We express our most sincere gratitude to the two anonymous referees who took the time to review our manuscript and provide edits as well as constructive feedback. We feel that all the recommendations acted to strengthen the manuscript and we were able to address and respond to each of the referees' comments. In the detailed response below we indicate how we modified the manuscript to address their comments and include relevant quoted revised text. The referees' comments are shown in *black italic text* and our response in blue with direct quotes from the revised manuscript indented.

RC1: 'Comment on egusphere-2025-643', Report #2 Submitted on 02 Oct 2025

The manuscript presents a method for classifying images from S-band ship radar into four categories based on the concentration of sea ice around a ship. The manuscript is well and logically written; it presents the data, method, and results in a transparent manner. A few minor comments listed below should be addressed to improve the text and figures.

However, the significance of the study is in question: do we really need a CNN for image classification in this case? Especially, in light of the conclusion that "the method is best suited for determining if there is ice or not". This question arises because the authors did not present the results from the simple, threshold-based classification, which, according to the authors, 'did not work' (Line 69). Unless the results from the more straightforward method are presented quantitatively (e.g., a confusion matrix, F1 score, etc) and compared to the CNN results, the manuscript cannot be recommended for publication.

We appreciate that we did not clarify our earlier methods to estimate sea ice conc from the S-Band images. We detailed our attempts at line XX of the revised manuscrtip.:

> "The classification of sea ice along the ship track of the SASSIE expedition was first attempted as a simple ratio of black to white pixels. Initial analysis revealed that this simple method was unstable and impacted by noise and changes in the radar settings. These issues became apparent when we investigated time series of pixel ratio sea ice concentration estimates and found that the variance was far greater than running mean estimates made at 1 and 5 minute intervals for almost all daily time series we studied. The simple pixel ratio technique was found to be unsuccessful for the following reasons: ."

Minor comments

Line 49. Although the details on how the images are geolocated are presented elsewhere (Drushka et al., 2024), it would be nice to go into details on how the geographic information is added to the images. In particular, the following questions arise, linked to the obtained results:

How latitude and longitude is assigned to image pixels (for further comparison with, e.g., sea ice concentration from grASIP)?

The images were gelocated using the ships GPS records. The image center is geolocated but due to the change in scaling of the radar image by the captain we cannot assign locations to each pixel. This is explained in line XX of the revised manuscript:

> Metadata include the spatial range of the image, which varied in size from 1.4 km to 22 km. The pixels outside of this range were removed from each image, and pixels within the range were georectified and stored as Geo-TIFF images. The images were geolocated using the ships GPS log in conjunction with heading and range information from the radar.

Is the slant range correction performed, and do the image pixels have equal area?

No further corrections were made beyond what was described in the manuscript.

If the image in Fig 1b is geolocated, where are the X- and Y- axes, scale, etc, typical for a geolocated map or a satellite image?

The image shown in Fig. 2 (we believe that is what was meant by the reviewer) is meant to show a final L3 image compared to the screen capture. As this image is meant to be an example of our methods we chose not to include axis labels or scales.

Line 69. 'This did not work' is not a very descriptive result. Please provide a quantitative assessment of a method more straightforward than the CNN to justify the application of the CNN. It can be presented in the form of a confusion matrix, F1-score, or a similar metric.

We have revised the text to further explain our choice to move to a CNN, which is now on line XX of the revised manuscript.

> The classification of sea ice along the ship track of the SASSIE expedition was first attempted as a simple ratio of black to white pixels. Initial analysis revealed that this simple method was unstable and impacted by noise and changes in the radar settings. These issues became apparent when we investigated time series of pixel ratio sea ice concentration estimates and found that the variance was far greater than running mean estimates made at 1 and 5 minute intervals for almost all daily time series we studied. The simple pixel ratio technique was found to be unsuccessful for the following reasons:

Line 78. 'the amount of ice in the floe' is always 100% because a floe represents a solid piece of ice floating in water. Needs rephrasing. Maybe "the amount of ice floes"?

We thank the reviewer for catching this error, the sentence in question has been revised as follows:

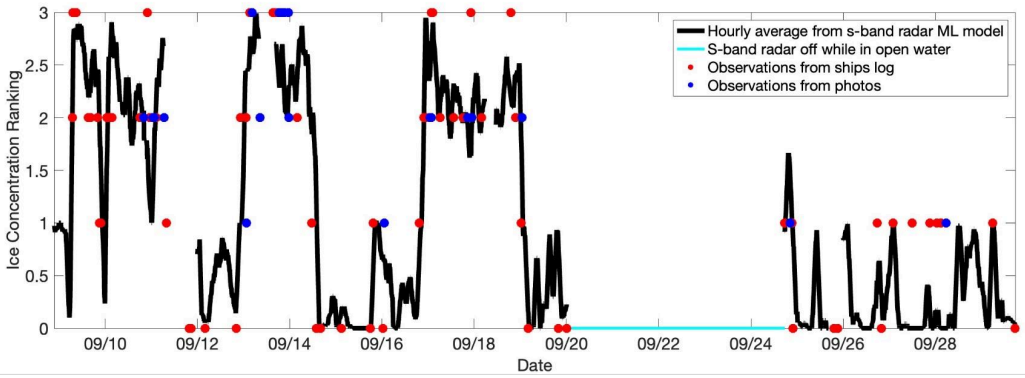
> When the ship approaches a patch of sea ice, the sea ice concentration is underestimated due to the radar's detection of the leading edge of the ice, that is, the ice floes that are closest to the ship.

Line 144. Figure 9 does not really reveal general agreement. Looking at the red dots, only 16 observations out of 28 seem to be close enough to the black line (60%). The blue dots are quite dark and are hard to identify on the plot near the black line, but it seems that there were only 7 photos for comparison, which match well. Please update the plot for easier identification of those points; compute and present the statistics (e.g., F1-score) using these independent observations, and provide a brief explanation for why there were only seven photos during a 20-day cruise.

We appreciate that the figure was hard to read in it's previous form. Please see our response below. As to statistics, we want to present the figure without any statistics because we do not believe that it's appropriate to report these number when the comparison is to qualitatively-labeled pictures and ship log entries and not any quantification of sea ice cover.

Line 147. On Fig 9 it looks like the log-book observations were rounded to the nearest noon or midnight, rather than nearest hour. Either the plots, or the text needs to be corrected.

We thank the reviewers for pointing out that the observations had been rounded to the nearest 12-hours. This has been corrected. Here is the new figure.



Line 154. The presented classification gives only four classes. How was the r2 score computed when comparing to the continuous SIC from grASIP? More details need to be provided here, as

well as a more detailed presentation of the comparison in the form of a scatter plot. This can shed more light on the developed model's applicability in various ice conditions.

We appreciate the comment and have modified the text to clarify our methods:

> To assess the validity of the model predictions of sea ice classification we compare our estimates to those of a derived daily gridded sea ice product generated from operational sea ice maps made by the National Weather Service Alaska Sea Ice Program (grASIP). The details of how the product is produced are given in \cite{pacini_2024}. Following the shiptrack, general agreement is found between sea ice concentration estimated from the model developed here and that of the grASIP (Fig. \ref{fig:asic}). Regression of our classification to the daily grASIP sea ice concentrations rounded to the nearest integer yields an $r^2=0.78$.

We opted not to include another figure as we think that the report of the correlation adequate summarized this comparison.

Line 180. The separation between C0-C1 and C2-C3 classes is not clearly seen in Fig 11. It looks like there is a strong peak in C1 (marked as L1 on the figure) in quite cold (0 °C) and fresh (24.5 PSU) waters. Please improve Fig. 11 either by tilting the view towards the reader or by providing four individual 2D histograms for each class in the T-S space.

We thank the reviewer for finding remnant mentions of our past L0-L4 classification. All figures have been updated to C0-C4. In regards to the reviewers request to change the “camera position” of this figure, we respectfully disagree.. We chose the position to highlight the mass of C3-C4 at the front of the image as this is the environment we expect to find high ice concentrations.

Figs 7 and 12: Classes C0 … C4 are shown as L0 … L1

Table 4: Should the values in this table be equal to the values on the confusion matrix presented in Fig. 7? Then the order of the values is wrong.

We are very grateful for the reviewer pointing out this error. In fact we decided that the info in the table was duplicative and we removed it.