The authors have addressed my previous comments. Further comments given below:

1. Although the authors did carry out a comparison between the NIC charts and the OSI-SAF charts, their results do not go into depth and are inconclusive. For example, it is stated that the accuracy of open water decreases because of the different ice concentration thresholds used for the OSI-SAF charts and for their classification method, but this is not shown. For example, it could be shown by retraining their classifier to use a different threshold, or by systematically comparing the misclassification rates for points in the NIC charts with ice concentration between 20% and 30% or 40% with misclassification rates for points with higher ice concentrations. On a similar note, for the high misclassification rate of first year ice, only one date is shown to demonstrate that the ice classes are different between the NIC charts and the OSI-SAF charts. It would be good to know how widespread these differences are.

2. page 10 lines 10-14 - here the authors compare their scores for the various classifier configurations for DS1 and DS2 (Tables 6 and 7). The fact that FC3 is overfitting for DS1, means that the accuracies for mFYI are very low (disregard the column for FC3 for Table 6), and there is considerable mFYI that is misclassified as old ice. It is not clear at all as to whether this problem (eg. 15% for mFYI for DS1) is due to the classifier itself or the data. Similar to the previous comment, it would be ideal to have a more thorough analysis of the ice charts themselves to unravel this problem, although I agree this could be left to a later study.

3. page 2 lines 25-24 - I believe the image analyses used in (Wang 2017) and those used in the present study (NIC) charts are similar in that both are manual analyses. The image analyses used in (Wang 2017) are each a direct manual interpretation of a SAR image by a trained ice analyst. They do represent analyses by 'certified ice analysts' (bottom of page 2). These image analyses are visually similar to the SAR data, as expected, and this is in agreement with the steps carried out in the submitted manuscript, where the authors visually check the agreement between the SAR imagery and the NIC charts. I think the only two differences are 1) the image analyses used in (Wang 2017) are not open to the public, and 2) they are an interpretation of the SAR image only. It's not clear from the manuscript what data sources the NIC charts are based on (though I may have missed it), and this information should be included.

4. I don't find the present manuscript demonstrates 'minimized manual work in the preparation of training and validation reference data' as the NIC charts are manual analyses, and the method itself also uses a manual step to select the training data.

- page 1 line 12 works should be work

- page 1 line 12 'reduce inconsistent decisions' - this was not shown. Please remove.

- page 2 line 20 and line 25 - what is the difference between a 'manual' interpretation and a 'raw' interpretation.