

Based on the provided data, we can observe that different machine learning models generated varying levels of carbon emissions during their execution. In this experiment conducted in the District of Columbia, United States, the Xgboost model emitted approximately 2.04×10^{-7} kg CO_2, while the Random Forest model emitted around 7.89×10^{-8} kg CO_2. The SHAP model produced about 6.32×10^{-9} kg CO_2, and the LIME model emitted roughly 3.29×10^{-8} kg CO_2. These figures indicate that different models generated different amounts of carbon emissions under the same environmental conditions, possibly influenced by factors such as model complexity and training duration. When selecting machine learning models, besides considering performance metrics, it is essential to factor in their environmental impact to make more sustainable decisions

4.1 Conclusion

This study has demonstrated the significant role that explainable machine learning models can play in enhancing the predictability and transparency of sea ice forecasting. By applying advanced models like Random Forest and XGBoost, and integrating XAI techniques such as SHAP and LIME, we not only achieved high accuracy in predictions but also gained deeper insights into the underlying factors driving these predictions. The ability to interpret model outputs and understand the influence of key variables, such as Sea Surface Temperature (SST), is crucial for advancing both the scientific understanding and practical application of sea ice forecasting.

One of the key findings from this research is the critical role that SST plays in determining sea ice extent. As the global climate continues to warm, understanding this relationship becomes increasingly important for predicting future trends in Arctic sea ice coverage. The ability to explain why certain predictions are made enables scientists and policymakers to develop more targeted and effective strategies for mitigating the impacts of sea ice loss. For instance, policies aimed at reducing greenhouse gas emissions could be informed by the clear evidence of the impact of rising temperatures on sea ice dynamics.

Moreover, the integration of carbon emission tracking into the model evaluation process highlights an often overlooked aspect of machine learning: its environmental footprint. By considering both the performance and the sustainability of the models used, this study underscores the importance of aligning scientific research with broader environmental goals. This holistic approach is especially relevant in the context of climate change research, where the tools used to understand and mitigate the problem should themselves be sustainable.

From a policy perspective, the insights gained from this study have several implications. First, there is a need for continued investment in climate research that leverages advanced, interpretable machine learning models. These models not only enhance the accuracy of predictions but also improve the transparency and trustworthiness of the results, making them more actionable for policymakers. Second, the results suggest that adaptive management strategies, which account for the dynamic and region-specific nature of sea ice loss, could be more effective. By understanding the local and temporal variability in sea ice extent, policies can be tailored to address the most vulnerable areas and times of year.

Furthermore, the ability to communicate complex model outputs in an understandable way is essential for effective policy-making. Tools like SHAP and LIME can bridge the gap between technical research and practical decision-making, allowing policymakers to see the rationale behind predictions and to make informed decisions based on robust evidence. This transparency is critical for gaining public trust and ensuring that policies are both scientifically sound and socially accepted.

In addition to policy-making, the findings of this study have direct implications for industries that operate in or near the Arctic, such as shipping, fishing, and natural resource extraction. Predictive models that are both accurate and interpretable can help these industries plan more effectively, reducing risks and ensuring sustainable practices. For instance, understanding the factors that lead to sea ice retreat can help shipping companies optimize routes to minimize the risk of ice-related accidents, while also reducing fuel consumption and emissions.

In conclusion, the integration of explainable machine learning models into sea ice forecasting represents a significant step forward in both scientific understanding and practical application. By improving the transparency of predictions and aligning model development with environmental sustainability, this research provides a valuable foundation for more informed and effective decision-making in climate policy and resource management. To protect sea ice extent and mitigate the impacts of climate change, it is essential that future research continues to refine these models and that policymakers leverage these insights to develop adaptive, evidence-based strategies that address the complex challenges of a warming Arctic.