Reviewer 1

The clarity and logics of the manuscript has significantly improved during the revision. I have only few minor comments and can recommend the manuscript for publication after the corrections.

L210. Detailed steps … are …

Edited.

L304 - 316. This paragraph became quite controversial. The logarithmic fit can indeed be seen, but as you write below, the spread of the scatter plot is quite wide. I don't think it is correct to attribute this spread only to the difference between surface/bottom roughness seen by GEM-2 and the roughness articulated on the SAR image. The spread can also appear because the HH intensities (especially on X-band SAR) are controlled NOT only by geometric ice roughness. Therefore, separating FYI and MYI into different degrees of deformation cannot be justified globally. I'm afraid, that the results obtained in a relatively limited dataset can be misleading and should be presented with caution. I would suggest the following phrasing:

This indicates that TSX SC HH intensities in these particular transects are controlled largely by geometric ice roughness because other sea ice surface properties (micro-roughness, salinity, snow, etc.) are quite similar.

This relationship, …, illustrates that under similar environmental conditions FYI and MYI can be attributed to different degrees of deformation.

Edited.

Figure 11. Y-labels for the upper part of the plot of 'Class fractions' are missing.

Edited.

Reviewer 2

Review of the second version of the submitted manuscript; "*Sea ice classification of TerraSAR-X ScanSAR images for the MOSAiC expedition incorporating per-class incidence angle dependency of image texture".* The authors have made comprehensive revisions to the manuscript and included many of the previous suggestions from both reviewers as well as elaborated on key points requested in the previous manuscript version increasing the readability.

General Comments

The paper has been shortened by two pages, however, I must admit that I would still prefer if the manuscript could be trimmed further. For instance, Figure 2 and the associated text in L117-131 serve the purpose to argue for the need to split young ice into multiple classes.

However, the story regarding high winds that may have caused an opening in the sea ice, I see as less relevant to this narrative.

The sentence regarding the wind effect on sea ice breakup has been deleted, and this paragraph has been edited for better clearer and more concise, as follows. The whole text has been read through and edited again for conciseness.

Young ice shows a wide range of HH intensities due to differences in surface characteristics, which affects ice type classification. Fig. 2 shows an example of the progression of young ice on overlapping TSX SC and S1 EW scenes ~~, both in HH . High winds were observed during this period (Krumpen et al., 2021), which presumably contributed to the ice opening and deformation events.~~ in HH polarization. On 2019.11.20, wide-spread lead openings occurred around the CO. Between 2019.11.20 and 2019.11.21, more openings appeared which quickly re-froze into young ice. ~~For~~ On the TSX scenes, on 2019.11.21, most of the young ice areas appear very bright. ~~While the leads gradually close up~~Subsequently, young ice gradually darken to similar brightness to the surrounding ice. On the S1 ~~EW~~ scenes, HH intensities ~~for~~ of young ice gradually increase, from similar or lower brightness to nearby level ice to very bright on 2019.11.23 and 2019.11.24~~, until they again reach~~ . Afterwards, they again darken to similar brightness to the surroundings~~later~~. The changing young ice intensities ~~through time is~~ are presumably due to evolving surface roughness, e.g., influenced by the ~~growing and disappearing~~ formation and evolution of frost flowers which are highly saline with varying sizes, ~~and causes changing~~ leading to varying scales of surface

roughness ~~which strongly influences SAR signals~~ (Martin et al., 1995; Barber et al., 2014; Isleifson et al., 2018; Johansson et al., 2018). The delayed increase and decrease in young ice ~~intensities~~ backscatter in C-band (5.405 GHz) compared to X-band (9.65 GHz) ~~data~~ is then presumably due to different interactions between changing surface roughness scales and different SAR wavelengths (Isleifson et al., 2010; Dierking, 2010; Barber et al., 2014; Park et al., 2020). These observations confirm the need to split young ice into separate classes for ice type classification, which is described below.

I am happy that several of my concerns regarding the methodology are addressed in the revision, particularly increase in reference polygons from 10 to 15 per scene as well as ensuring that the spatial separation for testing was taken into account to give a more generalistic evaluation of the classifier. In addition, the inclusion of previous work and referencing highlighting the classifier as a established method is appreciated. Though personally, I would have preferred to label a much larger number of pixels but I will accept the approach in its given form.

There are still a few things, which I would like to have elaborated as well as a number of minor adjustments, e.g. grammatical suggestions and references in need of edits.

Major Comments

L13, there was a change in accuracy from 86.05% to 83.70%. Was this a result of the increase of the number of reference polygons? (L19-20 in the changes document).

This is a result of the overall change of workflow as suggested by both reviewers, i.e., changed number of reference polygons, the removal of speckle filtering, and the resulting change in the chosen combination of GLCM textures to use. Qualitatively this change of accuracy does not significantly change the conclusions

of this work or the quality of the product, but the workflow is now more theoretically sound.

Figure 7. Is this showing the distribution of classification accuracy for the different classes or the accuracy distribution for the test scenes? In case of the latter, what is the average accuracy based on, all pixels or average scene accuracy?

In addition, I think you should consider investigating the raw numerical performance of your classifier more. For instance, you could examine what incorrect classifications are labelled as. Are there patterns?

The latter, based on all the testing pixels of each scene.

An explanation of the misclassification patterns is now added to the text: 'For the final classification with MRF contextual smoothing, the confusion matrix (not shown) indicates that remaining misclassifications mostly happen between the difficult class pairs, as expected. Leads and level ice are mostly correctly classified.' We feel this clarification is sufficient to elaborate the classification performance without adding too much to the text.

In relation to the beginning of the 'Results and Discussion' section on page 13, I think it would be beneficial to the reader if a short summary of what you go through, e.g. "first we investigate the performance of the classifier.. Then we compare with sea ice roughness estimates.. After which we present the time-series of ice class fractions.. And then we present the limitations and future steps...

This is added to the text: 'In this section, we first present qualitative and quantitative evaluation of the performance of our classification product. Then, we compare the classification maps with sea ice roughness estimates from MOSAiC in-situ data, and accordingly evaluate our classification scheme splitting FYI and MYI into different deformation states. To evaluate the consistency of the classification, temporal development of areal fractions of each class is then presented and compared with indicators of ice openings from in-situ data and other MOSAiC studies. Finally, we list several limitations of our workflow and give potential directions for future studies following this work.'

Sentences with similar information in the preceding paragraph has accordingly been deleted.

Minor Comments

Table 2. The GLCM textures are defined but variables are not described. I think you should consider adding the names of variables for instance in the table caption. E.g. i, j, $\boldsymbol{\mu}$, P, HXY, etc..

Added to the text.

L373: You have not explained what the abbreviation of CNN is.There are also some works that use CNNs to classify sea ice types. Perhaps it could be useful for you to investigate.

E.g. Boulze, H., Korosov, A. & Brajard, J. Classification of Sea Ice Types in Sentinel-1 SAR Data Using Convolutional Neural Networks. *Remote Sensing* **12**, 2165 (2020).

(though there are probably more recent developments)

Full name and citation added.

Additional suggestions for grammatical adjustments can be found in the attached PDF.

Edited according to the comments.

References:

L422:: the DOI link does not work

Edited.

L 427: the link contains two https://doi.org.. This should be fixed.

Edited.

L32: the DOI link does not work, though I believe it is due to the .ch13. Perhaps this should be written instead of included in the DOI?

Edited.

L444: please fix the DOI link

Edited.

L454: the scihub link does not work

Edited.

L560: the link contains two https://doi.org.. This should be fixed.

Edited.