Deep Learning Benchmarks for short-term Arctic Sea Ice Forecasting

Bo-Ram Kim [1], Junhwa Chi [1, 2]

[1] Dept. of Data Engineering, Pukyong National University, Busan, Korea – marob@pukyong.ac.kr
[2] Major of Big Data Convergence, Division of Data Information Sciences, Pukyong National University, Busan, Korea – jchi@pknu.ac.kr

Commission III, WG III/7

Keywords: Arctic sea ice, Short-term forecasting, Deep learning, CNN, Transformer, Northern Sea Route.

Abstract

The rapid decline of Arctic sea ice has increased spatiotemporal variability in the core marginal seas along the Northern Sea Route (NSR), thereby hindering reliable short-term forecasting. While deep learning models offer computationally efficient alternatives to physics-based numerical models, previous studies have often relied on global average errors and have provided limited assessment of boundary-preservation performance. In addition, few studies have systematically compared spatial feature extraction strategies within non-recurrent spatiotemporal architectures for sea ice forecasting. To address this gap, this study benchmarks CNN backbone and Transformer backbone families for 10-day sea ice forecasting. The evaluation focuses on five dynamic marginal seas along the NSR. Metrics included Integrated Ice Edge Error (IIEE), Mean Boundary Error (MBE), Intersection over Union (IoU), and Anomaly Correlation Coefficient (ACC). Results indicate that CNN-based models, including PoolFormer, generally show more favorable performance in boundary preservation and prediction stability than Transformer-based models. Regionally, the Barents and Kara Seas were more difficult to predict, whereas the Laptev and East Siberian Seas were relatively more predictable. The Chukchi Sea exhibited particularly high uncertainty during rapid summer ice retreat. Across all models, boundary-based performance degraded significantly as lead time increased, with MBE exceeding 30 km from T+5 onward. These results suggest that, under the current univariate setting, boundary-preservation performance remains relatively stable at short lead times, but degrades thereafter. Overall, CNN-based non-recurrent architectures show relative strengths over the NSR test regions, although future work incorporating multivariate inputs is needed to extend reliable lead times.

1. Introduction

In recent decades, the Arctic has warmed at a significantly faster rate than the global average, leading to an overall decline in both the extent and duration of Arctic sea ice (Laliberté et al., 2016). This transformation has heightened interest in environmental changes in Arctic waters while further amplifying the interannual and seasonal variability, as well as local uncertainty, of sea ice. In particular, marginal seas adjacent to the Northern Sea Route (NSR) experience rapid sea ice advance and retreat, with spatial distributions varying drastically over time and space. Consequently, short-term forecasting methods capable of accurately reproducing the position and morphology of the ice edge—beyond merely capturing average changes—are crucial. Therefore, sea ice forecasting in these dynamic regions should be evaluated based on its ability to accurately reproduce nonlinear spatiotemporal fluctuations and spatial boundary structures, rather than simply estimating changes in total sea ice extent.

The accumulation of satellite-based Earth observation data has opened new possibilities for sea ice forecasting. Specifically, passive microwave-derived sea ice concentration (SIC) data provide consistent long-term time series and serve as a core input for data-driven forecasting models (DiGirolamo et al., 2022). Traditionally, sea ice forecasting has relied on statistical approaches and physics-based numerical models. However, statistical models are limited in their ability to represent nonlinear spatiotemporal patterns of sea ice, whereas physics-based numerical models involve high computational costs and are sensitive to errors in initial conditions and data assimilation (Kvanum et al., 2025; Palerme et al., 2024). As an alternative, deep learning-based models that directly learn from large observational datasets have recently been widely adopted. Among early efforts, Chi and Kim (2017) presented the first application of deep neural network to Arctic SIC prediction using only observational data, demonstrating that fully data-driven approaches can effectively capture large-scale sea ice variability. Since then, a number of follow-up studies have further advanced this direction. For example, Andersson et al. (2021) utilized IceNet, a U-Net-based probabilistic deep learning model, achieving seasonal forecasting performance superior to conventional physical models. In addition, Chi et al. (2021) proposed a two-stream convolutional long- and short-term memory model with perceptual loss for sequence-to-sequence Arctic SIC prediction, highlighting the potential of deep learning architectures to better preserve spatial structure and temporal evolution in sea ice fields. More recently, Palerme et al. (2024) and Kvanum et al. (2025) applied deep learning to high-resolution short-term SIC forecasting, illustrating the potential for accuracy improvements at local scales.

However, the key challenge in short-term sea ice forecasting lies not only in reducing average errors, but in how reliably a model can reproduce dynamic spatiotemporal patterns and ice edge boundaries. Previous sea ice forecasting studies have often relied on qualitative evaluations of difference maps between forecasts and observations or on quantitative assessments using aggregate statistics such as total Pan-Arctic sea ice extent. However, these global average errors (e.g., Root Mean Square Error, RMSE) or simple area-based metrics are insufficient to fully explain how well a model captures climatological variability or preserves the position and morphology of the ice edge during periods of rapid melting. To address this, Goessling et al. (2016) introduced the Integrated Ice Edge Error (IIEE) as an objective spatial score for discrepancies in ice edge location, proposing a more refined approach to evaluating sea ice forecasts. Thus, assessing short-term sea ice forecasting performance requires considering not only average errors but also boundary-centric spatial characteristics.

A model-architecture perspective is also important. In the spatiotemporal forecasting literature, recurrent neural network (RNN)-based models, such as ConvLSTM (Shi et al., 2015), have served as a foundational approach. However, due to their sequential processing of past states, recurrent architectures can incur cumulative inference latency. Furthermore, as lead time increases, spatial information may gradually degrade, potentially leading to blurred and less distinct boundaries. Accordingly, non-recurrent spatiotemporal forecasting frameworks that enable parallel computation have recently gained attention as viable alternatives. Within these non-recurrent families, architectures can be divided into two main streams based on their spatial feature extraction methods. One group is the Convolutional Neural Network (CNN) family, which learns local spatial patterns and boundary structures through convolution operations (Gao et al., 2022; Guo et al., 2023; Li et al., 2024; Liu et al., 2022; Rao et al., 2022; Tan et al., 2025; Trockman and Kolter, 2022; Yu et al., 2022). The other is the Transformer family, which captures global spatial context and long-range dependencies through self-attention mechanisms (Dosovitskiy et al., 2021; Li et al., 2023; Liu et al., 2021; Tan et al., 2023; Tolstikhin et al., 2021). Therefore, to better understand sea ice forecasting performance, it is necessary to go beyond a general comparison of recurrent and non-recurrent structures and instead examine the distinct forecasting characteristics arising from different spatial feature extraction methods within non-recurrent models—namely, the CNN and Transformer families.

Accordingly, rather than establishing a general model hierarchy for the entire Pan-Arctic domain, this study presents a benchmark designed to evaluate the boundary-preservation performance of non-recurrent spatiotemporal forecasting models. This is achieved by designating five core marginal seas along the NSR—regions characterized by high spatiotemporal variability and prediction difficulty—as testbeds. To this end, the models were configured to forecast sea ice for the next 10 days (T+1 to T+10) based on an initialization date (T), using input data from the preceding 30 days. Furthermore, to provide a more comprehensive evaluation of model performance, this study integrates multiple types of metrics rather than relying solely on global average errors. First, the Anomaly Correlation Coefficient (ACC) is used to evaluate how well the models capture sea ice anomaly patterns relative to climatology. Second, building on the framework proposed by Goessling et al. (2016), boundary-centric metrics such as IIEE, Mean Boundary Error (MBE), and Intersection over Union (IoU) are used to quantify the ability to preserve the position and morphology of the sea ice edge.

The contributions of this study are threefold. First, non-recurrent spatiotemporal forecasting model families are compared under identical data and training conditions across five core marginal seas along the NSR. Second, sea ice forecasting performance is evaluated from multiple perspectives by jointly analyzing pattern reproduction using ACC and boundary preservation using IIEE, MBE, and IoU. Third, by analyzing performance degradation with increasing lead time (T+1 to T+10), this study identifies the approximate range over which boundary-based forecasting performance remains relatively stable.

2. Data and Methods

2.1 Sea Ice Data

This study utilizes daily SIC data from the "Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data (NSIDC-0051, Version 2)" provided by the National Snow and Ice Data Center (NSIDC) (DiGirolamo et al., 2022). These data provide consistent long-term daily SIC information across the Pan-Arctic region and are available on a 25 km grid using the NSIDC Sea Ice Polar Stereographic North projection. To ensure temporal continuity, the period from February 1, 1988, to December 31, 2023, which is free of long-term data gaps, is selected for this study.

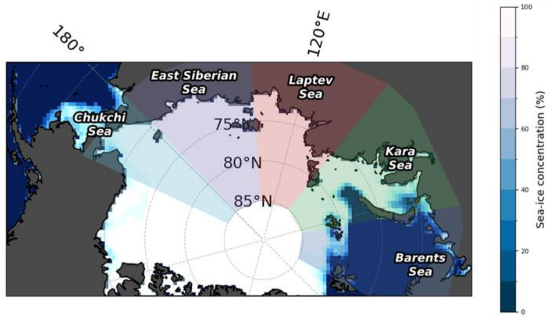

Figure 1. Study domain and five target regions along the NSR: the Barents, Kara, Laptev, East Siberian, and Chukchi Seas.

During data preprocessing, land and coastal areas are masked and excluded from training and evaluation. Additionally, the data void near the North Pole (the pole hole), caused by the lack of satellite observations, is filled using linear interpolation. For computational efficiency and to enable fair comparison across diverse model architectures, the original SIC data are downsampled to a 50 km × 50 km grid. This spatial resampling reduces computational burden while preserving the large-scale spatiotemporal patterns required for short-term sea ice forecasting. The forecasting models are designed to use the past 30 days of SIC data to predict SIC for the subsequent 10 days. As shown in Figure 1, the spatial domain of the study is restricted to five core marginal seas along the NSR—the Barents, Kara, Laptev, East Siberian, and Chukchi Seas—which are characterized by high spatiotemporal variability and prediction difficulty. This restriction aligns with the study's objective: rather than addressing general Pan-Arctic sea ice forecasting performance, the goal is to rigorously evaluate the models' ability to preserve ice edges in highly dynamic regions where prediction is most challenging.

While strictly maintaining chronological order, the entire dataset is partitioned into a training and validation pool spanning 1988–2020 and a fixed test set covering 2021–2023. This chronological split is designed to prevent data leakage and to maintain a validation structure consistent with real-world time-series forecasting conditions.

2.2 Benchmark Models

This study focuses on non-recurrent spatiotemporal forecasting architectures, which are broadly categorized into two backbone families. The first group is the CNN backbone family. This family is designed for parallel computation and stable spatial representation by extracting spatiotemporal features through convolution-based spatial encoding and token mixing, without relying on recurrent structures. The CNN-based models considered in this study include ConvMixer (Trockman and Kolter, 2022), ConvNeXt (Liu et al., 2022), HorNet (Rao et al., 2022), VAN (Guo et al., 2023), MogaNet (Li et al., 2024), SimVP (Gao et al., 2022), SimVPv2 (Tan et al., 2025), and PoolFormer (Yu et al., 2022).

The second group is the Transformer backbone family. This family excels at modeling global context and long-range spatial dependencies through self-attention or related token-mixing mechanisms. The Transformer-based models considered in this

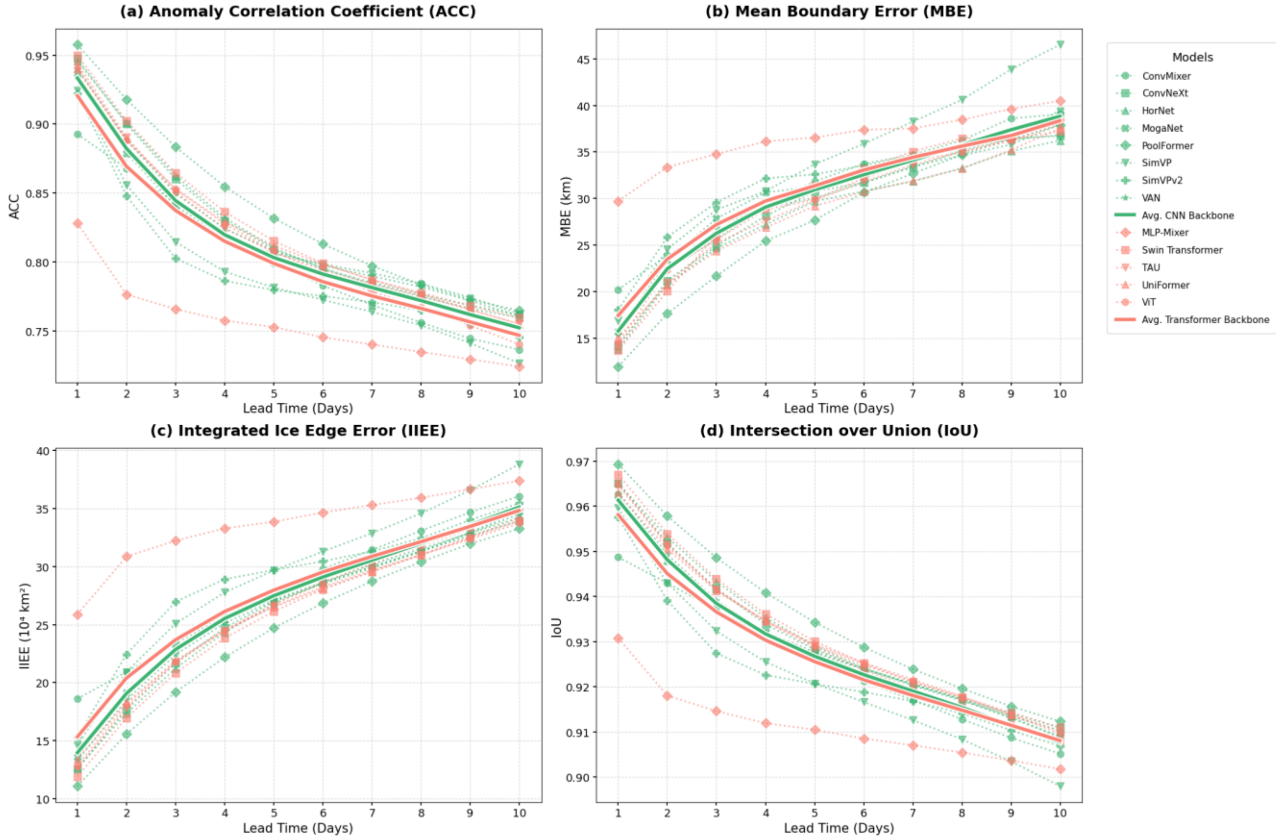

Figure 2. Lead time-dependent performance of benchmarked non-recurrent model families over the five NSR seas during the 2021–2023 test period. Thin lines denote individual models, while thick lines represent family means. The panels show (a) ACC, (b) MBE, (c) IIEE, and (d) IoU from T+1 to T+10.

study include ViT (Dosovitskiy et al., 2021), Swin Transformer (Liu et al., 2021), UniFormer (Li et al., 2023), TAU (Tan et al., 2023), and MLP-Mixer (Tolstikhin et al., 2021). In total, 13 non-recurrent models are compared, comprising eight CNN-based and five Transformer-based models. Rather than claiming the absolute superiority of any single model, this study focuses on comparing the common strengths and weaknesses of these two backbone families for short-term sea ice forecasting over the NSR test regions.

2.3 Evaluation Metrics

To rigorously evaluate model performance, this study employs boundary-centric metrics that directly assess the preservation and overlap of sea ice boundaries, rather than relying on conventional global average errors such as Mean Absolute Error (MAE) and RMSE. The sea ice edge is defined using a 15% SIC threshold, a widely used standard for determining the presence of sea ice (Cavalieri et al., 1996). Prior to calculating the boundary metrics, morphological filtering, including the removal of small objects, is applied to prevent sporadic small-scale ice patches or near-coastal noise from distorting distance-based measurements.

First, MBE represents the average distance error between predicted and observed boundaries, and serves as a direct indicator of spatial positional accuracy. Second, IIEE quantifies the total area of disagreement between predictions and observations regarding the presence of sea ice, thereby measuring the spatial magnitude of boundary errors. Third, IoU indicates the spatial overlap between predicted and observed sea ice areas, jointly evaluating boundary shape preservation and regional agreement. Fourth, ACC is used as a supplementary metric reflecting predictive performance in capturing anomaly structures relative to climatology. In this study, ACC calculations are restricted to the Marginal Ice Zone (MIZ) (15% < SIC < 85%), where concentration changes are most active. These four metrics evaluate model performance from the perspectives of distance, area, morphology, and variability patterns, respectively, thereby capturing sea ice boundary prediction capabilities more directly than simple average errors.

2.4 Experimental Setup

To enable a controlled comparison of the inherent spatiotemporal pattern-learning capabilities of the deep learning architectures, the input variable is restricted to SIC. For validation, a walk-forward validation scheme is adopted to strictly maintain chronological order. Because sea ice time series exhibit both strong seasonality and long-term trends, random data splitting can lead to overly optimistic performance estimates. Therefore, to ensure fairness in the relative comparison among models, identical time-series splitting rules and optimization conditions are uniformly applied across all models.

Z-score normalization is applied to the SIC data, which serves as both the input and the target variable. All models are implemented within a unified framework, and the Mean Squared Error (MSE) over the entire 10-day forecast horizon is used as the common loss function. Optimization is performed using the AdamW optimizer with an initial learning rate of 10^{-4}, coupled with a Cosine Annealing scheduler. Validation is conducted using a 5-fold walk-forward validation scheme. To ensure the reliability of the inter-model comparisons, consistent experimental settings—such as batch size, number of epochs, and hardware configuration—are maintained across all models.

3. Results

3.1 Lead Time-Dependent Forecasting Performance

Figure 2 shows lead time-dependent changes in forecasting performance across the five NSR seas during the 2021–2023 test period. Overall, all models exhibit a gradual decline in forecasting performance as lead time increases. A comparison of family-mean performance across lead times shows that the CNN backbone family generally achieves lower boundary errors and higher overlap than the Transformer backbone family. For ACC, both families maintained high values at shorter lead times; however, the advantage of the CNN backbone family is more consistently evident in the boundary-centric metrics.

At the individual model level, PoolFormer shows the best performance in boundary-based metrics. It records the lowest IIEE across the entire forecast range, with values of 11.09 10^4 km² at T+1 and 33.26 10^4 km² at T+10. It also achieved the highest IoU, with values of 0.9693 at T+1 and 0.9123 at T+10. For ACC, PoolFormer achieves the highest values, reaching 0.9576 at T+1 and 0.7646 at T+10.

Furthermore, as shown in Figure 2b, the family-mean MBE for the CNN backbone family exceeds 30 km at T+5 and surpasses 35 km at T+8. At the same time, IIEE increases and IoU decreases, indicating that both boundary-position accuracy and spatial overlap deteriorate as lead time increases.

3.2 Performance Differences Across Specific NSR Regions

Table 1 compares the average performance of the CNN backbone family and the Transformer backbone family across the five core marginal seas of the NSR. The absolute values of ACC, MBE, IIEE, and IoU vary significantly by region, indicating clear differences in prediction difficulty despite identical experimental settings. The Barents Sea records the lowest IoU (approximately 0.845–0.848) and the highest MBE among all regions. The MBE is 23.48 km for the CNN backbone family and 24.11 km for the Transformer backbone family, indicating substantially larger boundary errors compared to other seas. The Kara Sea also exhibits relatively high boundary errors, with MBE ranging from 12.42 to 13.36 km.

In contrast, the Laptev and East Siberian Seas show comparatively lower MBE and higher IoU. For the Laptev Sea, MBE ranges from 8.54 to 9.47 km and IIEE from 2.64 to 2.68 × 10^4 km², representing the lowest errors among all regions. In the East Siberian Sea, the CNN family slightly outperforms the Transformer family, with an ACC of 0.786, an MBE of 8.58 km, and an IoU of 0.949.

Region	Family	ACC (↑)	MBE (km) (↓)	IIEE (10^4km ²) (↓)	IoU (↑)
Barents Sea	CNN	0.715	23.48	7.22	0.848
	Trans-former	0.713	24.11	7.33	0.845
Kara Sea	CNN	0.746	12.42	5.15	0.920
	Trans-former	0.738	13.36	5.29	0.918
Laptev Sea	CNN	0.711	8.54	2.64	0.951
	Trans-former	0.711	9.47	2.68	0.951
East Siberian Sea	CNN	0.786	8.58	3.91	0.949
	Trans-former	0.778	8.96	4.05	0.948
Chukchi Sea	CNN	0.865	12.41	7.30	0.920
	Trans-former	0.860	12.57	7.38	0.919

Table 1. Regional mean performance of the CNN backbone family and the Transformer backbone family across the NSR

The Chukchi Sea maintains relatively high ACC vlaues (0.860–0.865); however, it exhibits larger boundary errors than the Laptev and East Siberian Seas, with IIEE values of approximately 7.30–7.38 × 10^4 km² and MBE ranging from 12.41 to12.57 km.

3.3 Qualitative Spatial Error Analysis

Figure 3 presents a 10-day prediction example from PoolFormer, the best-performing model in the quantitative evaluation. The case is initialized on July 29, 2023, and covers the forecast period from July 30, 2023 to August 8, 2023. In each panel, observed sea ice edge contours are overlaid on the predicted SIC fields, with yellow indicating overestimation and red indicating underestimation.

During the early forecast period (T+1 to T+3), the predicted boundaries generally align well with the observations, and the areas of over-prediction and missed ice remain limited. However, as lead time increases, error regions gradually expand near the ice edge. Notably, after T+7, both over-predicted and missed areas increase simultaneously along some coastal

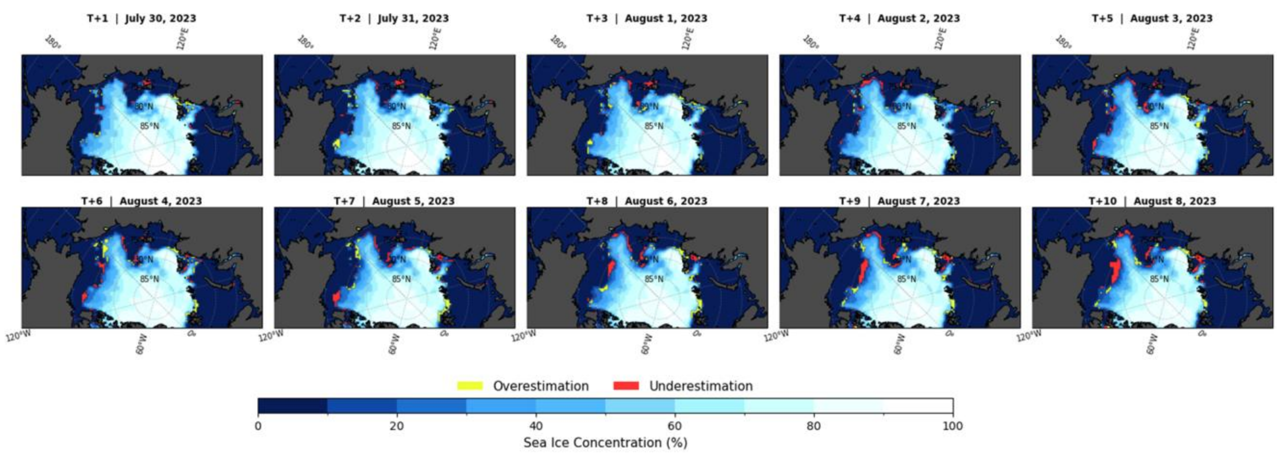

Figure 3. Example of 10-day sea ice forecasts produced by PoolFormer, initialized on July 29, 2023 and covering the period from July 30, 2023 to August 8, 2023 (T+1 to T+10). Observed ice edge contours are overlaid on the predicted SIC fields; yellow indicates overestimation, and red indicating underestimation.

sections and regions experiencing rapid ice retreat. By T+10, boundary disagreement becomes much more extensive compared to early lead times. These qualitative results are consistent with the quantitative trends shown in Figure 2—namely, increasing MBE and IIEE and decreasing IoU with longer lead times—and clearly illustrate the temporal limits of forecast reliability.

4. Discussion

Across the five core marginal seas of the NSR, the CNN backbone family generally shows more stable performance in boundary-centric metrics than the Transformer backbone family. This tendency is most evident in family-mean lead time comparisons, although several regional metrics are comparable between the two families and, in some cases, slightly favor the Transformer family. These results suggest that, under the short-term forecasting setup of this study, the ability to represent local morphological changes and spatial transitions of the sea ice edge plays a more direct role than capturing broader spatial context.

Performance degradation with increasing lead time is one of the key findings of this study. In the quantitative results, the family-mean MBE for the CNN group exceeds 30 km at T+5 and surpasses 35 km at T+8, accompanied by increasing IIEE and decreasing IoU. The qualitative case study also shows that, while predicted and observed boundaries align well at shorter lead times, areas of over- and under-prediction near the boundary expand as lead time increases. This suggests that boundary sharpness and positional alignment—critical factors in sea ice forecasting—deteriorate before large increases in average error become evident. Therefore, under the current univariate setting, boundary-based performance remains relatively stable at short lead times but degrades more rapidly thereafter.

The regional analysis also provides important insights. The Barents and Kara Seas exhibit larger boundary errors than other regions, whereas the Laptev and East Siberian Seas show more stable performance. The Chukchi Sea shows high ACC but relatively large boundary errors. This suggests that, even under identical experimental settings, prediction difficulty varies depending on regional sea ice dynamics and spatial variability. In other words, even within the NSR, model performance differs by region, and overall average performance alone is insufficient to explain these differences. These differences suggest that future regional-scale sea ice forecasting studies should consider regional characteristics when selecting models and interpreting performance.

Furthermore, to enable a controlled comparison among architectures, this study restricts the input variable to SIC and considers only non-recurrent models under identical training conditions. While this setup enhances the interpretability of the benchmark, it does not capture all environmental factors influencing sea ice variability. Including auxiliary variables such as meteorological fields, sea surface temperature, wind vectors, and ocean currents may reduce uncertainties, particularly at longer lead times. Future research should explore multivariate or hybrid architectures that incorporate additional environmental variables while retaining the boundary-preservation strengths of the CNN backbone family.

5. Conclusions

This study compares two non-recurrent spatiotemporal forecasting families with different spatial feature extraction mechanisms—CNN and Transformer backbones—for short-term sea ice forecasting over five core marginal seas along the NSR. To this end, rather than relying on global average errors, boundary-centric metrics such as IIEE, MBE, IoU, and ACC are used to evaluate the preservation of ice edge location, spatial overlap, and anomaly patterns.

The results show that, in family-mean lead time comparisons, the CNN backbone family generally demonstrates more stable boundary forecasting performance than the Transformer backbone family, with PoolFormer achieving the best overall performance among individual models. Moreover, the regional analysis reveals clear differences in prediction difficulty within the NSR. The Barents and Kara Seas exhibited relatively large boundary errors, whereas the Laptev and East Siberian Seas show more stable performance. The Chukchi Sea shows high overall pattern agreement but recorded relatively large boundary errors.

Additionally, boundary-based performance degrades significantly in all models as lead time increases, with boundary position errors rising rapidly, particularly from T+5 onward. This suggests that, under the current univariate setting, boundary-based forecasting performance remains relatively reliable at short lead times but degrades progressively beyond the mid-range forecast horizon. In conclusion, using the highly variable core seas of the NSR as a testbed, this study demonstrates that the CNN backbone family has relative strengths in short-term sea ice forecasting. Future research should incorporate multivariate inputs and structural improvements to reduce uncertainty at longer lead times and better account for regional characteristics.

Acknowledgements

This work was supported by the Korea Polar Research Institute (KOPRI) grant funded by the Ministry of Oceans and Fisheries (KOPRI PE26911), and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2023-00242528).

References

Andersson, T. R., Hosking, S., Ballance, W. T., de Burgh-Day, A. N. B. G., Graham, E., et al., 2021. Seasonal Arctic sea ice forecasting with probabilistic deep learning. *Nature Communications*, 12, 5124. doi.org/10.1038/s41467-021-25257-4.

Cavalieri, D. J., Parkinson, C. L., Gloersen, P., Zwally, H. J., 1996. Sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS passive microwave data (NSIDC-0051, Version 1). NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, CO, USA. doi.org/10.5067/8GQ8LZQVL0VL.

Chi, J., Bae, J., Kwon, Y.-J., 2021. Two-Stream Convolutional Long- and Short-Term Memory Model Using Perceptual Loss for Sequence-to-Sequence Arctic Sea Ice Prediction. *Remote Sensing*, 13(17), 3413. doi.org/10.3390/rs13173413.

Chi, J., Kim, H., 2017. Prediction of Arctic Sea Ice Concentration Using a Fully Data Driven Deep Neural Network. *Remote Sensing*, 9(12), 1305. doi.org/10.3390/rs9121305.

DiGirolamo, N., Parkinson, C. L., Cavalieri, D. J., Gloersen, P., Zwally, H. J., 2022. Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data, NSIDC-0051, Version 2. NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, CO, USA. doi.org/10.5067/MPYG15WAA4WX.

Dosovitskiy, A., et al., 2021. An image is worth 16x16 words: Transformers for image recognition at scale. In: *Proc. ICLR*.

Gao, Z., et al., 2022. SimVP: Simpler yet better video prediction. In: *Proc. IEEE/CVF CVPR*, pp. 3170–3180.

Goessling, H. F., Tietsche, S., Day, J. J., Hawkins, E., Jung, T., 2016. Predictability of the Arctic sea ice edge. *Geophysical Research Letters*, 43(4), pp. 1642–1650. doi.org/10.1002/2015GL067232.

Guo, M.-H., et al., 2023. Visual attention network. *Computational Visual Media*, 9(4), pp. 733–752. doi.org/10.1007/s41095-023-0364-2.

Kvanum, A. F., Palerme, C., Müller, M., Rabault, J., Hughes, N., 2025. Developing a deep learning forecasting system for short-term and high-resolution prediction of sea ice concentration. *The Cryosphere*, 19, pp. 4149–4166. doi.org/10.5194/tc-19-4149-2025.

Laliberté, F., Howell, S. E. L., Kushner, P. J., 2016. Regional variability of a projected sea ice-free Arctic during the summer months. *Geophysical Research Letters*, 43(1), pp. 256–263. doi.org/10.1002/2015GL066855.

Li, K., et al., 2023. UniFormer: Unifying convolution and self-attention for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(10), pp. 12581–12600. doi.org/10.1109/TPAMI.2023.3282631.

Li, S., et al., 2024. MogaNet: Multi-order gated aggregation network. In: *Proc. IEEE/CVF CVPR*, pp. 6239–6248.

Liu, Z., et al., 2021. Swin Transformer: Hierarchical vision transformer using shifted windows. In: *Proc. ICCV*, pp. 10012–10022.

Liu, Z., et al., 2022. A ConvNet for the 2020s. In: *Proc. IEEE/CVF CVPR*, pp. 11976–11986. doi.org/10.1109/CVPR52688.2022.01167.

Palerme, C., Lavergne, T., Rusin, J., Melsom, A., Brajard, J., Kvanum, A. F., Sørensen, A. M., Bertino, L., Müller, M., 2024. Improving short-term sea ice concentration forecasts using deep learning. *The Cryosphere*, 18, pp. 2161–2176. doi.org/10.5194/tc-18-2161-2024.

Rao, Y., et al., 2022. HorNet: Efficient high-order spatial interactions with recursive gated convolutions. In: *Proc. NeurIPS*, pp. 10353–10366.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., Woo, W.-C., 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In: *Advances in Neural Information Processing Systems 28 (NeurIPS)*, pp. 802–810.

Tan, C., et al., 2023. Temporal attention unit: Towards efficient spatiotemporal predictive learning. In: *Proc. IEEE/CVF CVPR*, pp. 18770–18782.

Tan, C., et al., 2025. SimVPv2: Towards simple yet powerful spatiotemporal predictive learning. *IEEE Transactions on Multimedia*. doi.org/10.1109/TMM.2025.3543051.

Tolstikhin, I. O., et al., 2021. MLP-Mixer: An all-MLP architecture for vision. In: *Proc. NeurIPS*, pp. 24261–24272.

Trockman, A., Kolter, J. Z., 2022. Patches are all you need? In: *Proc. ICLR*.

Yu, W., et al., 2022. MetaFormer is actually what you need for vision. In: *Proc. IEEE/CVF CVPR*, pp. 10819–10829.