Review: Extended seasonal prediction of Antarctic sea ice concentration using ANTSIC-UNet

This is a well written paper exploring the development and application of a convolutional neural network (CNN), ANTSIC-UNet, for seasonal predictions of Antarctic sea ice concentration (SIC). The paper demonstrates how ANTSIC-UNet outperforms two benchmark models, as well as the SEAS5 numerical sea ice forecasting model. This paper also explores variable importance via the use of the explainable AI tool, permute and predict. I recommend this paper for publication subject to the major revisions outlined below.

Whilst the research presented here is of high quality, the abstract, introduction and discussion need to emphasise the novelty brought by this paper. This is currently not clear to the reader. The introduction highlights that fewer studies have predicted SIC in the Antarctic compared with the Arctic. Although the application of sea ice forecasts to the Antarctic provides some novelty, greater clarification of the methodological novelty provided by this study is also required. For example, previous studies have already applied CNNs for sea ice forecasting, undertaken analysis of feature importance and compared ML model outputs to SEAS5. This clarification of methodological novelty will make it easier for the reader to follow the paper.

Linked to this point, whilst it is true that far fewer papers have forecast Antarctic SIC, some key publications are missing. Please cite these and contextualise the findings of this paper to these manuscripts:

Dong, X., Yang, Q., Nie, Y., Zampieri, L., Wang, J., Liu, J. and Chen, D., 2024. Antarctic Sea Ice Prediction with A Convolutional Long Short-Term Memory Network. *Ocean Modelling*, p.102386.

Lin, Y., Yang, Q., Li, X., Dong, X., Luo, H., Nie, Y., Wang, J., Wang, Y. and Min, C., 2025. Ice-kNN-South: A lightweight machine learning model for Antarctic sea ice prediction. *Journal of Geophysical Research: Machine Learning and Computation*, *2*(1), p.e2024JH000433.

Wang, Y., Yuan, X., Ren, Y., Bushuk, M., Shu, Q., Li, C. and Li, X., 2023. Subseasonal prediction of regional Antarctic sea ice by a deep learning model. *Geophysical Research Letters*, *50*(17), p.e2023GL104347.

As a general point, I also believe this paper requires some restructuring. The comparison of ANTSIC-UNet to SEAS5 is not mentioned until the discussion section on line 395. The use of SEAS5 requires mentioning in the introduction, methods, and results. Further, it is not common for the discussion section to provide new results and figures. I suggest Figures 10, 11 and 12, alongside the supporting text and equations describing these results, are moved to the results section. This will allow the discussion section to focus on the relevance and contextualisation of the results, making the paper easier to follow for the reader.

Overall, the paper reads very well with very few typographical errors, I suggest these further minor corrections:

Line 77 – 79: please provide detail on the algorithm used to convert from passive microwave brightness temperatures to sea ice concentration values.

Section 2.2. Please justify the use of a CNN with UNet architecture. Some recent papers have shown generative models or other AI approaches to outperform UNets. Were other ML algorithms and architectures considered?

Line 81 – 84: “A linear least-squares trend was fit….” This information does not fit under the subsection 2.1, as these lines are describing a method applied to the passive microwave data, rather than the data itself. Please create a new subsection in the methods section on the benchmark models.

Line 95 - 105: Please refer to Table 1 when listing all the variables.

Line 110: Please make clear here or somewhere else the temporal resolution of the forecasts. Is it monthly, daily, seasonal or some other resolution?

Line 135: Please describe the hyperparameter selection and tuning process you employed.

Figure 2: Please provide some background in the introduction on how and why the Southern Ocean is split into these five regions.

Figure 2 caption: typo: “based on the same calenda**t** month”.

Figure 3 a) and e), please flip the colour ramp so white is ice and water is blue, or use a separate colour ramp altogether. Same for Figure 7.

Figure 3: Please make clear whether these are February and September means for a particular year or the whole date range.

Figure 4 f1 and f2: please make clear that A and B stand for Amundsen and Bellingshausen.

Discussion section: Due to the large similarities between this paper and the IceNet model published in Andersson et al. (2021), please provide detailed discussion on the relative performance of ANTSIC-UNet and IceNet.

Figure 6 line 262-263: “ANTSIC-UNet (anomaly persistent model) at different lead times up to 6 months….” How are different lead times represented in this figure?

Section 3.3: As a general discussion point, what is the suitability of using the anomaly persistence model as a benchmark model for forecasting extreme events? Isn’t it always destined to underrepresent these extremely anomalous cases? Are there more appropriate benchmark models that could be used for these circumstances?

Table 3 line 298: Typo “Here, **O**bserved..” change to lower case.

Figure 8: Help the reader- in which month when did the extreme event(s) occur.

Section 3.4. There are some points that are more appropriate for the discussion, particularly where references are made to other papers. For example from line 320 “previous studies…..”

Line 334: Typo: Circulation**.** (Raphael…) – remove full stop.

Discussion: Please contextualise your findings on feature importance for sea ice forecasting with the following paper that also carried out a similar study:
Uebbing, L., Joakimsen, H.L., Luppino, L.T., Martinsen, I., McDonald, A., Wickstrøm, K.K., Lefèvre, S., Salberg, A.B., Hosking, S. and Jenssen, R., 2025, January. Investigating the Impact of Feature Reduction for Deep Learning-based Seasonal Sea Ice Forecasting. In *Northern Lights Deep Learning Conference 2025*.

Discussion: Why does the performance differ between the different regions of the Southern Ocean? For example, the disparities between 4 b1 – f1. There is mention of this on lines 367 – 369, but please expand further. Also, please comment on the better predictive performance of the tool in the Austral summer.

There is no conclusion section. Please check if this is required.