This manuscript primarily describes the release of a deep learning focused glacial feature tracking benchmark dataset on multi-mission SAR data. Additionally, it describes a reference deep learning implementation that performs automatic glacial termini extraction/land classification on the benchmark dataset, and provides tools for further use/standardization for easier comparison across different deep learning methodologies. The literature review is robust, and covers existing work done in the field regarding glacial feature datasets and automated extraction algorithms. This study uses data sources with high spatio-temporal and satellite variety for its benchmark dataset, exemplifying robustness and rigor in the final data product. Specifically, the dataset consists of 7 diverse tidewater glaciers from Greenland, Alaska, and the Antarctic Peninsula, from 1985-2020, across 6 separate SAR satellite families. The dataset itself consists of 681 total image pairs and accompanying metadata, which is a significant contribution to the field. This will encourage the development of more accurate and robust deep learning models for glacial feature extraction and land classification of cryospheric regions. Furthermore, the reference deep learning methodology provides a baseline for easy cross-comparison across future studies in glacial feature extraction. Robust pre- and post-processing techniques are employed to produce usable data products from the deep learning classification module. Accuracy is determined using the Mean Distance Error and Intersection over Union metrics that are commonly used by other studies in the field. While the deep learning model on the dataset testing set shows room for improvement, it fulfills the goals of this study to provide benchmarks for future models.

Overall, this study provides contributions to the field of deep learning as applied to automated glacial feature extraction and classification. Ultimately, this data will help encourage further development in the field, allow for increases in accuracy in derived data products, and ultimate enable better projections/modeling of sea level change and in AOGCMs. The overall manuscript is well done, free of major grammatical errors, with only
minor remarks (see below and specific comments) to be made and addressed before acceptance at the editor’s discretion.

**Major Comments:**

- It would be helpful to include the type of data covered by each study in Table 1, such what types of glacial features are provided by each dataset. These may include (but are not limited to) one or more of the following: full-line delineation (ESA), a centerline position (King), or glacial outline (GLIMS).
- Since the primary data of interest is the training/testing benchmark dataset, it would be beneficial to further emphasize this in some way throughout the manuscript, such as including the total number of image pairs/metadata in the abstract. This will enable readers to see the primary contribution more easily, and enhance visibility/usage of this work within the field.

**Specific Comments:**

**P5 L108:** "Lansat“ -> “Landsat”

**P19 L364:** “So there is a trade-off between patch size and batch size” Rephrase.

**P24 L463:** “are“ -> “is”

**P32 L539:** “25rd“ -> “25th“