Westbrook et al. (2025) propose a machine learning-based approach for classifying sea ice concentration using ship-board S-band radar imagery. This is a novel contribution with substantial scientific merit. The study is methodologically sound, particularly due to its use of multiple validation datasets, which significantly enhances its robustness.

However, I recommend **major revisions** before the manuscript is ready for publication. One of the key shortcomings is the **absence of a dedicated discussion section**. The current results section does not engage in analytical interpretation, and the conclusion is overly brief, leaving the manuscript feeling incomplete. More critical reflection on the implications, limitations, and broader context of the findings is necessary to fully realize the potential of this work.

Also, The quality of figures need to be improved before publication. The captions need to have more information.

Some general comments that need to be addressed:

Abstract: I feel the abstract needs to establish the uniqueness of the research as well as the ending sentence should put the applicability of this work in broader context.

Introduction:
The introduction is too short and needs more details to set the context and uniqueness of the study. Particularly, focused should be given on why classification of S-band radar and any past studies that has looked into this? Also, how is this helpful from a satellite remote sensing context, is this relevant for upcoming missions such as NISAR? Additionally, previous machine learning based studies that addresses sea ice classification needs to be addressed.

Method:
Section 2.1: Just curious, is there a reason by the original images were not used instead of screen captures? How does the resolutions compared? Could better results be expected if the original images were used?

Line 48/50: Is there a reference you can use where this approach was used before? I think the method of using RGB for classification used here needs to be backed by past studies or justified using patterns drawn from the images collected.

Section 2.2
Some basic information about the L4 data needs to be provided at the beginning to make it easier for the reader to follow along.

Figure 1, 2 The captions needs more details. The quality of figure needs to improve. There seems to be unnecessary white space. The size of the three images need to be made consistent. Please include what the white dots represent and is this relative to the ship in the center?

Line 64/65: Can you further explain what you mean by the leading edge effect here?

Line 85: Which issues are you referring to here? Be specific.

Line 90: Extra ‘a’ here..

Line 100: There needs to be a justification of why you are using this particular model for this classification task.

Line 117: Are there past studies that has made this observation about CNNs?

Line 130: What was the size of this ground truth dataset? How many images were used?

Figure 9: Needs improvements. Very difficult to undertsand this plot.

Table 1: Its not clear to me whether a part of the original dataset used for training and testing was used for validation or a separate dataset using the ground truth collocated photos were used to validate the model? Are these the results from the validation test using the collocated photos?

Section 3.3: This is actually a very interesting analysis. I feel like there could be further insights and interesting observations made from this which would further strengthen the study. I recommend adding more content to further justify the performance of the model compared to the salinity and temperature measurements.