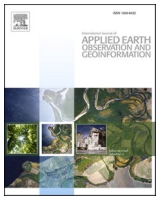
International Journal of Applied Earth Observation and Geoinformation 126 (2024) 103616

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Retrieving freeze-thaw states using deep learning with remote sensing data in permafrost landscapes

Yueli Chen [a,*], Shile Li [b], Lingxiao Wang [c], Magdalena Mittermeier [a], Monique Bernier [d], Ralf Ludwig [a]

[a] Department of Geography, Ludwig-Maximilians-Universität München (LMU), 80333 Munich, Germany
[b] Algolux, Canada
[c] School of Geographic Science, Nanjing University of Information Science and Technology (NUIST), Nanjing 210044, China
[d] Centre Eau, Terre & Environnement, Institut National de la Recherche Scientifique, Quebec City, QC G1K 9A9, Canada

ARTICLE INFO

Keywords:
Freeze-thaw state
Permafrost landscape
Sentinel-1
Backscatter
Deep Learning
CNN

ABSTRACT

The soil freeze–thaw (FT) cycle is a critical component of the terrestrial cryosphere and plays a significant role in hydrological, ecological, climatic, and biogeochemical processes within permafrost landscapes. The FT states can be monitored with in-situ field measurements, but these procedures are costly and limited to single chosen sites. Remote sensing data provides the opportunity to collect information repeatedly across extensive geographical areas. To explore a more effective way to monitor the FT states in the terrestrial cryosphere, in this study, we used microwave and optical remote sensing data and introduced the Deep Learning approach to simulate the soil FT states in the western part of Nunavik, Canada.

Two networks, Multilayer Perceptron (MLP) and Convolutional Neural Network (CNN), were trained and tested with over 35,000 and approximately 54,000 randomly selected data samples, respectively. The trained CNN networks outperformed the MLP networks, achieving the highest testing accuracy of 95.67% and the highest validation accuracy of 87.28% based on ground truth data from 32 measurement stations from all seasons across the year. This study proposed the reference periods concept for convenient labeling in data preparation and tested different combinations of influence variables to achieve better transferability of the method for future studies. Our findings offer a more effective way to monitor FT states in the terrestrial cryosphere, offering valuable insights into the consequences of climate change on permafrost landscapes. Moreover, the suggested deep learning approach can be easily expanded when additional input sources are accessible. This expansion has the potential to further improve the model's performance for the FT retrieval.

1. Introduction

The freeze–thaw (FT) cycle of soil, a crucial component within the terrestrial cryosphere, significantly influences climatic, hydrological, ecological, and biogeochemical processes in permafrost landscapes.

The FT states transitions can be accurately observed using traditional ground observation methods. The research community expended significant effort to establish observation stations for monitoring ground status. For example, the globally distributed stations of the Global Terrestrial Network for Permafrost, known as GTN-P, along with its circumpolar active layer monitoring (CALM) program, have been continually providing air and soil temperature data for decades (Burgess et al., 2000; Brown et al., 2000; Nelson et al., 2004). There are also a series of local-specific research programs. For example, on the Tibetan Plateau, long-year measurements of soil temperature, soil water content, heat flux etc., are continued by establishing automatic weather stations (Zhao et al., 2021). However, the information is low in density and only limited to the location of the stations, and the establishment and maintenance of stations are difficult and costly due to the infrastructural challenges in complex cryospheric landscapes.

Besides the in-situ measurements, remote sensing can also support the monitoring, providing physical observations of various variables in ever-increasing spatiotemporal coverage. In particular, active and passive microwave remote sensing has demonstrated significant potential in

* Corresponding author at: Teaching and research unit for Physical Geography and Complex Environmental Systems, Luisenstr. 37, Room AU106, 80333 Munich, Germany.
E-mail address: chen.yueli@lmu.de (Y. Chen).

https://doi.org/10.1016/j.jag.2023.103616
Received 30 May 2023; Received in revised form 8 November 2023; Accepted 9 December 2023

effectively monitoring landscape FT dynamics. Microwave remote sensing is sensitive to changes in the dielectric constant on the surface, which are caused by changes in soil water content and phase. It offers several advantages, including weather and light independence, as well as the ability to penetrate snow, ice and low to medium vegetation coverage. Additionally, optical satellite sensor systems can offer supplementary information on terrain, snow area, and land cover, while thermal sensors can be utilized to capture the fluctuations in land surface temperature.

Due to the specific characteristics and advantages, microwave remote sensing data have been used and evaluated for retrieving the FT states of soil. Very widely used are the L-band data, especially since missions such as the Soil Moisture and Ocean Salinity (SMOS) of ESA and the Soil Moisture Active Passive (SMAP) of NASA were successfully launched (Derksen et al., 2017; Rautiainen et al., 2018; Roy et al., 2015; Rautiainen et al., 2014; Rautiainen et al., 2016). The FT detection using different L-band data was assessed in a series of works globally (Kim et al., 2019) and regionally with special focuses. For example, extensive evaluations have been conducted in the northern terrestrial areas, such as Alaska, Finland, and large regions in Canada, specifically emphasizing the boreal forest and tundra (Derksen et al., 2017; Roy et al., 2015; Roy et al., 2017; Colliander et al., 2012; Davitt et al., 2019; Lyu et al., 2018). Generally, the L-band spaceborne missions can provide valuable information to support the FT states' monitoring across expansive spatial scales, with frequent revisits typically occurring every 2–3 days.

The potential of C-band microwave data to deliver information on FT state in the topsoil was also recognized by the researchers. In 1994, Rignot and Way (1994) proposed monitoring the FT cycle in high-latitude terrestrial ecosystems using C-band synthetic aperture radar (SAR) data. Subsequently, in 2012, Naeimi et al. (2012) explored the use of C-band scatterometer data to extract FT conditions through an empirical threshold-analysis algorithm; while Zwieback et al. introduced a sensor fusion algorithm incorporating Ku- and C-band scatterometer data to determine the FT state. In the 2020 s, studies using the new open accessible Sentinel-1 (S1) SAR data appeared. This data set has offered continuous global coverage at a 10 m resolution since 2016. Bergstedt et al. (2020) and Cohen et al. (2021) tested the S1-based FT retrieval in Finnland; Chen et al. (2022) tested the S1 data for retrieving the surface FT state in the 0–10 cm depth over a large region in northern Quebec spanning all typical terrestrial permafrost zones; also on the Qinghai-Tibet Plateau, the data was tested by Zhou et al. in 2022.

With the measurements obtained from satellite radiometers and scatterometers, numerous algorithms have been employed or specifically developed over the course of several decades to monitor the FT states and their periodic transitions, such as the double-index algorithm (Jin et al., 2015; Judge et al., 1997; Zhang and Armstrong, 2001; Zuerndorfer and England, 1992, Zuerndorfer et al., 1990), decision tree algorithm (Jin et al., 2009), and standard deviation algorithm (Han et al., 2015). Additionally, there are algorithms aimed at identifying patterns in seasonality by establishing freeze and thaw state references. These include the seasonal detection method (Derksen et al., 2017; Rautiainen et al., 2018; Kim et al., 2011) and the polarization ratio (PR)-based algorithm (Roy et al., 2015; Roy et al., 2017), and their variations, such as the discriminant function algorithms (Kou et al., 2018; Wang et al., 2018a; Zhao et al., 2011) and the modified seasonal threshold algorithm (MSTA) (Kim et al., 2017). These methods improved the monitoring and understanding of FT cycles based on microwave data but still have limitations. For example, it is difficult to take complex environmental information into account for the deterministic algorithm. Previous works found that the land cover types and the general heterogeneity in the surrounding area influenced the estimation of the FT states. However, it was barely possible to find clear rules that could be directly built into the algorithm to optimize the processing (Chen et al., 2022; Johnston et al., 2022).

Overall, microwave remote sensing has been extensively researched for monitoring soil freeze–thaw dynamics, resulting in useful algorithms and products, such as the SMAP's L3_FT_P with 36 km resolution (O'Neill et al., 2019). However, these products are limited to large-scale monitoring at levels ranging from a few kilometers to several tens of kilometers. Additionally, most methods focus solely on microwave signals. Although it has been found that environmental factors like land cover type can significantly impact results, it is still hard to quantify their effects clearly for use in processing. To address these limitations, we propose using deep learning models on high-resolution radar data from the Sentinel 1 mission, which provides data at a resolution of 5 m. Our approach aims to provide stable and highly accurate results while considering environmental factors and spatial relations, bridging the gap between existing methods and the need for more fine-grained monitoring of soil freeze–thaw dynamics.

By transitioning from a deterministic approach to machine learning and extensive training with a labeled, large dataset (approximately 50–80 k data samples), the deep learning approach enables the model to discern superior mappings from various inputs to the output result, automatically assigning weights to each input variable based on its importance and contribution. This approach is expected to help us discover hidden relationships between features that may have been overlooked or could not be effectively summarized using existing methods.

The primary objective of this work is to explore a more efficient and easy-to-use framework for monitoring FT states in the terrestrial cryosphere using the deep learning approach. The goal is to develop a well-performing framework that can be applied over large areas and is transferable to further research in related fields. With the ability to consider environmental factors and provide fine-grained monitoring, our proposed approach has the potential to advance our understanding of soil freeze–thaw dynamics and improve our ability to monitor this critical aspect of the terrestrial cryosphere.

2. Study area and database

2.1. Nunavik

The western part of Nunavik, located in Quebec, Canada, was chosen as the study area (refer to Fig. 1). Nunavik is a vast region situated north of the 55th parallel and is home to the Inuit communities of Quebec, with 14 villages scattered along the coastal areas of Hudson Bay and Ungava Bay. This region exhibits high complexity and diversity and garnered significant attention in various research domains, including permafrost studies, coastal geology, and geomorphological characterization (Beck et al., 2015, Wang et al., 2018b, Wang et al., 2020, Pelletier et al., 2018, Ropars et al., 2015, Calmels et al., 2008, Allard and Seguin, 1987; Hachem et al., 2009).

Nunavik experiences two main climate types: arctic in the north and subarctic in the south. These regions are characterized by prolonged, extremely cold winters and brief, cool summers with extended daylight hours. The area spans all four primary permafrost zones, namely continuous, discontinuous, sporadic, and isolated permafrost, extending from the northern to the southern regions. Notably, there are discernible spatial variations in vegetation cover, such as the presence of the northern treeline at the boundary between the Arctic and sub-Arctic regions and the transitional zone from discontinuous to continuous permafrost.

In the context of global climate change, climate and permafrost conditions in Nunavik are undergoing significant transformations. As highlighted in the study by Allard et al. (2007), temperatures in the region have been consistently above the long-term average since 1995, with a continued rise observed since the early 2000s (KRG, 2011). This warming trend has also been observed as increasing ground temperatures since the early 1990s (Romanovsky et al., 2010), resulting in a warmer vertical temperature profile within the permafrost and a deeper extent of summer thaw (Allard et al., 2012).

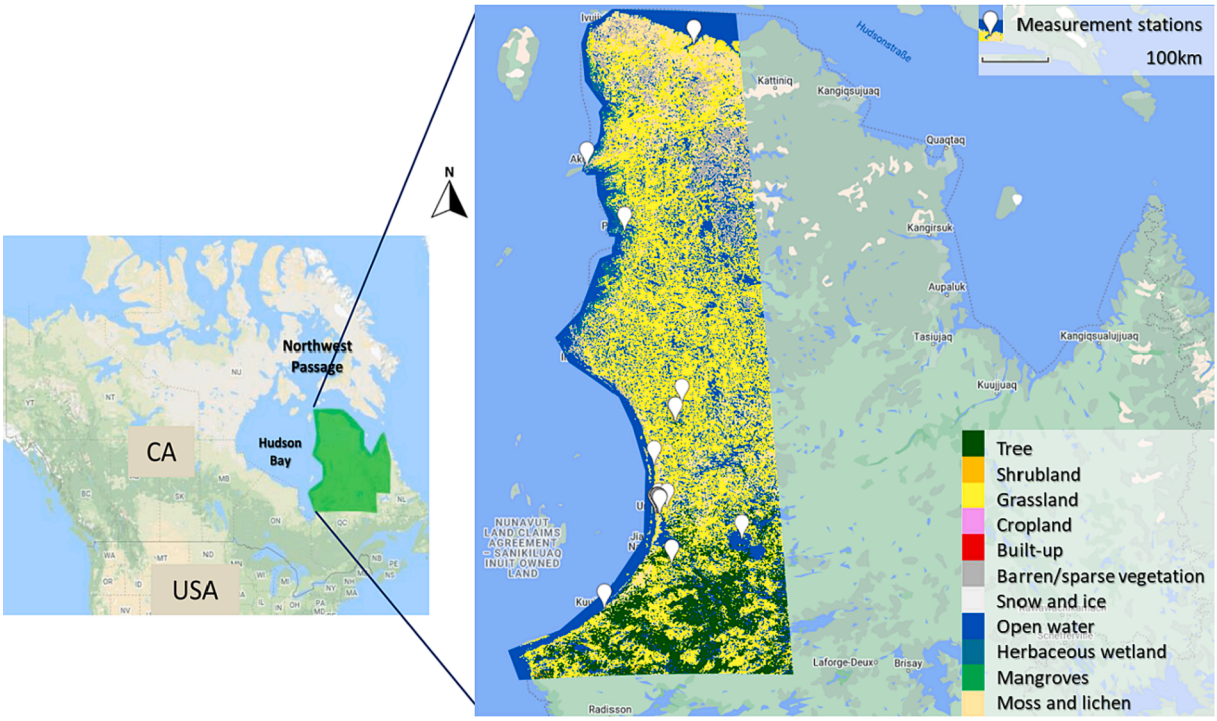


Fig. 1. Study area (Western Nunavik) with metrological stations (white). The map was created on Google Earth Engine, land cover type from WordCover 10 2020. A total of 32 measurement stations are depicted on the map, and there may be some overlapping symbols due to scaling issues. For detailed information on each measurement station, please refer to the appendix.

The chosen study area (Fig. 1, shown with the land cover base map) crosses all relevant climate types, vegetation covers, and permafrost states introduced above, which makes it possible to test the approach under different environments. The western coastline (Eastern Hudson Bay) is home to seven Inuit villages, which are quite vulnerable to the changing permafrost conditions, making improved monitoring of the FT cycles especially important. Additionally, the selected study area encompasses a significant portion of the operational research stations established within the Nunavik region, thereby offering the most ground truth data for implementing and validating the Deep Learning algorithm.

2.2. Remote sensing data

2.2.1. Sentinel-1

The Sentinel-1 (S1) mission, initiated by the European Commission (EC) and the European Space Agency (ESA) as part of the Copernicus joint initiative, has been instrumental in providing consistent, all-weather, day-and-night imagery of the Earth's surface. With the launch of Sentinel-1A (S1A) in April 2014, followed by Sentinel-1B in April 2016 (S1B), this mission employs a dual-polarization C-band synthetic aperture radar (SAR) instrument, offering high spatial resolution (10 m) and temporal resolution (6–12 days). Notably, the free and open-access availability of S1 data contributes to its widespread usage and stability.

For this study, the satellite imagery utilized originated from the S1A and S1B (mission ended in August 2022) satellite constellations [Copernicus Sentinel data 2022, processed by ESA]. The data collection period encompassed observations acquired before January 2022. The images obtained were level-1 ground range detected (GRD) products. Only ascending mode data were utilized due to the limited availability in descending mode across the study area. The data collection involved the use of the interferometric wide swath mode (IW) and included both vertical–vertical (VV) and vertical-horizontal (VH) polarizations.

2.2.2. WorldCover

This work also included land cover information from a recent product provided by the WorldCover project, which was initiated by ESA (Zanaga et al., 2021). The ESA WorldCover map (EWC) was developed at a resolution of 10 m for 2020 and 2021. This product provides a full spatial coverage over the chosen study area and was generated within the temporal range of the S1 data set used for this study (2016–2022). Moreover, this product has good compatibility with the chosen S1 data, since it was developed based on data from Sentinel-1 and Sentinel-2.

The accuracy and reliability of the EWC were ensured through near-real-time validation processes, the version 2020 reached an overall accuracy of 74.4%. The map includes 11 land cover types, detailed in the legend in Fig. 1. The typical land cover type in the cold region "moss and lichen" was included. As shown in Fig. 1, the transition of vegetation types across the northern treeline has been well interpreted on the map. The product version for 2020 was utilized for this study.

2.3. Ground truth from in-situ measurements

Besides the remote sensing data, soil temperature data measured at 32 meteorological stations (see Fig. 1) was collected to provide ground truth information for the validation process. The stations are distributed across the study area, covering multiple diverse environments. Some of them were established by ourselves (Bernier et al., 2019). They provide data that overlap with the service period of the S1 mission. Considering the penetration depth of S1 data, only the temperature data, aggregated to daily values, in the top soil layer (0–10 cm) were used. All data were provided by the Nordicana D – Mission, published by the Centre for Northern Studies (CEN) since 1964; unfortunately, all the used stations stopped updating between 2020 and 2022. On the Nordicana D website, one can access already published data and raw data, including unpublished data and their updates. Complete metadata and all useful information about the data from different measurement stations can be found on the site https://nordicana.cen.ulaval.ca.

Regarding the temporal availability and penetration depth of the S1 data, this study utilizes soil temperature data measured from the topsoil layer up to a depth of 10 cm between 2014 and 2019. Detailed information of measurement stations including soil layer depth are attached in the appendix.

3. Methods

In this work, we intended to accurately retrieve the soil FT states in the permafrost landscape from the remote sensing data by developing a deep learning model. The model is designed to answer whether the observed soil is frozen or thawed at the acquisition time, which involves a binary classification scheme. We trained the model on a dataset of SAR signals and land cover information to learn how to map input variables to specific output class labels. The two output classes are "frozen" and "thawed". Detailed information on how the data was prepared for the model is provided in Chapter 3.2. Two distinct neural network architectures, Multilayer Perceptron (MLP) and Convolutional Neural Network (CNN), were employed and evaluated for the modeling process.

3.1. Deep learning algorithms

In this chapter, we present a comprehensive description of the deep learning algorithms employed in the study for classifying frozen/thawed soil states. Specifically, we tested and compared the performance of two well-established approaches: MLP and CNN, to develop an efficient procedure for our task.

MLP and CNN have been widely used in previous studies for various classification tasks (LeCun et al., 1989; Krizhevsky et al., 2017). MLP, a conventional neural network architecture, comprises interconnected nodes organized into multiple layers, while CNN is purpose-built for analyzing data with a grid-like topology, such as images. Both approaches are effective at learning complex relationships between input and output variables. The subsequent sub-chapters provide in-depth explanations of the specific application designs utilized in this study.

3.1.1. Multilayer Perceptron

Multilayer Perceptron (MLP) is a type of classical neural network that is very flexible and can be used to learn a mapping from inputs to outputs. MLP is often used as a starting point to determine whether a task can be accomplished with neural networks. The results can then be used as a baseline for a comparison to determine whether other models perform better and add value to the analysis.

In this work, a MLP model structure was used with two hidden layers to provide levels of abstraction (Fig. 2). Predictions were made on the output layer using a widely adopted loss function called Binary Cross Entropy. This loss function is commonly utilized for binary classification tasks. The activation function used was Rectifier Linear Unit (ReLU) (Agarap et al., 2018), and regularization techniques Batch Normalization (BatchNorm) and Dropout were also included in the network (Kukačka et al., 2017).

As a basic network, MLP has a necessary structure that is not overly complex to run. The usage of MLP allows for quick testing to determine whether using the DL algorithm to model the FT states is feasible and whether more complex networks are needed. With MLP, all the data was processed pixel-wise, meaning the FT states were classified for every pixel based on the input information only from the pixel itself. This pixel-wise processing allows for flexible use of data; for example, the data can be excluded according to land cover types, regardless of where it is located on the image.

To consider the environmental factors, variables, such as land cover types, can be added as input. However, it is important to note that in this approach, each pixel is processed independently without considering its spatial relations with neighboring pixels.

3.1.2. Convolutional neural networks

Convolutional Neural Network (CNN) is a commonly applied neural network for image processing. Compared to MLP, which processes each pixel independently, CNN utilizes convolution filters to detect patterns across adjacent pixels, thereby enabling a better understanding of the spatial relationships between pixels. This makes them particularly well-suited to tasks where understanding the relationships between pixels is essential, such as image segmentation. In light of the fact that the classification of the FT states is known to be influenced by surrounding environmental factors, such as land cover types and spatial heterogeneity, the utilization of CNN could be instrumental. In addition to its

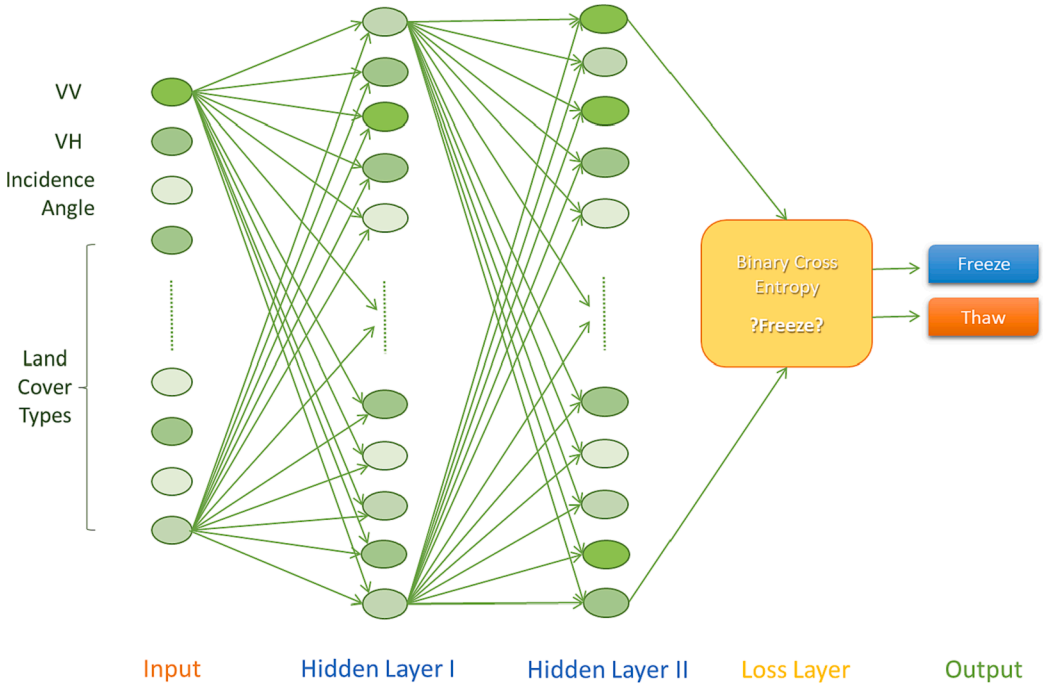


Fig. 2. Model construct MLP.

ability to capture spatial relationships, CNN also has the potential to prevent overfitting by using techniques such as pooling and Dropout. By introducing the CNN approach to this study, our goal is to leverage these advantages to improve the accuracy of our classification results. (Bishop, 1995, Bishop and Nasrabadi, 2006; Goodfellow et al., 2016; Botalb et al., 2018).

The CNN structure used in this study was built, as shown in Fig. 3. Two convolutional layers were built, and the activation function was again ReLU. Several regularization techniques were used and tested, including BatchNorm, Dropout layers, stepwise decreasing learning rate (Bengio, 2012), and weight decay in the Adam optimizer (Bengio, 2012, Zhang et al., 2018, Loshchilov and Hutter, 2017). The regularization techniques were used in deep learning training to prevent overfitting by adding constraints to the model during training, which limited its complexity and prevented it from memorizing the training data. It also helped to improve the overall model performance, such as weight decay, by reducing the complexity of the model and encouraging it to focus on the most important features. These techniques are important because they can help to improve the generalization performance of the model, allowing it to perform better on unseen data. The contribution of the different regularization techniques to the networks is presented in Chapter 4.3.

3.2. Model configuration and data preparation

3.2.1. Input and output

The input variables for the modeling process include backscatter signal VV (BScVV), backscatter signal VH (BScVH), Incidence Angle (Angle), and land cover types (Lc). These four variables were used in different combinations during the training process to test their contribution to the task. The best-performing combination was chosen and presented in Chapter 4.1. The output variables are two classes: “frozen” and “thawed”. Through the modeling process, a probability of the soil state being “frozen” was predicted.

BScVV, BScVH and Angle are extracted from S1 data, and Lc is extracted from EWC data. These four input variables were further reparameterized to 14 channels: the Lc with 11 types was converted to individual one-hot encodings that only contained integers as 0 or 1, while the other three channels, BScVV, BScVH, and Angle, contained continuous values and were generalized to a similar range (BScVV*0.1, BScVH*0.1, Angle*0.05). Every data sample consisted of 14 channels of input variables and one label in the form of integers representing frozen (1) or thawed (0) soil states.

The data preparation then differed between the two networks, with MLP using pixel-based data and CNN using patch-based data. Spatially, the data samples were randomly selected within the study area (Fig. 1). For CNN, a patch of 51x51 pixels was created around each sampling point, with each pixel containing individual values of input variables and the output label. This allows the model to consider the spatial patterns within the surrounding environment. However, a limitation of this approach is that it is impossible to exclude specific pixels from the patches, meaning that all land cover types were included in the processing, regardless of whether there was soil. To explore whether excluding non-soil areas is necessary during such training, we trained the MLP on datasets with and without soil-free land cover types (such as water bodies, infrastructures, etc.) since pixel-based processing can provide more flexibility in terms of data manipulation. The results of this analysis are presented in Chapter 4.1.

3.2.2. Reference periods

To determine the output label of each data sample, a concept of reference period was introduced. This involved defining two reference seasons for the frozen and thawed states, representing the entire frozen or thawed time in a year. The reference seasons were chosen following long-term meteorological measurements and empirical knowledge from the study area, during which the soil should definitely be frozen or thawed. This concept helps to cross the limited in-situ measurement data barrier and makes it possible to prepare accurate data over a large area. All the input data were collected only from these two periods every year, then labeled as “frozen” or “thawed” according to the time when they were acquired.

The samples were prepared based on a basic data set of S1 data, which was spatially delimited with the designated study area (see Fig. 1), and temporally delimited with the reference periods. All the S1 images acquired in the reference periods were first masked to the boundary of the study area, then combined with the EWC map and given a label. The data samples were then randomly selected in every preprocessed image and further configured as data pixels for MLP and data patches around chosen samples. On each image, 150 samples were taken. This is an amount we tested that can reach a good distribution over the imaged area without too much sample overlapping.

The study area exhibits a strong seasonality in the FT state changes,

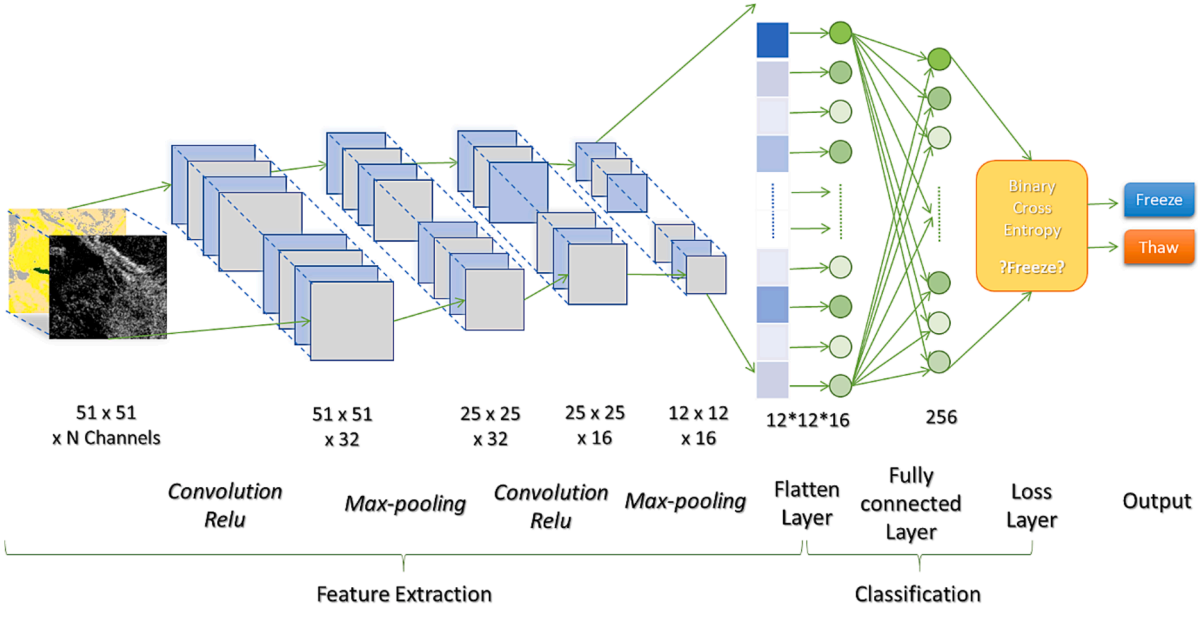


Fig. 3. Model construct CNN.

as demonstrated by the examples in Fig. 4. During the cold season, the soil remains frozen for an extended period, while the summer thawing season is relatively short. However, the duration and starting time of the FT seasons vary between stations and even between years at the same station. In some cases, there are quick and distinct transitions between the freeze and thaw states, such as in the spring of 2017 and fall of 2018 at station 9. In other cases, like in the spring of 2017 at station 37, soil temperature fluctuations around 0 °C for almost a month were exhibited. Sometimes multiple rounds of changes between freeze and thaw states can be observed by one transit before stabilizing. To ensure the accuracy of the labeling, the study only considered the central months of the empirical FT seasons. August (213th-243th day in a year, "Aug") was chosen as the thawed reference since the summer thawed season in Nunavik is typically short, with a high degree of interannual variability in its starting and end times. Correspondingly, February (31st-60th day in a year, "Feb") was chosen as the frozen reference with the same length as the thawed reference to ensure the data balance. Considering the overall long winter frozen season in the study area, an additional frozen reference was selected, spanning from mid-January to mid-March (20th-75th day in a year, "JFebM"). This interval also provided a safe window to capture the frozen signals. The results obtained using the references Feb and JFebM were compared to support the proofing of whether a short reference from the middle of the season suffices to represent the entire season for modeling purposes.

3.3. Evaluation

Three types of datasets were used in the workflow: the training set, the testing set and the validation set. Using the training set, multiple networks were trained by optimizing overall accuracy and minimizing loss. Subsequently, the performance of these networks was assessed using an independent testing set that was distinct from the data used for training. To avoid potential overfitting and confirm the transferability to practices, the performance was further measured on a third independent dataset called the validation set (details see Chapter 3.4).

The training and testing data sets were split from a basic data set built according to the reference periods concept described in Chapter 3.2.2. The splitting process was done randomly with regard to the balance between frozen and thawed proportions.

The validation sets were built based on ground truth data collected from in-situ measurements. The selection of validation data was not limited to the reference periods, but used every available measurement over the entire year. The total amount of data for each dataset and the respective amounts of data for each of the two classes (frozen and thawed) are listed in Table 1.

To visualize the misclassification error, the confusion matrix was used to break down the results into four categories: true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN) (Fawcett, 2006). The terms "positive/negative" refer to instances in the predicted class "1/0," while "true/false" indicates whether the predicted result matches the actual class. To present more clearly, the terms "positive/negative" were written directly as "Frozen(Fr)/Thawed(Th)" in the Result chapter (Table 4). The metric, Precision, which measures the proportion of items correctly classified as positives (TP) out of all items classified as positives, and Recall, which indicates the proportion of true positives out of all items that are indeed positives, were calculated. Additionally, the f1-score was calculated based on precision and recall to check for any significant imbalances in the classification results. In case of a strong imbalance, the classification results may be biased towards the majority class, leading to poor performance in identifying the minority class, which can affect the reliability of the model's performance.

3.4. Validation design

For validation, we abandoned the widely used approach, where the validation data is separated from the same dataset as the training and testing data before the training process. We introduced fully independent ground truth data from the in-situ measurement stations. The data is not only independent because it is not used during the training process, but it also shows temporal independence by including data from the time intervals in the year which are not contained in the reference periods. This concept was designed to help us confirm whether the model results are robust and to test whether the model can only work with the given references or all around the year.

Two validation sets were built, the Ground Truth set (GrTr), and the Ground Truth Extended set (GrTrEx), and are described in the following sub-chapters.

3.4.1. Ground truth set

The ground truth dataset (GrTr) was built on in-situ measurements from 32 chosen stations scattered over the entire study area. The frozen and thawed days were carefully defined individually on each station for every year, according to the measured soil temperature (Fig. 5). Considering the penetration depth of S1 data, only the soil temperature within 10 cm was used. 0°C was used as a threshold between frozen and thawed state; the dates were not immediately taken as samples, but firstly when the temperature stabilized above or below 0 °C for several consecutive days. After defining the frozen and thawed dates of each station, the data available within that range was collected from each relating station and configured to the same form as the training and testing data.

The ground truth dataset contained 1667 data samples, 948 labeled as frozen and 719 labeled as thawed (Table 1). The data scattered spatially over 32 locations and temporally differently over a maximum of three years (late 2016 to summer 2019); the data collection was done across the entire year, not limited to the reference periods used for training. This data set provided very exact and reliable ground truth information over different environments, but with a limited amount because of the short overlap of the in-situ measurements (ends in summer 2019) with remote sensing data (stable S1 data availability since late 2016). To overcome this limitation, an additional dataset was built to further explore ways to expand the validation set (see Chapter 3.4.2).

3.4.2. Ground truth extended set

The ground truth extended set (GrTrEx) was built with the goal of finding a useful way to create a larger database for validation. As in building the GrTr set, the soil temperature data measured in the top 10 cm were used to distinguish the state between frozen and thawed. Spatially, the data were collected at the location of the 32 stations. The freeze/thaw days defined for each site were no longer directly used as a temporal range of data collection, but two freeze/thaw reference periods were selected for each site individually through comparative observations over all available years. The references were then repeatedly used for each station separately for every year.

To ensure reliability, the reference periods were selected very carefully and kept relatively short, which means fewer dates per year are used compared to the GrTr set. This method does not require complete temporal overlapping between remote sensing data and measured data so that longer time series of soil temperature statistics can be considered, and the entire service span of S1 data that is still growing can be included for data collection. As shown in Table 1, the amount of validation data samples increases from 1667 to 4390.

In the later validation process, both the GrTr set and GrTrEx set were used. The results from GrTr set were used to determine the model performance, and the results of the GrTrEx set were compared to test whether the concept of the GrTrEx set was adding reliable value.

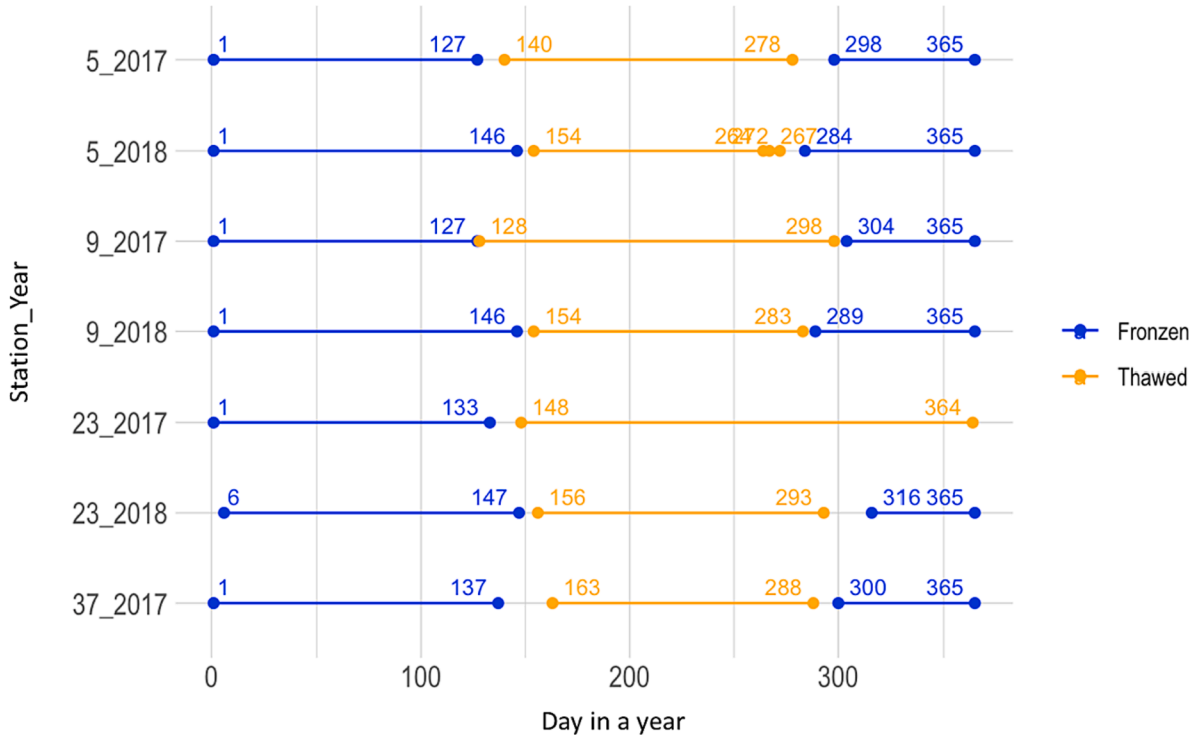


Fig. 4. Representation of frozen and thawed days at example measurement stations from example years. (The x-axis denotes the day number within a year, while the y-axis labels indicate the measurement station ID and the corresponding measurement year. For instance, "5_2017" represents data measured at the 5th station in the year 2017.).

Table 1
Listing of counts of data samples employed for various methods (MLP/CNN) across reference time periods and sets (Training, Testing and two Validation sets) for each class (Frozen, Thawed).

Method	Reference	Dataset	Frozen	Thawed	Sum
MLP (Pixel)	Feb. / Aug.	Training	13,688	12,708	26,396
		Testing	4561	4276	8837
CNN (Patch)	Feb. / Aug.	Training	20,777	19,692	40,469
		Testing	6916	6534	13,450
	m.Jan.-m.Mar. / Aug.	Training	38,580	19,692	58,272
		Testing	12,771	6534	19,305
	Validation	Ground Truth	948	719	1667
		Ground Truth extended	2848	1542	4390

4. Results

4.1. Data combinations

The results of various Deep Learning training experiments using different combinations of input variables are presented in Table 2. The top row represents the results obtained with the set of backscatter signals, Incidence Angle and land cover types, while each subsequent row reduces by one variable. The difference between polarizations is represented in columns. From left to right, the columns represent the overall accuracy of four types of training respectively: training with CNN using both VV and VH polarizations, with MLP using both VV and VH, with MLP using only VV, and with MLP using only VH. For the same method (CNN/MLP), the trainings were conducted based on the same settings.

The best result was obtained using all four input variables, as shown in the first row and first column. A clear trend can be observed in the table, where the further down and to the right the experiments are, the lower the accuracy achieved. In general, using only the backscatter data can already achieve decent accuracies; MLP achieved close to 70 %, while CNN achieved over 85 % accuracy.

The variables Lc and Angle each contribute additional benefits to enhance the model. The inclusion of Lc results in an over 8 % performance improvement when using CNN and up to 5 % when using MLP, whereas the addition of Angle leads to a slight 1.3–2.5 % enhancement. The highest accuracy is attained when all variables are utilized together. Among the polarizations, VV performed slightly better than VH, while the combined use of both polarizations resulted in significantly higher accuracy. Therefore, the most suitable combination for this training is all four input variables backscatter signal VV, backscatter signal VH, incidence angle and land cover types. In subsequent model optimization and evaluation, this input combination is applied.

We further conducted comparative experiments using MLP to investigate the effect of in- and excluding the soil-free data on the training results, as shown in Table 3. Four experiments were performed with the same model settings, including (a) using land cover as an input variable and excluding the soil-free data, (b) using land cover as an input variable and including soil-free data, (c) not using land cover information and including the soil-free data, and (d) not using land cover information and excluding the soil-free data. The concern behind this is, that the soil-free area is not necessary for a survey of the soil state, and might bring interference items to the classification. However, the patch-based CNN, which can achieve better results than MLP, does not allow

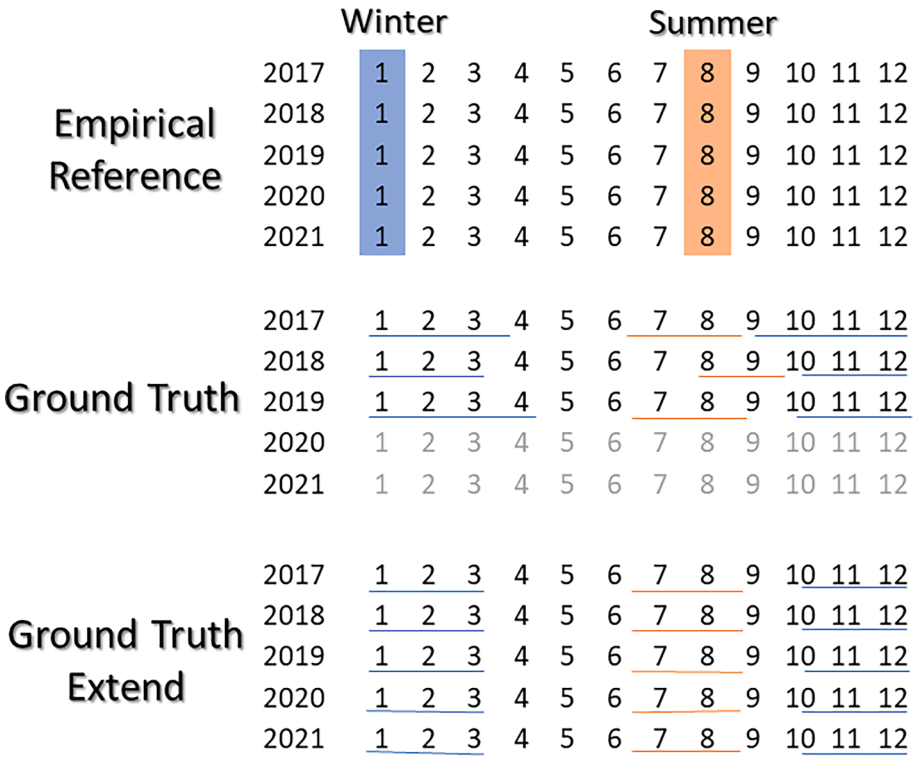


Fig. 5. Data Selection Overview: This figure illustrates the categorization of months as freeze (underlined in blue) and thaw (underlined in orange). January and August are chosen as reference months for freeze and thaw, respectively. Note that "Ground Truth" and "Ground Truth Extended" involve slight variations in the selection of months; additional details are provided in the accompanying text. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Training Accuracy (%) across various methods (MLP, CNN) with different input data combinations (Backscatter Signal = BSc; Incidence Angle = A; Land Cover = Lc); the used backscatter polarizations (VV, VH, VV + VH) showed in columns.

Model	CNN	MLP		
Polarization	VV + VH		VV	VH
BSc + A + Lc	94.17	75.97	72.59	70.74
BSc + Lc	93.40	74.68	71.14	69.26
BSc + A	86.37	72.44	68.95	68.99
BSc	85.08	69.97	67.64	67.24

Table 3
Accuracy for both the training and testing datasets under varied conditions (All LC: data with soil-free samples, SF: data without soil-free samples) with two combinations of input variables (VHAL: VV + VH + Incidence Angle + Land Cover Types, VHA: VV + VH + Incidence Angle).

Features	Accuracy %		limit of Samples
	Train	Test	
VHAL	74.67	75.32	NoWBI
	74.47	75.35	AllLc
VHA	71.88	72.60	
	71.87	72.49	NoWBI

for the exclusion of specific areas from the patches arbitrarily. Moreover, we know from previous studies that the radar signals might exhibit distinct behavior under different land cover types, for example, in forest versus on bare ground; and that the FT dynamic may be influenced by the surrounding environment, such as if a test site is located near a waterbody or a sealed road. We anticipated CNN to help to learn from these spatial patterns and therefore aimed to verify through this analysis the contribution of land cover information to the training and the acceptability of retaining soil-free data in the training.

The results show no significant difference between the results of trainings excluding and including soil-free data, with the results of experiments (a) and (b), (c) and (d) being very close. Adding land cover information significantly improved the results, demonstrating its contribution to the modeling, which is consistent with the findings shown in Table 2.

These findings support that the additional step of removing soil-free data can be omitted. This conclusion reduces doubts about the use of other methods, such as CNN, that require input in the form of patches.

4.2. Model evaluation and optimization

According to the results presented in the previous chapter, it is apparent that the training using CNN can achieve significantly better results than using MLP (Table 1) with 15–20 % higher accuracies. Therefore, the CNN was chosen for further optimization and evaluation. This evaluation of model performance is mainly based on the validation results using the both validation sets. The f1-score was used as a complementary metric, the training and testing accuracies were further calculated, but only as a supplement not used for any decision.

To improve the model performance, a series of regulations were tested, including BatchNorm, Dropout, learning rate and weight decay in Adam. In general, adding the Dropout layers is very recommended, while BatchNorm is less functional. We compared adding BatchNorm in convolutional Layers, adding BatchNorm in convolutional and fully connected Layers, and adding BatchNorm in combination with Dropout. In the training process, all the networks with BatchNorm reached significantly better accuracy than those without. Adding BatchNorm only in convolutional layers helped the model to reach the highest training accuracy. By evaluation with the test set, the networks with joint use of BatchNorm and Dropout achieved the best performance. However, the performance could generally not be confirmed by the

validation set, where all the networks with BatchNorm resulted in a 5–10 % lower validation accuracy than the others. This showed that the thereby resulted enhancement was not robust.

A stepwise learning rate can also be employed to optimize the model performance. This technique adjusts the learning rate, a hyperparameter that controls the magnitude of adjustments made with respect to the gradient of the loss function. Specifically, the learning rate can be reduced during the latter part of the training when the model approaches the minimum of the loss function. This approach prevents overshooting or deviating from the optimal solution, which can occur if the learning rate is too high in the final stages of training. Changing the steps and dropping factors lead to improvements, which were, however, not significant in this case; therefore, a deeper adjustment of the stepwise learning rate is useful but not urgently necessary. The use of weight decay in the Adam optimizer results in a reduced accuracy with respect to the training and testing set (only below 90 %). However, the performance tested with the validation sets is still close to the best results. In general, the contribution of weight decay is noticeable but can also be covered by other regulations.

Overall, the most suitable combination of setting for network regulations that can help to reach the highest accuracy both regarding the training and testing sets (unseen for the model) and at the validation sets are: Dropout in convolutional and fully connected layers, stepwise learning rate, no BatchNorm and no weight decay in Adam.[1]

In addition to the test and validation accuracy, the confusion matrix was also computed to check whether the classification results are balanced, mainly whether a large number of errors are concentrated in a specific class. Multiple networks with various performances were tested using a confusion matrix, covering networks with high validation accuracy and high testing accuracy, and also networks that performed well in the testing process but worse in validation. In total, no significant bias was found. The precision, recall and f1-score are always at a very high level. In Table 4, the confusion matrix of one network was shown as an example: the distribution of misclassification errors is slightly different regarding testing and validation sets, while the GrTr and GrTrEx results tended to classify more actual thawed cases as frozen, and the testing results classified more actual frozen cases as thawed. But in general, all three sets showed very high accuracy, and misclassification errors are not overly skewed.

As indicated in Table 5, the results from the model using data collected from two different reference periods (Feb&Aug vs. JFebM&Aug) are close to each other with no substantial differences found. The testing results indicate that the performance of the data with a frozen reference period from mid-January to mid-March (JFebM&Aug) was slightly better than from February (Feb&Aug), while the Feb&Aug set achieved even a higher accuracy and f1-score regarding validation results. These results, with high modeling accuracy, confirm that the concept of reference periods is an efficient solution for data preparation in areas with limited measurements and works well for FT state classification. A short interval from the center of a season is sufficient as a reference. The validation results with samples from the whole year confirm that the network's learning works for the entire targeted time span, not only for specific seen periods. For future studies in the same area, the combination of Feb&Aug can be a good choice as the reference periods. Generally, the tolerance for the length of the choosen reference period is high. If there is sufficient data available, longer time periods can be used as references for FT states classification, which is, however, not indispensable for achieving good performance.

[1] : All settings of the deep learning method were implemented using PyTorch and executed on a PC equipped with a single GPU, specifically the RTX 2070, for both training and testing.

4.3. Validation with ground truth data

Overall, the well-performing networks were confirmed by the validation sets. The best-performing network reached a validation accuracy of 87.28 % with the GrTr set and over 90 % with the GrTrEx set. During the experiments, the validation sets showed their necessity and importance. Especially during the model optimization process, the validation helped to eliminate the networks with false high accuracy. Fig. 6 shows the overall training, testing, and validation accuracy from 12 example networks. The networks 4, 5 and 6, for example, achieved very high accuracies during training and testing processes, with network 4 even achieved a training accuracy higher than 99 %. However, these were further overturned by validation results from both sets. Also, in the positive validated models, the validation results conveyed novel information than in the testing results alone.

The comparison between the testing and validation sets revealed that the GrTr and GrTrEx sets showed similar results, with the GrTrEx set showing larger fluctuations than the GrTr set. As shown in Fig. 6, while the training and testing accuracies achieved by the same network are mostly close to each other, the results from the validation with GrTr and GrTrEx also show a similar trend. The GrTr accuracies of networks 1, 4, 5, 6, and 11 showed relatively large differences comparing to their training and testing accuracies, with the GrTrEx accuracies dropping to a low level. On the other hand, networks 2, 3, 7, 8, 9, and 10 had GrTr accuracies closer to their training and testing accuracies, and the GrTrEx accuracies rise correspondingly as GrTr accuracies rise. Overall, the validation results of GrTrEx corresponded well with the results with GrTr and showed more sensitivity by making larger steps of changes.

These findings suggest that extending ground truth data to build a validation data set can be useful and feasible as a supplement for situations where there is not enough overlapping remote sensing and measured data to support the creation of a sufficient pure ground truth data set. This approach can provide a good supplement to build long time series of validation data or when many data gaps exist in the measurements, which is a common occurrence in the northern cryosphere.

5. Discussion

In total, the proposed method with the deep learning approach using remote sensing data performs well for the soil FT states classification task. The high overall accuracy achieved through training is stable throughout the study area and across the entire year, encompassing all seasons. The performance of the networks is confirmed through testing with an independent data set and validation with ground truth data collected from in-situ measurements. Our approach achieves better and more stable results compared to other existing methods. For instance, in the study by Shao and Zhang in 2020, the four most widely used near-surface soil FT detection algorithms were evaluated and showed a broad spectrum performance with the overall accuracy varying between 73.8 % and 86.2 %. Our networks outperform this range with testing accuracies of around 95 % and validation accuracies of up to 87.28 %. The study by Cohen et al. in 2021 reached the highest similarity of S1 retrieval to the air temperature by 94–99 % in boreal forest environments using a threshold-based algorithm, however, in one of their three study areas, the similarity was only 64 %. Also in our previous study (Chen et al., 2022), the accuracies of different test sites varied from around 70 % to over 90 %, and requested individual threshold optimization.

Furthermore, this high accuracy is achieved with high spatial resolution. Compared to the existing products, such as SMAP FT with the highest spatial resolution at the kilometer scale, our classification is processed on a meter scale and directly based on originally captured signals, not derived from coarser resolution data. However, the S1 data used for this study have lower temporal resolution comparing to some other widely used data, such as SMAP, especially in the high latitude

Table 4
Confusion matrix of an example network for the testing set, GrTr set and GrTrEx set Confusion Matrix for an example network applied to the Testing and both validation sets, which are Ground Truth (GrTr), and Ground Truth Extended (GrTrEx) sets. The metrics in the upper left part (Target) are shared among the remaining three subtables, where NPV represents negative predicted value, FPR represents false positive rate and ovrlAcc represents overall accuracy.

		Target		
		1	0	
Prediction	1	True Frozen	False Fr	Precision
	0	False Th	True Thawed	NPV
		Recall	1-FPR	ovrlAcc
				F-score

		Testing		
		F	T	%
Prediction	F	6417	237	96.44
	T	499	6297	92.66
	%	92.78	96.37	94.53
				94.58

		Ground Truth		
		F	T	%
Prediction	F	905	184	83.10
	T	43	535	92.56
	%	95.46	74.41	86.38

		Ground Truth extend		
		F	T	%
Prediction	F	1500	130	92.02
	T	42	1176	96.55
	%	97.28	90.05	93.96

Table 5
Accuracy of classification using various reference FT (frozen/thawed) time periods for different sets (Test, Ground Truth - validation set, Ground Truth Extended - extended validation set), where "Feb&Aug" corresponds to February and August, and "JFebM" denotes mid-January to mid-March.

Dataset	Test		Ground Truth		Ground Truth extend	
Ref. FT	Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score
FebAug	94.88	94.96	86.20	88.68	90.73	92.05
JFebMAug	95.67	96.77	84.10	87.23	89.82	91.34

region. This can result in limitations on the accurate identification of transitions between frozen and thawed seasons. In-situ measurements show that the transition usually takes a few days, and the soil temperature is close to 0 °C. However, there are cases where the transition process occurs quickly, with a large gradient of soil temperature change. Additionally, the soil temperature may repeatedly rise and drop below zero, leading to multiple changes between frozen and thawed states before reaching a stable period. In order to observe these phenomena accurately, a dense time series of data is needed. Although the S1 data has a large advantage in terms of spatial resolution, its 12-day repeat cycle per satellite (ascending or descending) can make it difficult to observe the exact temperature transition process. The unfortunate end of the Sentinel-1B satellite mission has further increased this difficulty. It would be ideal for establishing a database combining high spatial and temporal resolution to enhance the monitoring of FT cycles in the northern permafrost landscape. In future research, it would be beneficial to explore the integration of high spatial resolution radar data with high temporal resolution microwave data, such as fusing S1 data with SMAP data. This integration would enable more accurate monitoring to support a more comprehensive understanding of FT dynamics.

Moreover, this deep learning approach in this study allows for the convenient addition of various variables to the model. This characteristic helps combine different potential influence factors to increase the quality of retrievement. Especially with CNN, spatial patterns can be considered instead of learning from each pixel individually. The comparison between the pixel-based processing with MLP and patch-based processing with CNN shows that the latter adds significant value to solving the task. As an extension, this insight suggests that it could be interesting to further investigate the relationship between soil FT dynamic and land cover type on a patch basis, meaning to analyze the spatial patterns of land cover types over an area, rather than focusing solely on isolated points with individual land cover types. In this study, we used only land cover types to represent environmental factors and achieved high accuracies. This provides an effective working frame and a line of thought to improve the estimation of FT states. Future studies can use this method to test more possible variables, such as climate, the thickness of ground ice, and their effects on FT classification.

Within the proposed framework, we have overcome several difficulties to make the procedure more convenient and repeatable. In the case of supervised model training, which is the focus of this study, it is crucial to have an adequate amount of data with accurate labeling.

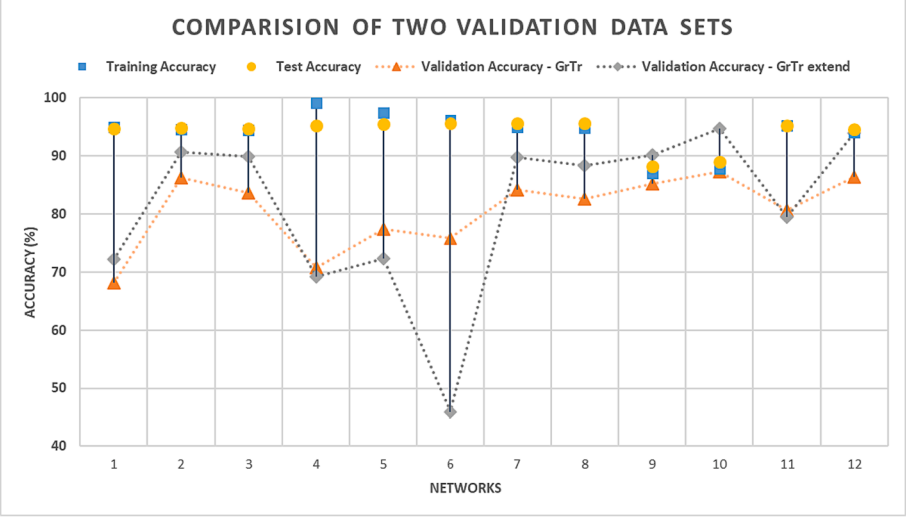


Fig. 6. Comparison of the validation accuracy of the GrTr and GrTrEx set by various condition (12 example networks with same input variables, differentiated by hyperparameter tuning).

Unfortunately, fulfilling this requirement is often challenging in cryospheric studies due to the difficulties involved in data collection. The concept of reference periods proposed in this study provides a much more convenient and feasible way to label the data. This approach enables the utilization of remote sensing data to build a training dataset by simply requiring knowledge of the seasonality in the area of interest, without the need for direct measurements for every sample. Consequently, this allows further data preparation that is not dependent on the amount and locations of measurement stations. The size of data is therefore not limited to the used amount in this study, users can freely create larger or smaller data sets after defining their needs. Moreover, the idea of extending the ground truth dataset helps create a larger validation data set based on the available in-situ measurements when it is challenging to obtain sufficient in-situ measurements. This strategy enables researchers to evaluate the accuracy of the model over a broader range of conditions, improving the reliability of the results. Overall, these advancements make the approach proposed in this study applicable and transferable for future studies over large regions by overcoming the challenges associated with limited in-situ measurements.

As a whole, this study found an effective way to classify soil FT states using the DL method and remote sensing data. From this study, we successfully tested that DL is suitable for this task and achieved high accuracy with high spatial resolution compared to existing methods. However, it is important to acknowledge that DL is a result-oriented black box method, which cannot directly provide specific feedback on soil temperature regimes and the mechanisms of FT processes. Studying these mechanisms is meaningful and indispensable for better understanding the dynamics of the permafrost landscape. Thus, we should consider DL as an efficient tool and make rational use of its strengths, but not unthinkingly rely on it and ignore the exploration of physical mechanisms. A successful retrieval achieved through the DL method could build a good foundation for further investigation of soil FT dynamic and its relation with the found influence factors, such as the land cover types.

6. Conclusion

The deep learning approach can be effectively used for FT estimation with remote sensing data. Both the pixel-wise training with MLP and the patch-wise training with CNN can simulate the FT states over the entire study area, while CNN showed significantly better performance with a test accuracy close to 95 % and a validations accuracy of over 88 %. Combined use of backscatter signals from both VV and VH polarizations is recommended; the incidence angle and land cover information can also contribute well to the modeling. Elimination of land cover types, that are not covered by soil, such as infrastructure and waterbodies, is not necessary for the processing. The concept of using seasonal references is a good solution for labeling training data. Defining references should be done carefully based on an investigation of local meteorological conditions. The reference period does not necessarily have to be long, but it should be robust and representative. The validation process with ground truth data is very important to help distinguish true and false well-performing networks. An empirical extension of Ground Truth data can be used carefully if necessary. For future work, we would like to explore methods for integrating microwave data with advantages in temporal resolution for instance L-band SMAP data set and data with high spatial resolution such as S1 to construct a better database for further improving the monitoring and understanding of the landscape FT cycles.

CRediT authorship contribution statement

Yueli Chen: Conceptualization, Methodology, Software, Data curation, Writing – original draft, Writing – review & editing, Visualization, Investigation, Validation, Project administration. **Shile Li:** Methodology, Software, Formal analysis, Validation, Writing – review & editing. **Lingxiao Wang:** Reviewing. **Magdalena Mittermeier:** Reviewing. **Monique Bernier:** Supervision. **Ralf Ludwig:** Reviewing and Editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

We acknowledge the contribution made by ESA in making Sentinel data easily available and at no cost, and the contribution made by Google Earth Engine in providing platform for easily processing of data. We also thank the Centre d'Études Nordiques for logistics support and all of our INRS and LMU colleagues, who were involved in the field campaigns. Yueli Chen especially thanks the Chinese Scholarship Council for generous support throughout the study. This research publication was also supported by LMUexcellent in form of the LMU Completion Grant, funded by the Federal Ministry of Education and Research (BMBF) and the Free State of Bavaria under the Excellence Strategy of the Federal Government and the Länder.

References

Agarap, A.F., 2018. Deep learning using rectified linear units (relu). arXiv preprint arXiv: 1803.08375. https://doi.org/10.48550/arXiv.1803.08375.

European Space Agency. Sentinel-1 GRD. 2021. Accessed June 14, 2022. https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar.

Allard, M., Seguin, M.K., 1987. The Holocene evolution of permafrost near the tree line, on the eastern coast of Hudson Bay (northern Quebec). Can. J. Earth Sci. 24, 2206–2222. https://doi.org/10.1139/e87-209.

Allard, M., Fortier, R., Sarrazin, D., Calmels, F., Fortier, D., Chaumont, D., Savard, J.P., Tarussov, A., 2007. L'impact du Rechauffement Climatique Sur Les Aeroports du Nunavik: Caracteristiques du Pergelisol et Caracterisation des Processus de Degradation Des Pistes. Quebec, Canada, Ouranos https://doi.org/10.13140/2.1.3889.6328.

Allard, M., Lemay, M., Barrette, C., L'Hérault, E., Sarrazin, D., Bell, T., Doré, G., 2012. In: Permafrost and Climate Change in Nunavik and Nunatsiavut: Importance for Municipal and Transportation Infrastructures. Publisher:, Quebec, Canada, pp. 171–197.

Beck, I., Ludwig, R., Bernier, M., Lévesque, E., Boike, J., 2015. Assessing permafrost degradation and land cover changes (1986–2009) using remote sensing data over Umiujaq, sub-arctic Québec. Permafr. Periglac. Processes 26, 129–141. https://doi.org/10.1002/ppp.1839.

Bengio, Y., 2012. Practical recommendations for gradient-based training of deep architectures. Neural Networks: Tricks of the Trade: Second Edition 437–478. https://doi.org/10.48550/arXiv.1206.5533.

Bergstedt, H., Bartsch, A., Neureiter, A., Höfler, A., Widhalm, B., Pepin, N., Hjort, J., 2020. Deriving a frozen area fraction from Metop ASCAT backscatter based on Sentinel-1. IEEE Trans. Geosci. Remote Sens. 58 (9), 6008–6019. https://doi.org/10.1109/TGRS.2020.2967364.

Bernier, M., Ludwig, R., Poulin, J., Touati, C., Kalantari, P., Wang, L., Chen, Y., Jacome, A., Ratsimbazafy, T. (2019). Mesures de température et teneur en eau du sol au voisinage d'Umiujaq et du bassin versant de la rivière Sheldrake, v. 1.0 (2011-2018). Nordicana D46, doi.org/10.5885/45567CE-639242EA518841D0.

Bishop, C.M., 1995. Neural networks for pattern recognition. Oxford University Press.

Bishop, C.M., Nasrabadi, N.M., 2006. Pattern Recognition and Machine Learning Vol. 4, No. 4, p. 738)..

Botalb, A., Moinuddin, M., Al-Saggaf, U. M., & Ali, S. S. Contrasting convolutional neural network (CNN) with multi-layer perceptron (MLP) for big data analysis. In 2018 International conference on intelligent and advanced system (ICIAS) (pp. 1-5). IEEE. https://doi.org/10.1109/ICIAS.2018.8540626.

Brown, J., Hinkel, K.M., Nelson, F.E., 2000. The circumpolar active layer monitoring (CALM) program: Research designs and initial results. Polar Geogr. 24, 166–258. https://doi.org/10.1080/10889370009377698.

Burgess, M.M.; Smith, S.L.; Brown, J.; Romanovsky, V.; Hinkel, K. Global Terrestrial Network for Permafrost (GTNet-P): Permafrost Monitoring Contributing to Global Climate Observations: Current Research Report 2000-E14; Geological Survey of Canada: Ottawa, ON, Canada, 2000; Available online: http://dsp-psd.pwgsc.gc.ca (accessed on 10 June 2022).

Calmels, F., Allard, M., Delisle, G., 2008. Development and decay of a lithalsa in Northern Quebec: A geomorphological history. Geomorphology 97, 287–299. https://doi.org/10.1016/j.geomorph.2007.08.013.
Centre for Northern Studies. NordicanaD. https://nordicana.cen.ulaval.ca/. (accessed on 10 June 2022).
Chen, Y., Wang, L., Bernier, M., Ludwig, R., 2022. Retrieving freeze/thaw cycles using sentinel-1 data in Eastern Nunavik (Québec, Canada). Remote Sens. (Basel) 14 (3), 802. https://doi.org/10.3390/rs14030802.
Cohen, J., Rautiainen, K., Lemmetyinen, J., Smolander, T., Vehviläinen, J., Pulliainen, J., 2021. Sentinel-1 based soil freeze/thaw estimation in boreal forest environments. Remote Sens. Environ. 254, 112267 https://doi.org/10.1016/j.rse.2020.112267.
Colliander, A., Mcdonald, K., Zimmermann, R., Schroeder, R., Kimball, J.S., Njoku, E.G., 2012. Application of QuikSCAT backscatter to SMAP validation planning: Freeze/thaw state over ALECTRA sites in Alaska from 2000 to 2007. IEEE Trans. Geosci. Remote Sens. 50, 461–468. https://doi.org/10.1109/TGRS.2011.2174368.
Davitt, A., Schumann, G., Forgotson, C., McDonald, K.C., 2019. The utility of SMAP soil moisture and freeze-thaw datasets as precursors to spring-melt flood conditions: A case study in the Red River of the North Basin. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12, 2848–2861. https://doi.org/10.1109/JSTARS.2019.2918947.
Derksen, C., Xu, X.L., Dunbar, R.S., Colliander, A., Kim, Y., Kimball, J.S., Black, T.A., Euskirchen, E., Langlois, A., Loranty, M.M., 2017. Retrieving landscape freeze/thaw state from Soil Moisture Active Passive (SMAP) radar and radiometer measurements. Remote Sens. Environ. 194, 48–62. https://doi.org/10.1016/j.rse.2017.03.007.
Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recogn. Lett. 27 (8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010.
Goodfellow, I., Bengio, Y., & Courville, A. Deep learning. 2016. MIT press. https://www.deeplearningbook.org.
Hachem, S., Allard, M., Duguay, C., 2009. Using the MODIS land surface temperature product for mapping permafrost: An application to Northern Quebec and Labrador. Canada. Permafr. Periglac. Processes 20, 407–416. https://doi.org/10.1002/ppp.672.
Han, M.L., Yang, K., Qin, J., Jin, R., Ma, Y.M., Wen, J., Chen, Y.Y., Zhao, L., Lazhu Tang, W.J., 2015. An algorithm based on the standard deviation of passive microwave brightness temperatures for monitoring soil surface freeze/thaw state on the Tibetan Plateau. IEEE Trans. Geosci. Remote Sens. 53, 2775–2783. https://doi.org/10.1109/TGRS.2014.2364823.
Jin, R., Li, X., Che, T., 2009. A decision tree algorithm for surface soil freeze/thaw classification over China using SSM/I brightness temperature. Remote Sens. Environ. 113, 2651–2660. https://doi.org/10.1016/j.rse.2009.08.003.
Jin, R., Zhang, T.J., Li, X., Yang, X.G., Ran, Y.H., 2015. Mapping surface soil freeze-thaw cycles in China based on SMMR and SSM/I brightness temperatures from 1978 to 2008. Arct. Antarct. Alp. Res. 47, 213–229. https://doi.org/10.1657/AAAR00C-13-304.
Johnston, J. M.; Houser, P. R. ; Maggioni, V.; Kim, R. S.; Vuyovich, C., 2022. Informing improvements in freeze/thaw state classification using subpixel temperature. IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1-19, 2022, Art no. 4301319, doi: 10.1109/TGRS.2021.3099292.
Judge, J., Galantowicz, J.F., England, A.W., Dahl, P., August 1997. Freeze/thaw classification for prairie soils using SSM/I radiobrightnesses. In Proceedings of the IEEE International Symposium on Geoscience and Remote Sensing, Singapore 3–8, 827–832. https://doi.org/10.1109/36.602525.
Kim, Y., Kimball, J.S., Mcdonald, K.C., Glassy, J., 2011. Developing a global data record of daily landscape freeze/thaw status using satellite passive microwave remote sensing. IEEE Trans. Geosci. Remote Sens. 49, 949–960. https://doi.org/10.1109/TGRS.2010.2070515.
Kim, Y., Kimball, J.S., Glassy, J., Du, J.Y., 2017. An extended global earth system data record on daily landscape freeze-thaw status determined from satellite passive microwave remote sensing. Earth Syst. Sci. Data 9, 133–147. https://doi.org/10.5194/essd-9-133-2017.
Kim, Y., Kimball, J.S., Xu, X., Dunbar, R.S., Colliander, A., Derksen, C., 2019. Global assessment of the SMAP freeze/thaw data record and regional applications for detecting spring onset and frost events. Remote Sens. 11, 1317. https://doi.org/10.3390/rs11111317.
Kou, X.K.; Jiang, L.M.; Yan, S.; Wang, J.; Gao, L.Y. Research on the improvement of passive microwave freezing and thawing discriminant algorithms for complicated surface conditions. In Proceedings of the IEEE International Symposium on Geoscience and Remote Sensing. Valencia, Spain, 22–27 July 2018; pp. 7161–7164. https://doi.org/10.1109/IGARSS.2018.8518731.
Krizhevsky, Alex; Sutskever, Ilya; Hinton, Geoffrey E. (2017-05-24). ImageNet classification with deep convolutional neural networks (PDF). Communications of the ACM. 60 (6): 84–90. https://doi/10.1145/3065386.
Kukačka, J., Golkov, V., & Cremers, D. (2017). Regularization for deep learning: A taxonomy. arXiv preprint arXiv:1710.10686. https://doi.org/10.48550/arXiv.1710.10686.
LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L. D., 1989. Backpropagation Applied to Handwritten Zip Code Recognition. Neural Comput. 1 (4), 541–551. https://doi.org/10.1162/neco.1989.1.4.541.
Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101. https://doi.org/10.48550/arXiv.1711.05101.
Lyu, H.B., McColl, K.A., Li, X.L., Derksen, C., Berg, A., Black, T.A., Euskirchen, E., Loranty, M., Pulliainen, J., Rautiainen, K., et al., 2018. Validation of the SMAP freeze/thaw product using categorical triple collocation. Remote Sens. Environ. 205, 329–337. https://doi.org/10.1016/j.rse.2017.12.007.
Naeimi, V., Paulik, C., Bartsch, A., Wagner, W., Kidd, R., Park, S., Elger, K., Boike, J., 2012. ASCAT Surface State Flag (SSF): Extracting Information on Surface Freeze/Thaw Conditions From Backscatter Data Using an Empirical Threshold-Analysis Algorithm. IEEE Trans. Geosci. Remote Sens. vol. 50, no. 7 https://doi.org/10.1109/TGRS.2011.2177667.
Nelson, F.E., Shiklomanov, N.I., Hinkel, K.M., Christiansen, H.H., 2004. The circumpolar active layer monitoring (CALM) workshop and the CALM II program. Polar Geogr. 28, 253–266. https://doi.org/10.1080/789610205.
O'Neill, P.E., Chan, S., Njoku, E.G., Jackson, T., Bindlish, R., Chaubell, J., 2019. SMAP L3 Radiometer Global Daily 36 km EASE-Grid Soil Moisture, Version 6. NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, Colorado USA.
Pelletier, M., Allard, M., Levesque, E., 2018. Ecosystem changes across a gradient of permafrost degradation in subarctic Québec (Tasiapik Valley, Nunavik, Canada). Arct. Sci. 5, 1–26. https://doi.org/10.1139/as-2016-0049.
Rautiainen, K., Lemmetyinen, J., Schwank, M., Kontu, A., Ménard, C.B., Mätzler, C., Drusch, M., Wiesmann, A., Ikonen, J., Pulliainen, J., 2014. Detection of soil freezing from L-band passive microwave observations. Remote Sens. Environ. 147, 206–218. https://doi.org/10.1016/j.rse.2014.03.007.
Rautiainen, K., Parkkinen, T., Lemmetyinen, J., Schwank, M., Wiesmann, A., Ikonen, J., Derksen, C., Davydov, S., Davydova, A., Boike, J., 2016. SMOS prototype algorithm for detecting autumn soil freezing. Remote Sens. Environ. 180, 346–360. https://doi.org/10.1016/j.rse.2016.01.012.
Rautiainen, K., Lemmetyinen, J., Aalto, T., Tsuruta, A., Kangasaho, V., Ikonen, J., Cohen, J., Kontu, A., Vehvildäinen, J., Pulliainen, J., July 2018. Smos retrievals of soil freezing and thawing and its applications. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain 22–27, 1463–1465. https://doi.org/10.1109/IGARSS. 2018.8519243.
Rignot, E., Way, J.B., 1994. Monitoring freeze—thaw cycles along North—South Alaskan transects using ERS-1 SAR. Remote Sens. Environ. 49 (2) https://doi.org/10.1016/0034-4257(94)90049-3.
Romanovsky, V.E., Smith, S.L., Christiansen, H.H., 2010. Permafrost thermal state in the polar Northern Hemisphere during the international polar year 2007–2009: A synthesis. Permafr. Periglac. Processes 21, 106–116. https://doi.org/10.1002/ppp.689.
Ropars, P., Lévesque, E., Boudreau, S., 2015. How do climate and topography influence the greening of the forest-tundra ecotone in northern Québec? A dendrochronological analysis of Betula glandulosa. J. Ecol. 103, 679–690. https://doi.org/10.1111/1365-2745.12394.
Roy, A.; Royer, A.; Derksen, C.; Brucker, L.; Langlois, A.; Mialon, A.; Kerr, Y.H. Evaluation of spaceborne L-band radiometer measurements for terrestrial freeze/thaw retrievals in Canada. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2015, 8, 4442–4459. https://doi.org/10.1109/JSTARS. 2015.2476358.
Roy, A., Toose, P., Williamson, M., Rowlandson, T., Derksen, C., Royer, A., Berg, A.A., Lemmetyinen, J., Arnold, L., 2017. Response of L-band brightness temperatures to freeze/thaw and snow dynamics in a prairie environment from ground-based radiometer measurements. Remote Sens. Environ. 191, 67–80. https://doi.org/10.1016/j.rse.2017.01.017.
Wang, L., Marzahn, P., Bernier, M., Ludwig, R., 2018a. Mapping permafrost landscape features using object-based image classification of multi-temporal SAR images. ISPRS J. Photogramm. Remote Sens. 141, 10–29. https://doi.org/10.1016/j.isprsjprs.2018.03.026.
Wang, L., Marzahn, P., Bernier, M., Ludwig, R., 2020. Sentinel-1 InSAR measurements of deformation over discontinuous permafrost terrain, Northern Quebec. Canada. Remote Sens. Environ. 248, 111965 https://doi.org/10.1016/j.rse.2020.111965.
Wang, P.K., Zhao, T.J., Shi, J.C., Hu, T.X., Roy, A., Qiu, Y.B., Lu, H., 2018b. Parameterization of the freeze/thaw discriminant function algorithm using dense in situ observation network data. Int. J. Digit. Earth 12, 980–994. https://doi.org/10.1080/17538947.2018.1452300.
Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., Wevers, J., Grosu, A., Paccini, A., Vergnaud, S., Cartus, O., Santoro, M., Fritz, S., Georgieva, I., Lesiv, M., Carter, S., Herold, M., Li, Linlin, Tsendbazar, N.E., Ramoino, F., Arino, O., 2021. ESA WorldCover 10 m 2020 v100. https://doi.org/10.5281/zenodo.5571936.
Zhang, T.J., Armstrong, R.L., 2001. Soil freeze/thaw cycles over snow-free land detected by passive microwave remote sensing. Geophys. Res. Lett. 28, 763–766. https://doi.org/10.1029/2000GL011952.
Zhang, G., Wang, C., Xu, B., & Grosse, R., 2018. Three mechanisms of weight decay regularization. arXiv preprint arXiv:1810.12281. https://doi.org/10.48550/arXiv.1810.12281.
Zhao, T.J., Zhang, L.X., Jiang, L.M., Zhao, S.J., Chai, L.N., Jin, R., 2011. A new soil freeze/thaw discriminant algorithm using AMSR-E passive microwave imagery. Hydrol. Process. 25, 1704–1716. https://doi.org/10.1002/hyp.7930.
Zhao, L., Zou, D., Hu, G., Wu, T., Du, E., Liu, G., Cheng, G., 2021. A synthesis dataset of permafrost thermal state for the Qinghai-Tibet (Xizang) Plateau. China. Earth System Science Data 13 (8), 4207–4218. https://doi.org/10.5194/essd-13-4207-2021.
Zhou, X., Zhang, Z., Shen, Q., Chen, Q., Liu, X., 2021. Identifying soil freeze/thaw states using scattering and coherence time series of high-resolution C-band synthetic aperture radar in the Qinghai-Tibet Plateau. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15, 519–532. https://doi.org/10.1109/JSTARS.2021.3137187.
Zuerndorfer, B.W., England, A.W., Dobson, M.C., Ulaby, F.T., 1990. Mapping freeze/thaw boundaries with SMMR data. Agric. for. Meteorol. 52, 199–225. https://doi.org/10.1016/0168-1923(90)90106-G.
Zuerndorfer, B., England, A.W., 1992. Radiobrightness decision criteria for freeze/thaw boundaries. IEEE Trans. Geosci. Remote Sens. 30, 89–102. https://doi.org/10.1109/36.124219.
Zwieback, S., Bartch, A., Melzer, T., Wagner, W. Probabilistic Fusion of K-u- and C-band scatterometer data for determining the freeze/thaw state. IEEE Trans. Geosci. Remote Sensing, 50(7), pp.2583-2594. https://doi.org/10.1109/TGRS.2011.2169076.