Towards representation learning of radar altimeter waveforms for sea ice surface classification

Lena Happ[1,3], Sonali Patil[2], Stefan Hendricks[1], Riccardo Fellegara[2], Lars Kaleschke[1], Andreas Gerndt[2,3]

[1] Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar- und Meeresforschung, Bremerhaven, Germany
- (Lena.Happ, Stefan.Hendricks, Lars.Kaleschke)@awi.de
[2] German Aerospace Center (DLR), Institute of Software Technology, Braunschweig, Germany
- (Sonali.Patil, Riccardo.Fellegara, Andreas.Gerndt)@dlr.de
[3] University of Bremen, Center for Industrial Mathematics, Bibliothekstraße 5, Bremen, Germany

Keywords: Satellite Radar Altimeter Data, Sentinel3/SRAL, Sea Ice Classification, self-supervised learning, variational auto-encoder, representation learning

Abstract

Satellite radar altimeters provide crucial insights into polar oceans and their sea ice cover, enabling the estimation of sea level, sea ice freeboard, and thickness. These retrieval algorithms depend on accurate discrimination between radar altimeter waveforms from sea ice and ocean surfaces in heterogeneous and dynamic surface conditions. A further and less mature step is classifying different sea ice types in addition to the ice/ocean discrimination. We aim to develop new methods for a novel multi-category sea ice and ocean surface classification directly from satellite radar altimeter data to improve sea ice climate data records. Traditional waveform representations are limited to a small set of parameters, leading to information loss. Moreover, machine learning models for sea ice classification often depend on supervised training, which is vulnerable to uncertainties in labeled data, especially in polar regions. To address these limitations, we explore self-supervised learning methods to optimize waveform representations, which can capture more detailed information for a classification with finer granularity. Furthermore, they do not require labeled data, which is not available at the spatial coverage and resolution of radar altimeter waveforms. We apply these techniques to SRAL data from the Sentinel-3 mission. We show that the information preserved in the latent space of an auto-encoder enhances the feature space of traditional waveform parameters, improving the subsequent classification process, when comparing our results to available sea ice charts and other remote sensing products. Our results demonstrate better generalization compared to supervised approaches.

1. Introduction

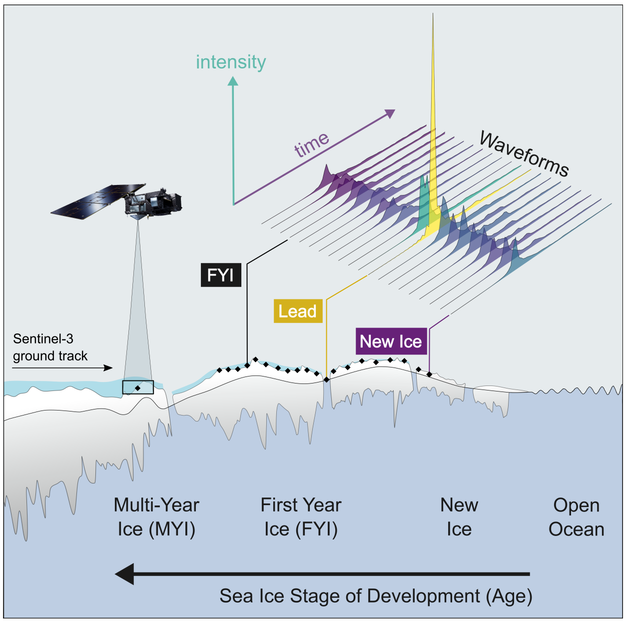


Figure 1. In the Arctic, satellite radar altimetry is used to estimate sea ice thickness. For each footprint a single intensity curve (called waveform) is generated. These contain valuable information about the sea ice type and snow cover (depicted in light blue) within the footprint.

In the sea ice domain, which consists of frozen seawater floating on the ocean surface, satellite radar altimeter data is used to estimate sea ice thickness and volume. To estimate the sea surface height along the ground track of the satellite, reference points of uncovered sea surface are required. This introduces the need to distinguish sea ice from open ocean and fractures in the ice, called leads. The sea surface height is then substracted from the ice surface heights yielding the fraction of the ice columns above local sea level (freeboard).

Sea ice thickness is derived from freeboard using the buoyancy of the ice and its snow load, which depends on the density of the sea ice and water as well as on the snow mass. Uncertainties of these parameters are critical for meeting the target uncertainty for sea ice thickness of climate data records.However, they are not routinely measured and information in retrieval chains rely on external data sources, climatologies or parametrizations.

Sea ice density and snow load develop with the life cycle of sea ice from formation to melt, which can be categorized as stage of development. It is therefore desirable to classify radar altimeter waveforms by sea ice type to improve the parametrization of sea ice density and snow load. Additionally, due to the ice drift causing substantial displacements of the ice on small timescales, it is desirable to retrieve information about different ocean surface types (including open ocean, leads, and different stages of sea ice development) directly from the same sensor used for the sea ice thickness estimates, i.e., the radar altimeter data (Quartly et al., 2019).

Improving sea ice thickness measurements is essential because sea ice conditions in the polar regions significantly influence the global climate system and data about its declining thickness is an important input to climate models. However, current radar altimetry-based sea ice thickness products do not yet meet the

https://doi.org/10.5194/isprs-annals-X-G-2025-333-2025

target accuracy requirements (World Meteorological Organization, 2022), and we consider sea ice type classification as an avenue towards this goal.

Radar altimetry, unlike most other Earth observation methods, is not an imaging technique. It produces single intensity curves for footprints of a narrow swath along the trajectory of the satellite. These signals are called waveforms and it is usually desired to extract information for each of the corresponding footprints. This is depicted in Figure 1.

Existing approaches (Section 3) do not use the radar altimeter data as direct input to their classification models. Instead, a few parameters are extracted from the waveforms that describe its shape. Only these extracted parameters are used for the classification process.

For thresholding methods, this representation by waveform parameters is a natural choice, since thresholds for these parameters are historically found empirically to distinguish between leads and sea ice. However, representing waveforms with only a few parameters may result in the loss of additional information inherent in the waveform signal. Although the increase of computational power and the development of more sophisticated machine learning methods now enable the use of more complex representations as input, the practice of using those waveform parameters continued with the exploration of machine learning methods.

On the other hand, using the whole waveform as input for classification would correspond to an unnecessarily high feature space and lead to computational inefficiencies, potentially causing over-fitting due to the curse of dimensionality. Finding an optimized representation for the classification process might also facilitate the adaptation of the complete classification pipeline to other satellite radar altimeter missions.

However, radar altimeter data differs significantly from other data sources for which extensive machine learning techniques have been developed. Unlike image-producing sensors, radar altimeters generate intensity curves that are distinct from typical time series signals, as they are very short and of lower frequency. Therefore, techniques from these two research fields cannot be directly applied, requiring adaptations of machine learning methods for this specific type of data.

Additionally, the existing literature on sea ice surface type classification primarily focuses on supervised classification methods, where the adjustment of internal parameters requires a labeled dataset. In the polar regions, this is challenging because these labels are derived from other remote sensing data (including synthetic aperture radar, passive microwave, visual, as well as thermal-infrared data) that come with inherent uncertainty. The differences of spatial resolution between data sources and collocation over drifting sea ice introduces additional uncertainties. This uncertainty is difficult to quantify due to the scarcity of available ground truth data, resulting from limited access to polar regions. Consequently, supervised learning is highly susceptible to the quality and reliability of its training data.

As an alternative approach, we explore self-supervised learning techniques to learn a representation of the radar altimeter waveforms that can be used for classifying polar ocean surfaces. Specifically, we train an auto-encoder and a variational auto-encoder to extract meaningful waveform representations and evaluate their suitability for classification. Our results show that the learned representations retain additional information to that contained in traditional waveform parameter representations and improve surface classification. This confirms that a learning objective focused on signal reconstruction can capture relevant information but cannot yet completely substitute traditional approaches. Additional loss functions and the incorporation of class labels in the training have to be investigated. We verify that our network architecture effectively extracts the necessary information directly from the waveforms by training a neural network classifier to distinguish between different polar ocean surfaces based on the radar altimeter waveforms. To the best of our knowledge, this is the first study to apply self-supervised representation learning to radar altimeter waveforms, providing new insights into representation learning for polar ocean surface classification.

2. Background

In the following sub-section 2.1, we give a general introduction to satellite radar altimetry, the kind of data it produces, and how this is traditionally represented by waveform parameters. We also provide the background on operational ice charts in subsection 2.2, which is the main source of class labels not only in this study but also in the existing literature.

2.1 Satellite Radar Altimeter Data

There are several satellite radar altimeter missions suitable for polar ocean observation. They span approximately the last 30 years with overlapping intervals. In this study, we are using data from the most reasoned missions. These are the Sentinel-3A and -3B missions of the same satellite series equipped with identical Sentinel-3 Radar Altimeters (SRAL) (ESA, 2022). Our long-term goal is to derive methods that are easily adaptable to older missions. For this reason, the following introduction is kept general. Satellite radar altimeter sensors are active sensors that are usually used to measure surface elevation. They do not produce image data, instead only single measurements for footprints (of a size of approximately 300 x 1500m for SRAL) along the satellites ground trajectory are produced. The sensor continuously sends radar pulses which get reflected at the surface and are received at the altimeter. Each of these single looks describe the received echo power as a function of time. Multiple looks are combined to form a synthetic aperture to increase the along-track ground resolution. Each so called multi-looked waveform is not a single intensity peak associated to one single timestamp. It is dispersed over a wider time frame of a few micro- to milliseconds (Wingham et al., 2006). This echo is sampled in a limited number of time bins (256 bins in case of the Sentinel-3 missions) called waveforms. The shape of the measured waveform depends on the properties of the altimeter sensor and the applied synthetic aperture, but also on the properties of the reflecting surface, specifically the roughness, backscatter properties from the ice and snow layers as well as the backscatter incidence angle dependence. For example, leads represent a very flat surface with specular reflection yielding a very narrow waveform. Older sea ice surfaces are rougher, with diffuse backscatter properties and higher height distribution thus yielding broader waveforms (see Figure 5b). This discrepancy is exploited for surface type classification. Additionally, sea ice surfaces evolve through the stage of development by dynamic and thermodynamic processes which are then reflected in waveform properties. This allows for a more finer distinction of surface types based on waveforms shapes.

2.1.1 Waveform Parameters In the existing literature a classification of surface types is not based directly on the altimeter waveforms. Instead, a few parameters are derived from the signal and used for the classification. These waveform parameters describe the shape of the waveforms. Below, we list a few waveform parameters used in this study:

Normalized Backscatter Coefficient (σ_0). The radar backscatter coefficient quantifies the radar signal's return strength from the Earth's surface. It is defined as the ratio of the power scattered back to the radar from a unit area of the surface to the power incident on that area. It provides insights into surface characteristics, especially roughness. Its calculation depends on a lot of internal parameters of the sensor (Dinardo, 2016).

Pulse Peakiness (PP). PP measures the peakiness of the waveforms. It is found by dividing the maximum power P_{max} by the total accumulated power of the waveform, which is the sum of the power P_i measured for each bin i (Wernecke and Kaleschke, 2015).

$$PP = \frac{P_{max}}{\sum_{i=1}^{n} P_i} \tag{1}$$

Leading Edge Width (LEW). LEW measures the length of the signal before its maximum peak. Specifically it counts the number of bins from the onset of intensity increase to the peak maximum. It is calculated using a smoothed curve fitted to the waveform (Wingham et al., 2006).

Late-Tail-to-Peak-Power ratio (LT2PP). LT2PP parameter measures the off-nadir power in the tail of the waveform (Rinne and Similä, 2016). It is calculated as the ratio of the accumulated echo power between the 50th and 70th bin after the signal's maximum (max) and the maximum power.

$$LT2PP = \frac{\frac{1}{21}\sum_{i=max+50}^{max+70} P_i}{P_{max}} \tag{2}$$

2.2 Operational Ice Charts

We derive class labels from the operational ice charts (OIC) of the U.S. National Ice Center (USNIC) (U.S. National Ice Center, 2022). These charts are manually created on a weekly or biweekly basis by experts using a combination of satellite images and observational data accumulated over the previous three days.

The ice charts define polygons with a homogeneous distribution of different stages of ice development. For labeling, we summarize these stages into the overall categories New Ice, First Year Ice (FYI) and Multi Year Ice (MYI). For each polygon, the total sea ice concentration is provided as a percentage, along with the concentration of the three thickest ice types present in the polygon. Therefore, they do not provide exact class labels but rather a mix of different ice types. They also do not include information about leads, so these labels must be extracted from a different source.

3. Related Work

The existing literature on detecting of leads using radar altimeter data can be categorized into two main approaches: thresholding methods (Wernecke and Kaleschke, 2015; Laxon et al., 2013; Passaro et al., 2018) and methods using machine learning techniques (Müller et al., 2017; Poisson et al., 2018; Longépé et al., 2019; Bij de Vaate et al., 2022; Dawson et al., 2022).

The classification of different ice types is focused on distinguishing between FYI and MYI. As for lead detection, they all use traditional waveform parameters to represent altimeter waveforms. Zygmuntowska et al. (2013) classified leads, FYI and MYI using a Bayesian approach. Rinne and Similä (2016) used a nearest neighbor approach to classify open ocean, thin FYI ($<$ 70 cm), thick FYI ($>$ 70 cm) and MYI.

While the focus of Zygmuntowska et al. (2013) is on an improvement of sea ice thickness and mass estimates, but using an airborn radar altimeter, ASIRAS, the study of Rinne and Similä (2016) aims at an automated ice charting to support the operational ice charting for navigation.

Shen et al. (2017) and Shu et al. (2020) classify ice types based on a Random Forest approach, but Shu et al. (2020) average features over segmented patches of homogeneous waveform information (Object-based Random Forest). All the above studies use operational ice charts as a label source. Fredensborg Hansen et al. (2021) on the other hand use labels based on sea ice type charts from Ocean and Sea Ice Satellite Application Facility (OSI SAF) and compared four different classification methods: threshold-based, Bayesian, Random Forest and k-nearest neighbour (kNN). Aldenhoff et al. (2019) do not derive a method for automated classification. Rather, they are investigating possibilities to combine SAR imagery and altimeter data. They compare the distribution of different waveform parameters for FYI and MYI, manually labeled based on SAR images. For a few representative scenes collocated with radar altimeter data, they have analyzed the concordance of the two data sources.

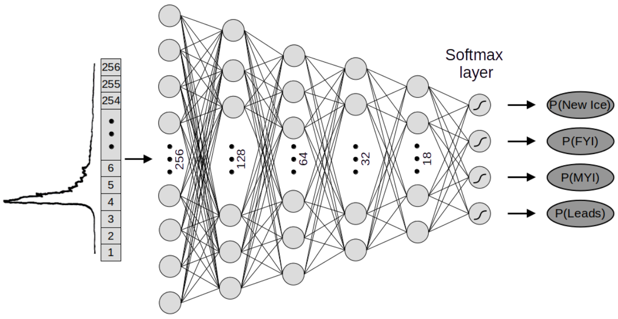
4. Methods

We explore different self-supervised learning methods to derive representations of the radar altimeter waveforms that can later be used to classify different sea ice surface types. To evaluate the quality of these waveform representations, we measure the inter-class and intra-class variability between polar ocean surface classes. This is done by applying the kNN algorithm to the feature space under analysis, which could be either the representation derived from a self-supervised method or its combination with traditional waveform parameters.

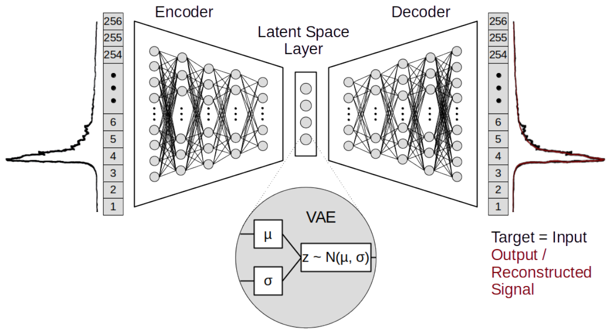
Since the kNN is sensitive to class imbalance, we use a labeled test dataset with a uniform distribution of surface classes. Class labels are assigned using a leave-one-out method, where each test sample is classified based on the majority vote of the $k = 1$ closest samples in the test set (excluding the sample in question). The proximity between samples is determined by Euclidean distance in the feature space. We use the kNN classification accuracy as an objective measure of the quality of waveform representations for surface classification. As a baseline, we use the four waveform parameters described in section 2.1.1 to represent waveforms.

Although our primary focus is on self-supervised methods, we first develop a simple multilayer perceptron (MLP) to classify the four different surface classes: New Ice, FYI, MYI and leads, as described in subsection 4.1. This step allows us to assess whether this architecture can effectively extract relevant information for surface classification from the radar altimeter waveforms.

Subsequently, we will train an auto-encoder (sub-section 4.2) and a variational auto-encoder (sub-section 4.3) with the same MLP architecture in the encoder. The latent space of the trained models will then be evaluated against our baseline dataset using kNN accuracy.



(a) MLP classifier with the waveform as input.



(b) AE and the VAE

Figure 2. Shown are sketches of the used neural network models.

4.1 Multilayer Perceptron Classifier

A multilayer perceptron (MLP) is a simple feed forward neural network (Hastie et al., 2009). We first train a MLP with a simple architecture in a supervised manner, using an One-Hot Encoding for the labels from the OIC as targets and the waveforms as inputs. Although this study focuses on self-supervised learning, this step verifies whether the chosen architecture can extract relevant information from the normalized waveforms. Additionally, we use the four selected waveform parameters as input to compare the results.
In both cases, we use a similar network architecture consisting of five fully connected linear layers, each followed by a rectified linear unit (*ReLu*) as activation function to introduce nonlinearity. The only difference is in the input layer size, which reflects the dimensions of the input vectors. In the first case, the linear layers are of sizes 256-128-64-32-18-4, where 256 corresponds to the number of bins in the waveform. In the second case, the input layer contains only four neurons, corresponding to the four waveform parameters. In the final layer, we are applying the *LogSoftmax* activation function to output the probabilities for each class. The architecture is shown in Figure 2a.
For training, we use the negative log-likelihood loss function, with class weights adjusted according to the number of samples in each class to address the imbalanced class distribution. Based on this, the model's learnable parameters are updated using the Adam optimizer, which combines momentum and an adaptive learning rate and therefore tends to converge faster compared to standard Stochastic Gradient Descent (Kingma and Lei Ba, 2014). The initial learning rate is $\lambda = 0.0001$.

4.2 Auto-Encoder

An auto-encoder (AE) is a neural network trained in a self-supervised manner, that can be used to learn a low-dimensional representation of the input data. Its architecture consists of two parts: (i) an encoder, which reduces the input signal to a minimal form (defined by the size of the last layer in the encoder, called *latent space*), and (ii) a decoder, which reconstructs the compressed signal back to an output with the same dimensions as the input data. The network is trained by applying a loss function that compares the input with the reconstructed signal, to preserve as much information as possible in the latent space (Bank et al., 2023).
We want to use the AE's latent space as a new representation of waveforms, with the normalized waveform as the input. The loss function used is the mean square error (MSE) between the input and the reconstructed waveforms, meaning no surface type information is required for training.
The encoder has a similar architecture to the MLP classifier, except it lacks the *LogSoftmax* activation and has a smaller last layer. We test different latent space sizes, ranging from two to ten neurons. The decoder mirrors the encoder's architecture, with layer sizes of 18-36-64-128-256. The AE is trained using the Adam optimizer, with a learning rate of $\lambda = 0.0001$.

4.3 Variational Auto-Encoder

A variational auto-encoder (VAE) is a type of generative model that extends the concept of a regular AE (Kingma and Welling, 2014). While in a standard AE, the encoder maps the input data to a deterministic latent vector, in a VAE, each latent variable is represented as a probability distribution. Specifically, the encoder outputs the mean μ and standard deviation σ of a Gaussian distribution for each latent variable. Samples are drawn from this probability distribution, which are then decoded. As a result, outputs of the VAE that are similar in terms to the reconstruction loss are also close to each other in the latent space. This property is not necessarily present in a standard AE.
The loss function of a VAE has two components. One is the reconstruction loss, as in a regular AE. The other is a regularization term that forces the latent space distribution to approximate a Gaussian distribution. This is achieved by applying the Kullback–Leibler (KL) divergence, a measure of the difference between two probability distributions (Kullback and Leibler, 1951). Figure 2b shows the concept of a VAE.
The VAE's latent space could potentially have an advantage for the classification of sea ice surfaces, as waveforms of the same surface type are likely to be more similar compared to those from different classes. The key architectural difference from the regular AE is the division of the latent space into its mean and variance components, with the KL-divergence added to the loss function.

5. Data

We use radar altimeter data from the SRAL sensor aboard the Sentinel-3A and -3B missions, as provided in the level 2 data from PySiral (Hendricks et al., 2024). This dataset also includes labels for leads, resulting from a thresholding approach. For dates on which operational ice charts are published by the USNIC, we retrieve ice class labels for footprints located within polygons where the sea ice concentration for a single stage of development is at least 90%. We labeled data for an entire season (October 2022 to April 2023) and distributed it across training, validation, and testing datasets. We summarize sub-classes into the four main categories: New Ice, FYI, MYI and Leads, with the distribution of these classes shown in Table 1. For self-supervised learning, no labels are needed for training. Thus, we augmented the training dataset with radar altimeter data of the day before and after the dates on which OIC data is released

Class Name	Training	Validation	Test
New Ice	3,561	4,121	729
FYI	163,319	66,228	93,216
MYI	253,665	92,303	109,468
Leads	126,114	49,688	12,029
Total	**546,659**	**212,340**	**214,835**

Table 1. Distribution of different surface classes in the training, validation, and test sets.

(independent of the dates assignment to training, validation and test set). This additional data improves Arctic coverage, as it is not limited to footprints within polygons with a high, single sea ice concentration. As a result, the number of training samples increases by 2,842,088 footprints. The training dataset is used for updating the model parameters based on the loss functions, while the validation dataset helps select the optimal model by epoch. The test set is reserved only for evaluating the selected models. Since we require a balance of class labels for the kNN algorithm, and the new ice class only contains 729 samples in the test set, we define a class-balanced test set of 700 randomly sampled footprints for each surface type.

5.1 Uncertainties in class labels

One challenge when working with remote sensing data in the Arctic is validating data products, and the OIC data hold an intrinsic inaccuracy. Additionally, there is uncertainty in the class labels due to inhomogeneities in the ice chart polygons. To estimate the uncertainties in these class labels, we compare FYI and MYI class labels derived from the operational ice charts with those from the OSI SAF Global Sea Ice Type Classification (Copernicus Climate Change Service, 2020), which also contains its own uncertainties. The comparison yields an accuracy of 0.87 in label agreement between OIC and OSI SAF in the test set, highlighting that part of the evaluated methods' inaccuracy comes from label inconsistencies in the test data.
To reduce this uncertainty, we select FYI and MYI samples with matching labels from both OIC and OSI SAF for the balanced test set, which is used for comparing different waveform representations. This dataset is referred to as the OIC + OSI SAF test set. To demonstrate the impact of data distribution shifts on supervised learning, we also validate the MLP classifier using a balanced test set based solely on the OIC labels (without the comparison to OSI SAF data). This is referred to as the OIC test set.

5.2 Waveform pre-processing

Before feeding the altimeter waveforms to the neural network models, we perform two pre-processing steps:

- First, we align the waveforms so that their maximum power is consistently at the 100$^{\text{th}}$ bin. If the signal is shifted to the right, the last few bins are truncated, and the leading ones are padded with zeros, and vice versa for leftward shifts, ensuring a fixed signal length of 256 bins. This alignment is important as the waveform's position within the recorded time frame does not contain information about the surface type. The time frame in which the signal is recorded depends on the surface elevation and is constantly adjusted.

- The second pre-processing step is normalization. Since neural networks usually require the normalized inputs, we normalize the waveforms by dividing each value by its maximum. This ensures all waveform values range between zero and one, though we lose the information about the absolute heights in the process.

6. Experimental Evaluation

In this section, we first present the baseline (section 6.1) against which we compare our results. To demonstrate that the chosen architecture is capable of extracting relevant information for classification, we train an MLP classifier (section 6.2). Finally, we extract waveform representations from both auto-encoder (AE) and a variational auto-encoder (VAE) (section 6.3).

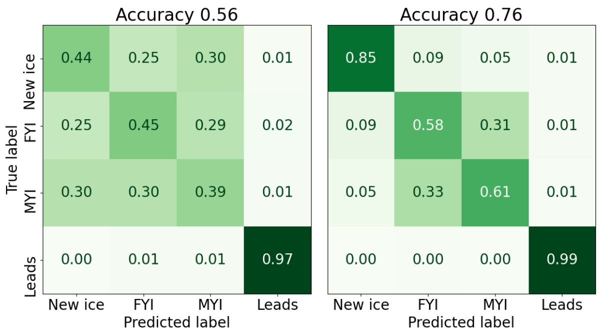
6.1 Baseline

As a baseline for a waveform representation, we use the traditional waveform parameters described in section 2.1.1. We select the three waveform parameters PP, LEW, and LT2PP because they are the ones used by Rinne and Similä (2016). While Rinne and Similä (2016) additionally uses the parameter stack standard deviation (SSD), this parameter is not available for older satellite radar altimeter missions. Since we aim for our methods to be transferable to these older missions, we omit this parameter.
Instead, we include the normalized backscatter σ_0, which has been shown to capture information about the surface roughness (Quartly et al., 2019). In Figure 3a, we present the confusion matrix from the kNN algorithm based on these waveform representations. The matrix on the left uses only the three waveform parameters PP, LEW and LT2PP, while the one on the right additionally includes σ_0.
It is evident that in both cases leads can be distinguished from sea ice. However, the classification of different ice classes improves significantly with the inclusion of σ_0. The overall accuracy increases from 0.56 to 0.76 when σ_0 is added to the feature space. This improvement is particularly noticeable in the classification of new ice, where the information provided by σ_0 plays a crucial role. Additionally, the distinction between FYI and MYI, which are otherwise difficult to separate, also improves.
This increase in accuracy by adding σ_0 clearly demonstrates that this parameter contains important information about ice classes that is not captured by the other three waveform parameters.

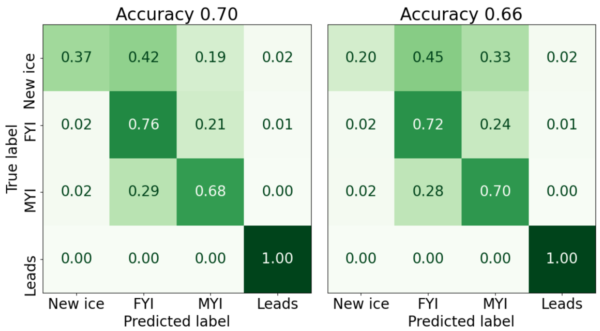
6.2 MLP Classifier

The MLP classifier is trained on the labeled training dataset, which is based only on the labels of the OIC data. To illustrate how inaccuracies in class labels or shifts in the data distribution between training and testing sets can impact the model's performance, we evaluate the MLP classifier using both the OIC test set and the OIC + OSI SAF test set.

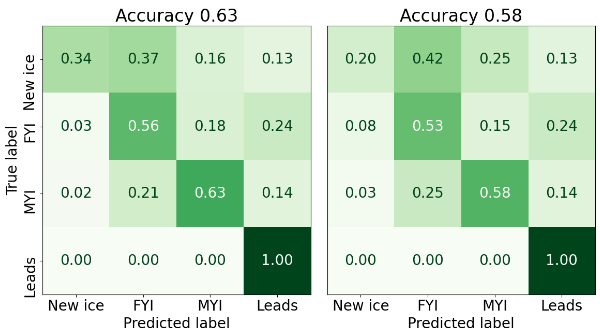
6.2.1 Evaluation using the OIC test set We first evaluate the model using the OIC test set, which is constructed similarly to the training dataset. Figure 3b shows the confusion matrix for the MLP classifier with the pre-processed waveforms as input (left) and with the four waveform parameters as input (right). The accuracy for the waveform inputs (0.7) is significantly higher than for the parameter inputs (0.66). While the ability to classify FYI, MYI, and leads improves compared to the baseline, the classification of new ice is notably lower, often misclassified as FYI or MYI. Although the overall accuracy



(a) kNN classification using parameter baseline with PP, LEW, LT2PP (left) and including σ_0 (right)



(b) MLP classifier for OIC test set with waveforms (left) and waveform parameters (right) as inputs



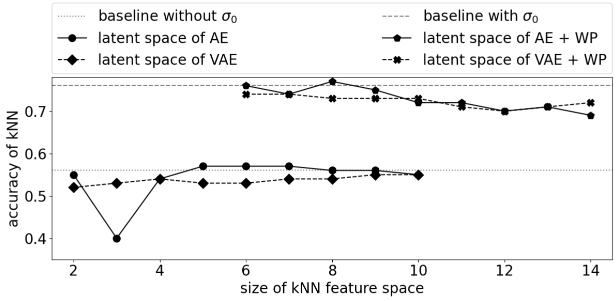
(c) MLP classifier for OIC + OSI SAF test set with waveforms (left) and waveform parameters (right) as inputs

Figure 3. The confusion matrices are displayed using the fractions of samples in the different categories.

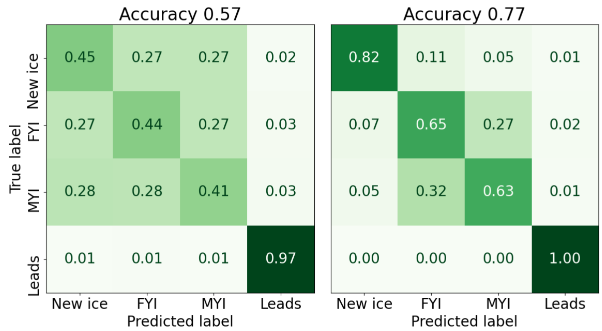
is lower than that of the baseline, the classification performance for FYI and MYI improves, especially when using the full waveform as input.

The difficulty in classifying new ice is likely due to the limited number of samples in the training dataset. However, the results show that the chosen architecture is generally effective in extracting relevant information from the waveform input. Furthermore, using the full waveform as input to the neural network model is superior to using only the waveform parameters, as it contains more information for classification.

6.2.2 Evaluation using OIC + OSI SAF test set When evaluating the MLP classifiers using the OIC + OSI SFA test set (Sub-Figure 3c), we observe similar trends as with the OIC test set. However, the accuracies drop to 0.63 for the full waveform input and to 0.58 for the waveform parameter input. This



(a) accuracy vs. size of kNN feature space (consisting of the waveform parameters (WP) in combination with different sizes of the latent space of the AE or VAE



(b) confusion matrix of the kNN for the six dimensional latent space of the AE (left) and the combined feature space of latent space size four together with the waveform parameters (right) given as fractions

Figure 4. The confusion matrices for the feature spaces with highest accuracies are displayed.

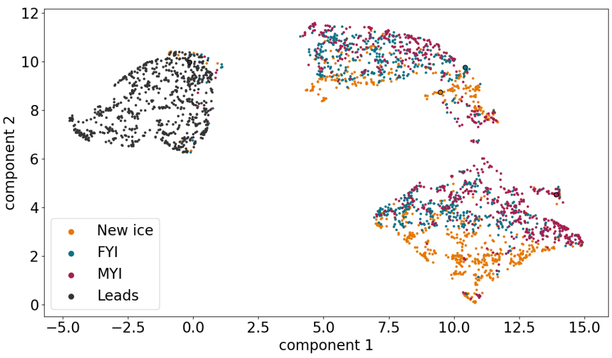
demonstrates the MLP classifier's limited ability to generalize to data with a different underlying distribution. This finding further motivates the exploration of self-supervised learning. Although these methods also struggle in generalizing to unseen data, they do not rely on potentially incorrect labels, offering a key advantage.

6.3 (Variational) auto-encoder

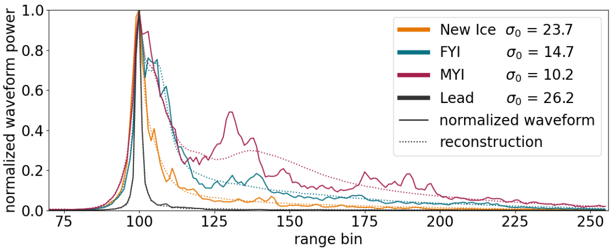
Figure 4a shows the accuracy of the kNN algorithm based on the latent space of both the AE and VAE for different dimensions of the latent space, which correspond to the feature space of the kNN. The accuracy increases as the latent space expands, peaking at 0.57 for dimensions five, six, and seven, but then declines to 0.55 for larger latent spaces.

The plot also shows the accuracy for a feature space combining the latent space of the AE with the four normalized waveform parameters. In this case, the accuracy is highest for smaller latent space dimensions and decreases with larger latent spaces. Comparing this with the baseline, the combined feature space performs better when the the latent space is of size four (leading to a feature space of size eight) in the case of the standard AE. For the VAE, while accuracies are consistently lower than those of the AE, they follow a similar trend.

The confusion matrices for the AE's latent space of size six and the combined feature space (latent space of size four with waveform parameters, resulting in a feature space of size eight) are shown in Figure 4b. In both cases, leads are well-separated from other ice classes, but the separation between different ice classes, especially FYI and MYI, remains challenging, similar to the baseline without σ_0. In the combined feature space,



(a) UMAP representation of combined feature space



(b) Original waveforms and their reconstruction

Figure 5. Sub-Figure 5a shows UMAP visualization of the combined feature space (AE latent space size four and the four waveform parameters) from the OSI SAF test set. Representative waveforms and its reconstruction (highlighted in 5a with a wider marker) are shown in Sub-Figure 5b, with the σ_0 parameter, which is logarithmic to the maximum received power, provided for reference.

new ice is the second-best separated class after leads. However, FYI and MYI remain difficult to distinguish. This trend is also visible in the UMAP representation (McInnes et al., 2020) of the feature space consisting of the AE's latent space of size four combined with the waveform parameters (Sub-Figure 5a), where leads form a distinct cluster, while FYI and MYI are more intermixed.
The results indicate that the latent space of the AE conserves additional information beyond what is captured by the waveform parameters, as demonstrated by the superior accuracy of the combined feature space compared to the baseline. The latent space, being trained to preserve information necessary for signal reconstruction, captures different aspects of the waveform compared to the traditional parameter calculation. However, this information alone is not enough to adequately separate the ice classes. The accuracy increases with latent space size, suggesting that the information content in the latent space increases as well. In the combined feature space, accuracy decreases for larger latent spaces, likely due to the curse of dimensionality or because the features from the larger latent space dominate the waveform parameters and the latent space information overlaps with the waveform parameters.
Contrary to expectations, the VAE did not outperform the standard AE. One reason could be the inherent difficulty in separating FYI and MYI. The randomness introduced by sampling from a probability distribution in the VAE's latent space may have increased the overlap between these classes, rather than enhancing their separation, making it harder to distinguish between those two classes effectively. Another reason could be the added complexity of balancing two loss functions during training. The reconstruction loss focusing on preserving information is crucial for reconstructing waveforms and is shown to enhance class separability. Introducing the KL-divergence as a second loss complicates this process. The model must balance two objectives, which may dilute the reconstruction-related information in the latent space. Additionally, enforcing the latent space to conform to a Gaussian distribution could smooth class boundaries, leading to less distinct separations compared to the deterministic latent space of the standard AE.

7. Conclusion

The major finding of this study is that self-supervised learning methods can effectively provide additional useful features to the representation of radar altimeter waveforms, enhancing their potential for sea ice surface classification. Compared to supervised learning methods, these methods offer the significant benefit of not depending on correctly labeled training data, which is a crucial advantage in polar ocean remote sensing, where ground truth data is often limited and uncertain.
While the methods proposed in this study do not yet replace the traditional parameter-based representations, they have demonstrated potential as valuable enhancements to the classification process.
Several challenges have emerged during the study. First, there is the class imbalance in the training data, which complicates the classification process. Second, distinguishing between first-year ice (FYI) and multi-year ice (MYI) remains particularly challenging. Both the auto-encoder and variational auto-encoder are designed to cluster similar waveforms together in the feature space. However, they do not explicitly enforce the separation of dissimilar waveform shapes and surface classes. The reconstruction loss employed by these models primarily focuses on keeping information needed for reconstructing the original waveforms, which may not be sufficient for effectively distinguishing between surface types.
This shows that while self-supervised methods hold promise, relying only on a reconstruction loss might not lead to optimal separation of different surface classes. Incorporating class label information into the training process may be necessary to improve performance, though this would diverge from the self-supervised approach.
In the light of these findings, future research will focus on contrastive learning, a method that can be applied in both supervised and self-supervised contexts (Le-Khac et al., 2020). Contrastive learning explicitly focuses on the separation of dissimilar samples in the embedded space by incorporating this objective into the loss function, potentially overcoming the limitations identified in the current study.

Acknowledgements

The first author, Lena Happ, is funded through the Helmholtz School for Marine Data Science (MarDATA), Grant No. HIDSS-0005. We acknowledge the Helmholtz Information & Data Science Academy (HIDA) for providing financial support enabling a short-term research stay for the first author at the Earth Observation Center (EOC) of DLR to get in exchange about data science methods in Earth observation. We thank the people from the EO Data Science Department of the Remote Sensing Technology Institute at the DLR EOC for their valuable scientific discussions and exchange of ideas during this research stay, as well as Stephan Paul for his support during the initial phase of the project.

References

Aldenhoff, W., Heuzé, C., Eriksson, L. E. B., 2019. Sensitivity of Radar Altimeter Waveform to Changes in Sea Ice Type at Resolution of Synthetic Aperture Radar. *Remote Sensing*, 11(22).

Bank, D., Koenigstein, N., Giryes, R., 2023. *Machine learning for data science handbook*. Springer Cham, chapter Autoencoders, 353–374.

Bij de Vaate, I., Martin, E., Slobbe, D. C., Naeije, M., Verlaan, M., 2022. Lead Detection in the Arctic Ocean from Sentinel-3 Satellite Data: A Comprehensive Assessment of Thresholding and Machine Learning Classification Methods. *Marine Geodesy*, 45(5), 462–495.

Copernicus Climate Change Service, C. D. S., 2020. Sea ice edge and type daily gridded data from 1978 to present derived from satellite observations. Accessed on 2023-06-01.

Dawson, G., Landy, J., Tsamados, M., Komarov, A. S., Howell, S., Heorton, H., Krumpen, T., 2022. A 10-year record of Arctic summer sea ice freeboard from CryoSat-2. *Remote Sensing of Environment*, 268, 112744.

Dinardo, S., 2016. Guidelines for reverting waveform power to sigma nought for cryosat-2 in sar mode. ESA Technical Note 23/06/2016 Issue 2 Rev 2.

ESA, 2022. Sentinel-3 SRAL/MWR Land User Handbook. S3MPC-STM-RP-0038.

Fredensborg Hansen, R. M., Rinne, E., Skourup, H., 2021. Classification of Sea Ice Types in the Arctic by Radar Echoes from SARAL/AltiKa. *Remote Sensing*, 13(16).

Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, NY.

Hendricks, S., Brockley, D., Paul, S., Sallila, H., 2024. python sea ice radar altimetry toolbox (version 0.11) (v0.11). Zenodo. https://doi.org/10.5281/zenodo.10727996.

Kingma, D. P., Lei Ba, J., 2014. Adam: A method for stochastic optimization. *ICLR 2015*, International Conference on Learning Representations (ICLR).

Kingma, D. P., Welling, M., 2014. Auto-Encoding Variational Bayes. *Conference Track Proceedings*, 2nd International Conference on Learning Representations (ICLR 2014).

Kullback, S., Leibler, R., 1951. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79–86.

Laxon, S. W., Giles, K. A., Ridout, A. L., Wingham, D. J., Willatt, R., Cullen, R., Kwok, R., Schweiger, A., Zhang, J., Haas, C., Hendricks, S., Krishfield, R., Kurtz, N., Farrell, S., Davidson, M., 2013. CryoSat-2 estimates of Arctic sea ice thickness and volume. *Geophysical Research Letters*, 40(4), 732–737.

Le-Khac, P. H., Healy, G., Smeaton, A. F., 2020. Contrastive Representation Learning: A Framework and Review. *IEEE Access*, 8, 193907-193934.

Longépé, N., Thibaut, P., Vadaine, R., Poisson, J.-C., Guillot, A., Boy, F., Picot, N., Borde, F., 2019. Comparative Evaluation of Sea Ice Lead Detection Based on SAR Imagery and Altimeter Data. *IEEE Transactions on Geoscience and Remote Sensing*, 57(6), 4050–4061.

McInnes, L., Healy, J., Melville, J., 2020. UMAP: Uniform manifold approximation and projection for dimension reduction.

Müller, F., Dettmering, D., Bosch, W., Seitz, F., 2017. Monitoring the Arctic Seas: How Satellite Altimetry Can Be Used to Detect Open Water in Sea-Ice Regions. *Remote Sensing*, 9(6), 551.

Passaro, M., Müller, F., Dettmering, D., 2018. Lead detection using Cryosat-2 delay-doppler processing and Sentinel-1 SAR images. *Advances in Space Research*, 62(6), 1610–1625.

Poisson, J.-C., Quartly, G. D., Kurekin, A. A., Thibaut, P., Hoang, D., Nencioli, F., 2018. Development of an ENVISAT Altimetry Processor Providing Sea Level Continuity Between Open Ocean and Arctic Leads. *IEEE Transactions on Geoscience and Remote Sensing*, 56(9), 5299–5319. Conference Name: IEEE Transactions on Geoscience and Remote Sensing.

Quartly, G. D., Rinne, E., Passaro, M., Andersen, O. B., Dinardo, S., Fleury, S., Guillot, A., Hendricks, S., Kurekin, A. A., Müller, F. L., Ricker, R., Skourup, H., Tsamados, M., 2019. Retrieving Sea Level and Freeboard in the Arctic: A Review of Current Radar Altimetry Methodologies and Future Perspectives. *Remote Sensing*, 11(7), 881.

Rinne, E., Similä, M., 2016. Utilisation of CryoSat-2 SAR altimeter in operational ice charting. *The Cryosphere*, 10(1), 121–131. Publisher: Copernicus GmbH.

Shen, X.-y., Zhang, J., Meng, J.-m., Zhang, J., Ke, C.-q., 2017. Sea ice type classification based on random forest machine learning with cryosat-2 altimeter data. *2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP)*, 1–5.

Shu, S., Zhou, X., Shen, X., Liu, Z., Tang, Q., Li, H., Ke, C., Li, J., 2020. Discrimination of different sea ice types from CryoSat-2 satellite data using an Object-based Random Forest (ORF). *Marine Geodesy*, 43(3), 213–233.

U.S. National Ice Center, 2022. U.S. National Ice Center Arctic and Antarctic sea ice charts in SIGRID-3 format.

Wernecke, A., Kaleschke, L., 2015. Lead detection in Arctic sea ice from CryoSat-2: quality assessment, lead area fraction and width distribution. *The Cryosphere*, 9(5), 1955–1968.

Wingham, D., Francis, C., Baker, S., Bouzinac, C., Brockley, D., Cullen, R., de Chateau-Thierry, P., Laxon, S., Mallow, U., Mavrocordatos, C., Phalippou, L., Ratier, G., Rey, L., Rostan, F., Viau, P., Wallis, D., 2006. CryoSat: A mission to determine the fluctuations in Earth's land and marine ice fields. *Advances in Space Research*, 37(4), 841–871.

World Meteorological Organization, 2022. The 2022 GCOS ECVs Requirements (GCOS 245). https://library.wmo.int/doc_num.php?explnum_id=11318.

Zygmuntowska, M., Khvorostovsky, K., Helm, V., Sandven, S., 2013. Waveform classification of airborne synthetic aperture radar altimeter over Arctic sea ice. *The Cryosphere*, 7(4), 1315–1324.