A Visualization Framework for Unsupervised Analysis of Latent Structures in SAR Image Time Series

Chandrabali Karmakar ⬤, Corneliu Octavian Dumitru, Nick Hughes, and Mihai Datcu ⬤

Abstract—**Openly available satellite image time series (SITS) are considered an important resource for spatiotemporal change monitoring. However, obtaining semantically annotated datasets for such tasks is an expensive affair. To alleviate this problem, this article presents a novel framework to model and understand the image dynamics by discovering latent information in Sentinel-1 SITS, even with limited ground truth data. The framework suggests how to use visualizations to efficiently integrate domain knowledge both for execution and evaluation of the machine-learning pipeline in the absence from ground truth data in SITS change studies. In a case study at a Polar region, we extend a limited amount of ground truth data and then discover its temporal evolution at image patch level, in an unsupervised manner. The trustworthiness of the framework is ensured by integration of domain knowledge and intelligent visual verification strategies. A visualization tool is also implemented for this purpose. The proposed framework contains two modules: a classifier and a change modeler. Our experiments show that a domain-knowledge-based classifier gives the best accuracy. The classifier semantically labeled the complete dataset of 24 study months, containing 153 600 patches with a size of 256×256 pixels by extending the available semantic labels from just three months. The temporal sequence of these sematic labels are then recorded and fed to a Bayesian model called Latent Dirichlet Allocation (LDA) to discover the underlying patterns. LDA generates a change map containing the dominant dynamic patterns to give a consolidated view of the evolution without having to browse the whole dataset. Further, color-coded change signatures explain the change classes.**

Index Terms—**Change maps, color-coded change signatures, domain knowledge, Latent Dirichlet Allocation (LDA), satellite image time series (SITS), unsupervised, visualization.**

I. INTRODUCTION

MOST regions on Earth undergo many transformations due to reasons like climate change, human activity, disasters, as well as geographical and geological processes. It is, therefore, of great importance for remote sensing (RS) researchers to monitor such changes and look into the phenomena causing such changes. The foremost thing is to observe changes over a period of time and understand any underlying patterns. Satellite image time series (SITS) contains huge amount of information and is frequently used to identify trends and changes in an area, e.g., assessing forest health, crop monitoring, land cover changes, urban planning, etc. However, such a task becomes more challenging in hard-to-access areas of the Earth, for example, the Polar regions, due to lack of ground truth data. Ironically, the Polar regions, especially the Arctic, are critically important areas to study surface cover changes due to their severe impact on the global climate [1]. However, collecting labeled datasets for any task, including that of SITS change analysis, is an expensive affair. Therefore, many researches aim to alleviate the problem of manual labeling by implementing semi-supervised or unsupervised machine-learning (ML) methods of semantic annotation. In particular, active learning has shown promising results in reducing human effort in labeling important image datasets. In this article, we focus on implementing an optimal framework for surface cover change analysis for areas where only a limited amount of domain knowledge is available, and contribute to a case study concerning change analysis of sea-ice cover in Greenland. In fact, we extend some limited amount of domain knowledge obtained from active learning research [2] to create semantically annotated datasets for a longer period of time and, thereafter, propose a completely unsupervised method for analyzing the temporal evolution of the changes. We use SAR (synthetic aperture radar) images. Among others, weather-independence and day—night coverage are the advantages of using SAR data.

In essence, our objective is to propose a simple yet robust framework to analyze the dynamics in SITS, considering some important issues not being addressed by most researchers in the area, which are as follows.

1) The framework for change analysis should work even in areas with limited ground truth data.
2) It should be possible to manually verify the intermediate results and/or integrate a domain expert's knowledge into the results from each module of the framework to detect anomalies.
3) Any highly resource-consuming method should be avoided to allow practical applications.
4) A single consolidated change map for the entire study period is preferred to several change maps resulting from comparison between consecutive study years.

Based on these objectives, we propose a framework that contains two modules: a classifier and a change modeler. After several experiments, a rule-based support vector machine (SVM) chi-squared kernel was chosen as the classifier to extend a set of domain knowledge rules from an active learning research [2]. More details on the classification is given in Sections III and IV. The postclassification change analysis was performed with LDA (Latent Dirichlet Allocation), a Bayesian model originally developed for textual data. A temporal sequence of semantic labels for each image patch was input to LDA, which discovers the underlying patterns in such sequences. We avoid the time-consuming dictionary creation process of traditional LDA by implementing a random dictionary assignment strategy based on [3]. As promised in the objectives, the outcomes from each of the two modules are illustrated by visualizations. To this end, we implemented a tool to visually verify the classification results. LDA generates a single change map with a consolidated view of the changes, segregated into a number of change classes, while explaining these classes with further class-specific visualizations, referred to as *change signature*.

We applied the proposed framework in a case study in the north-eastern part of Arctic Greenland for a study period of two years (from January 2018 to December 2019). Specifically, the two main results from the case study are as follows.

1) *A benchmark dataset obtained by forwarding a limited amount of domain knowledge*: The classifier unit of the proposed framework propagates some domain knowledge from [2] to predict the land cover classes for 24 months of study; thus, obtaining a benchmark dataset for the study area.

2) *Analysis of surface cover dynamics in SITS of 2018 and 2019 supported by visualizations*: Following our objectives, a single change map is generated to present the surface-cover dynamics of the sea-ice classes. The change classes in the map are explained with 10 color-coded spatiotemporal matrices showing the occurrence of six semantics classes, namely, *Glacier, Marine glacier, Mountains, Old ice, First-year ice, and Young ice*, and tally with a classification map of the area. We also gathered a domain expert's view on our findings. Interestingly, one of our visualizations helped the domain exert detect an anomalous appearance of a specific semantic class during the autumn months.

The rest of the article is organized as follows: Section II positions this article in the state of the art, Section III presents the framework, the dataset for the case study is described in Section IV, Section V details the methodology being used for semantic classification and SITS evolution analysis with LDA, Section VI presents the experimental results, and Section VII discusses the novelty of the article and finally Section VIII concludes the article.

II. RELATED STUDY

The availability of frequent revisit observations by satellites has opened a direction in research with SITS. Due to their size and complexity, it is extremely difficult for human users to find changes just by browsing through the time-indexed satellite image dataset. Therefore, researchers have devised various automatic methods to extract temporal information. The extraction of temporal information can be performed at pixel or image patch level, while the change modeling algorithm may work in a supervised or unsupervised manner. Pixel-based change modeling approaches are differed from image patch-based techniques. The difference lies in applications, running time, and outcomes. Pixel-based techniques are usually aimed to detect changes and result in a per-pixel change or no change outcome. Pixel-based algorithms may take longer to execute but provide more accurate results; however, considering the loss of contextual information, it can be a riskier option compared to its patch-based counterpart. Regarding the algorithmic point of view, we already mentioned that this can be a supervised spatiotemporal change modeling or an unsupervised one. By supervised we mean that the change to look for is specified in the beginning, e.g., biomass changes; this requires knowledge of the dataset, and more often used to see abrupt changes like natural disasters. On the other hand, in a completely unknown dataset, spatiotemporal evolution analysis has to be addressed as a data mining problem. For example, Julea et al. [4] proposed a novel method for identifying evolutions and subevolutions at pixel level. These researchers monitored changes at pixel level by considering an SITS as a set of symbolic and temporal sequences, and found grouped frequent sequential patterns, a data mining pattern defined by them. The method finds groups of pixels in the SITS that might be of interest to the end users.

A. Integration of Domain Knowledge

Next, we want to draw the attention of the research to another important aspect in earth observation studies: domain knowledge. Domain knowledge can be understood as subject-specific knowledge and understanding of the essential aspects of a specific field of inquiry and is considered essential for any ML task. The importance of integrating appropriate and adequate domain knowledge in various stages of an ML pipeline has been discussed by a number of researchers [5], [6]. While recent literature of explainable ML emphasizes the role of domain knowledge in enhancing model trustworthiness; especially, in data mining tasks like ours, where the model learns directly from the data, it is important to guide the learning of such unsupervised models [7]. Conventionally, the use of domain knowledge in earth observation studies has been rather limited to validation. For example, Datcu et al. [8] had a similar objective to discover latent structures in SITS but used domain knowledge only to verify the results at the end of the workflow, whereas we do a continual application of it to enhance trustworthiness of our approach.

However, this task of integrating domain knowledge becomes challenging when it comes to explain the outcomes to the domain experts and users not aware of the ML model used.

B. Visualization as Gateway of Domain Knowledge

Visualizations come to the rescue here. Spatiotemporal change studies have often used visualizations to pictorially

present the dynamics in the study area which are commonly referred to as change map. Depending on the ML model used, researchers have used visualizations ranging from graphs [4] to change map times series [5]. However, the importance of having a consolidated evolution map for Earth observation applications is evident. Our approach stresses upon creating a single evolution map showing the evolution dynamics and novel visualization techniques to further explain the patterns of evolution found by the change modeling algorithm. Visualizations are, although, in general considered important in ML as gateways of human-algorithm interaction and/or visual verification and can be used to communicate with end-users and domain experts with limited knowledge of ML methods [5], they can be even more essential in case of unavailability of ground truth information. This is a commonly occurring problem in RS research, but surprisingly overlooked yet. In our approach, we propose some novel designs for space-time-aware visualizations verifiable by a domain expert. We argue that the scarcity of ground truth information in RS applications can be compensated with problem-specific visualizations and domain knowledge.

However, we demonstrate several potential entry points to apply available knowledge at every stage of our workflow and at some stages, support the idea with visualizations as follows.

1) *Active Learning:* We applied the method described in [2] to retrieve, group, and semantically label together the classes based on their similarity.
2) *Patch sampling strategy:* We selected only about 20% of the data to be labeled for computing the features for classification. We defined a particular size and moving pattern for a patch selection window based on the size of the scenes and our understanding of the classes.
3) *Visual verification of classification results:* We implemented a visualization tool to verify the results of classification against product quick-look and KL-divergence graph of the SITS patches. As per suggestions of our domain expert, an inspection strategy was established to catch any potential incoherence between the semantic label, backscatter image, and feature representation (in the KL-divergence map).
4) *Selection of number of evolution classes:* We input a particular number of evolution classes to our change modeler algorithm as per our understanding of the study area, whereas other researchers use a complete data-driven approach.
5) *Verification of the evolution patterns:* A domain expert visually verifies the evolution patterns and eventually finds some anomalies.

C. Scalability and Resource Requirements

Since spatiotemporal change studies find many applications in the industry, it is important to consider some implementation level details, for example, the execution time of the algorithm, resource requirements, scalability with data volume, and, least but not the least, the software. Julea et al. [4] presented some impressive results by adapting a novel unsupervised method to the respective spatiotemporal context. However, not much has been commented about the scalability of their algorithm or the

aspect of visualizations. Although the authors have mentioned that their experiments work for 20 scenes of 100 000 pixels, we argue that it is less than what we can expect in real scenarios, e.g., like ours. We have 24 scenes of 426 250 066 pixels on average, and still plan to model more scenes. Also, the authors of [4] evaluated the results with ground truth data, which vividly differs from situations where no ground truth information is available. On the other hand, the problem of change analysis is addressed by [8] with a Bayesian model, namely, LDA. LDA was applied to four similarity measures calculated on consecutive pairs of multitemporal images. Hence, from a single SITS, one can generate four change map time series based on the computation of four similarity measures between every pair of images. In our opinion, this can be prohibitively resource-consuming in real-world applications. In essence, we recommend more experiments as well as methodic improvements to tackle the scalability issue. To set an example for research, we have developed a random dictionary assignment approach for fast creation of a set of features for semantic classification, also the problem of high memory requirements of classical *k-means* algorithm is handled by this.

D. Why Not Deep Learning?

Deep-learning algorithms are being successfully used in spatiotemporal evolution analysis. However, rather than blindly applying state-of-the art deep-learning frameworks, we recommend considering some practical issues as follows.

1) *Black-boxedness:* Deep-learning algorithms are often criticized to be not explainable and the results are not reproducible [6]. Therefore, it is often necessary to run a surrogate model to prove model's trustworthiness, which is extremely resource consuming. With this modern-day demand for explainable methods in artificial intelligence applications [6], we find it crucial to state that our framework is end-to-end explainable. The end-to-end explainability we mean the following: 1) Each constituent model is intrinsically explainable and computation follow the principles of explainability [6]. Our chosen feature representation (BoW—bag-of-words [3]), classifier (SVM) and change modeler algorithm are demonstrated to be explainable [9]. 2) Model interpretability is ensured with the help of visualizations. 3) Integration of domain knowledge support trustworthiness of intermediate and final outcomes.
2) *Lack of sufficient amount of labeled data:* Most deep-learning models need a huge amount of labeled data to perform well, which is often a hindrance in use cases involving areas (e.g., Polar areas).
3) *Discovery of latent structures:* In our approach, LDA observes the temporal sequence of surface cover classes and discover hidden structures called "topics," which is distinguished from distance-metric-based clustering approaches as followed by [10].

III. FRAMEWORK

Our framework contains two modules (see Fig. 1): a classification module and a change modeler. The detailed proposed
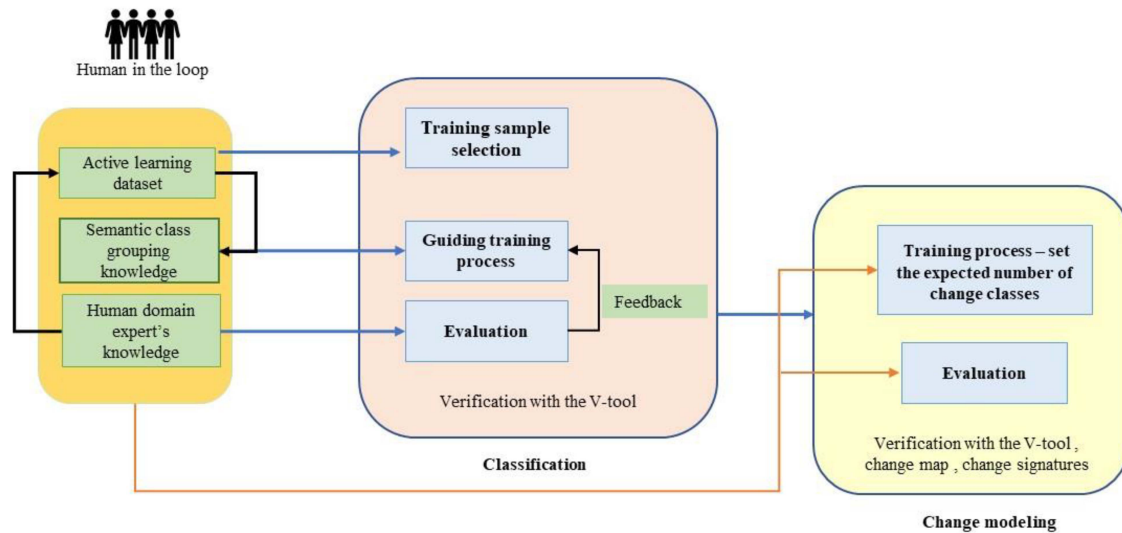
Fig. 1. Interplay and integration of domain knowledge in the proposed framework.
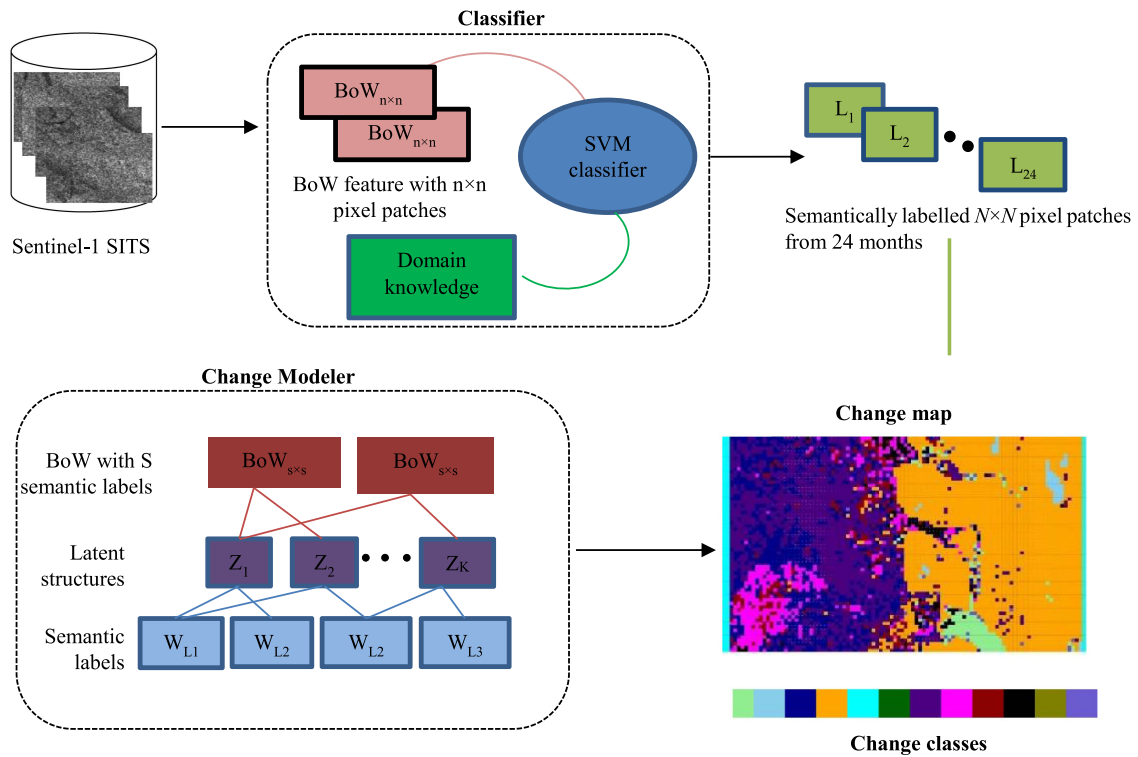

Fig. 2. Proposed framework contains two modules: the classifier and the change modeler.

mechanism of the framework is shown in Fig. 2. As per our proposed framework, the first task is to classify the SITS. As per our case study with SAR SITS, we look for best methods to classify SAR images. Literature show many efforts toward sea-ice classification using SAR data. A neural-network-based classifier was proposed in [11] to characterize SAR polarimetric data acquired by the ALOS-2, RADARSAT-2, and TerraSAR-X/TanDEM-X satellites. However, we chose openly available Sentinel-1 data. Sentinel-1 data were already successfully used in several sea-ice studies. For example, Tan et al. [12] made a semi-automated

segmentation of Sentinel-1 imagery, and evaluated the results with an open water segmentation approach. The workflow gave 95% accuracy in potential water identification. Among the many classification methods used by researchers, the ones based on SVMs are found to be widely accepted [13]. Also, a novel method based on SVM classifiers was proposed in [14] for the extraction of sea-ice cover from Sentinel-1 images. Motivated by these approaches, we tested several types of SVM kernels to classify our 24 scenes, together with three scenes from an active learning research project [2] used for training the classifier. We

found the best results with a rule-based, chi-squared kernel by using BoW features.

Compared to the abovementioned examples, our classification approach needs much less training examples by proposing an optimal method to use existing domain knowledge and also eliminates the need of preprocessing the SITS. We computed BoW features computed directly from the pixel brightness values in the scenes; this avoids any bias caused by low-level feature computation. The classifier uses an interclass relationship found in [9] as a source of domain knowledge. Also, by using three semantically labeled scenes from an active learning research [2] and by creating semantic labels for our 24 scenes, we actually proposed a method to extend the limited amount of available domain knowledge for applications with larger datasets.

In our case, LDA was chosen to model the spatiotemporal changes considering its reasonable resource requirements and explainability properties [9]. The proposed framework is presented in Fig. 2. The classifier module first tiles the SITS containing S scenes into $N \times N$-pixel patches, containing M ($M = (N \times N)/(n \times n)$) $n \times n$-pixel patches represented as BoW features (BoW$n \times n$). Next, the chi-squared SVM classifier semantically labels every $N \times N$ pixel patch with S unique labels. This gives a sequence of 24 labels. The change modeler represents the labels of SITS as BoW of S semantic labels (BoW$_{S \times S}$) by handling each semantic label as a word (WL$_1$ …WL$_S$), and discovering C latent structures (Z_1 …Z_C) as change classes being used to create an evolution map with C classes.

IV. DATASET DESCRIPTION

The selection of the data considers three criteria as follows.
1) Not to have commercial satellite data (so that other users can also benefit from the data/results).
2) To cover our area of interest defined by the users in the ExtremeEarth project.
3) To have as many acquisitions as possible regardless of the prevailing weather conditions.

For this, we chose the European Copernicus Programme which provides free satellite data captured by Sentinel-1 and Sentinel-2 (which have rather high resolution when compared with other satellites).

The Sentinel-1 SAR data have the advantage of operating at wavelengths not impeded by thin cloud cover, or a lack of solar illumination, and the satellites can acquire data over large areas during day or night with nearly no weather condition restrictions. There are twin satellites (Sentinel-1A and 1B) with a repeat period of 12 days for each satellite, that means every 6 days there may be an acquisition by one of the satellites.

Please note that Sentinel-2A/2B multispectral (optical) data are affected by weather conditions (e.g., cloud cover) prevalent both in the polar regions, mid-latitude, and equatorial regions, and as an optical sensor using visible wavelengths, it is also unavailable during the polar night which at the mid-latitude of this article lasts from late October to mid-February.

The area of our investigation is one of the areas analyzed by the ExtremeEarth project. Based on [14], this area was affected by melting and a reduction of the ice volume in 2018 and 2019.

Fig. 3. Location of the Sentinel-1 satellite images projected onto Google Maps.

The dataset used covers the area of Belgica Bank in the northeast of Greenland (see the location in Fig. 3) which is an area of extensive fast land-locked ice, the Norske Øer Ice Barrier. This makes the area ideal for using SITS to monitoring seasonal variations of the ice cover and icebergs as these remain stationary for a large part of the year.

The prevailing conditions in the area for RS retrievals are already known from previous articles [15].

For this demonstration, we selected the Sentinel-1, level-1 Ground Range Detected "amplitude" data in interferometric wide swath mode with dual polarization (HH and HV for polar areas). The products are geocoded with a resolution of 20×22 m (range $\times$ azimuth) and a pixel spacing of 10×10 m.

Considering our requirements for the location and the Sentinel-1 parameters, we retrieved 122 products from the Sentinel hub [16] (the list of products is shown in Table I).

In this article, we selected 24 product-images, i.e., one image per month for a period of 2 years (2018 and 2019) thus covering the two seasonal cycles (winter and summer).

V. METHODOLOGY

The proposed framework contains two modules: the classifier module and the change modeler. In this section, we thoroughly discuss the components of this novel framework in detail. The first module is a classifier. Remember that our SITS cover the same area at different dates and every pixel in our image data is the reflectance intensity of the geographic zone that it represents. Here, the classifier is employed to map these intensity values into a known set of labels.

A. Classifying to Extend Existing Domain Knowledge

The classified SAR images obtained from Dumitru et al. [2] were inadequate for our evolution study. We needed a classifier to forward this knowledge to create a benchmark dataset of adequate size to apply a method for evolution analysis. Technically, the availability of the small classified dataset allows us to choose a supervised ML model for classification; however, there are other factors to be cautious of as follows.
1) We cannot blindly rely on the classifier as there are neither enough number of labeled images, nor there are ground

TABLE I
ACQUISITION DATES OF OUR SENTINEL-1 IMAGE PRODUCTS OVER THE AREA OF BELGICA BANK, GREENLAND

Acquisition dates							
05.01.2018	11.01.2018	17.01.2018	23.01.2018	29.01.2018	04.02.2018	10.02.2018	16.02.2018
22.02.2018	28.02.2018	06.03.2018	12.03.2018	18.03.2018	24.03.2018	30.03.2018	05.04.2018
11.04.2018	17.04.2018	23.04.2018	29.04.2018	05.05.2018	11.05.2018	17.05.2018	23.05.2018
29.05.2018	04.06.2018	10.06.2018	16.06.2018	22.06.2018	28.06.2018	04.07.2018	10.07.2018
16.07.2018	22.07.2018	28.07.2018	03.08.2018	09.08.2018	15.08.2018	21.08.2018	27.08.2018
02.09.2018	08.09.2018	14.09.2018	20.09.2018	26.09.2018	02.10.2018	08.10.2018	14.10.2018
20.10.2018	26.10.2018	01.11.2018	07.11.2018	13.11.2018	19.11.2018	25.11.2018	01.12.2018
07.12.2018	13.12.2018	19.12.2018	25.12.2018	31.12.2018	01.01.2019	06.01.2019	12.01.2019
18.01.2019	24.01.2019	30.01.2019	05.02.2019	11.02.2019	17.02.2019	23.02.2019	01.03.2019
07.03.2019	13.03.2019	19.03.2019	25.03.2019	31.03.2019	06.04.2019	12.04.2019	18.04.2019
24.04.0219	30.04.2019	06.05.2019	12.05.2019	18.05.2019	24.05.2019	30.05.2019	05.06.2019
11.06.2019	17.06.2019	23.06.2019	29.06.2019	05.07.2019	11.07.2019	17.07.2019	23.07.2019
29.07.2019	04.08.2019	10.08.2019	16.08.2019	22.08.2019	28.08.2019	03.09.2019	09.09.2019
15.09.2019	21.09.2019	27.09.2019	03.10.2019	09.10.2019	15.10.2019	21.10.2019	27.10.2019
02.11.2019	08.11.2019	14.11.2019	20.11.2019	26.11.2019	02.12.2019	08.12.2019	14.12.2019
20.12.2019	26.12.2019						

The selected samples for our investigations are shown in red.

truth for validating the classification results. As the performance of most supervised ML models suffer due to lack of a good quality training data to contain the variability of classes, we incorporated domain knowledge to select a good set training data.

2) In a little-known dataset like ours, selecting right feature is not easy. A better approach is to compute feature representation on the pixel intensity values [3].

3) Rather than trusting the classification accuracy metrics, a visual verification of the classification results would be necessary.

B. Selecting Training Data for Classification

Selecting good training samples is crucial for employing any supervised ML model. Researchers have explored various methods to select small yet general set of training set; however, selection of training data based on domain knowledge is often recommended. Therefore, we designed a moving window of patch selection that traverses the entire scene and samples patches from "important" parts the scene. Based on a domain expert's view, some important parts of every scene are marked beforehand. Selecting equal number of samples from the known classes also improves the quality of the training dataset.

C. Features for Classification

We used BoW [3] features directly with pixel intensity values for classifying the SAR SITS for two main reasons: 1) the simplicity of computation and 2) its explainable properties [3]. The BoW technique was proposed for video search [16] and since then has been successfully used for several tasks in RS like image annotation [17], image object classification [18], target detection [19], etc.

We use the BoW technique to retrieve the features for the semantic classification of SITS. BoW is a histogram-like representation that generates vectors specifying the number of times a word appears in a document, normalized with the number of words; this vector is a probability vector of words. The BoW technique relies on the creation of a dictionary of words. It is a common approach to use vector quantization in low-level feature space for learning the dictionary. Next, the higher level BoW features can be learnt based on this dictionary.

Researchers have commonly used *k-means* as the vector quantization method. There have been efforts to estimate the natural number of clusters which include rate-distortion analysis [20] to estimate the size of the dictionary in BoW approaches. In an application-specific environment, it is also possible and even recommended to follow the state of art to decide the size of the dictionary for BoW, due to the high computational demand of experiments concerned with clustering huge datasets.

Vector quantization to create a dictionary with the BOW method is usually made with clustering. *k-means* is a centroid-based clustering method, originated in signal processing, and aims to partition O observations into k clusters by assigning a cluster label to each observation based on its distance to the K clusters. In the present case of *k-means* clustering with the pixel brightness values, each observation can be a vectorized image patch of size $(n \times n)$ containing n^2 numeric values; here, n can be chosen based on previous articles or domain knowledge.

For a database of I images, the first task is to sample P patches of size $(n \times n)$ pixels. This can be done by dense or sparse sampling. The next step is to extract local descriptor vectors x^i_j i $(j = 12, \dots P)$ from all patches. The third step is to learn the dictionary with V words by clustering the local descriptor vectors. The dictionary $D = (d1, \dots, dV) \in R\ D \times V$. Each word in the dictionary is the center of a cluster.

Next, each local descriptor vector is assigned to the index of the nearest dictionary word. This assignment can be a "hard feature assignment" which assigns a single label. Hard assignment can formally be defined as

$$v(x) = \begin{cases} 1, & \text{if } k = \text{argminn}_i \|x - d_i\|^2 \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

The other kind of feature assignment is called "soft feature assignment," which assigns a local descriptor to multiple words

from the dictionary based on the proportion of distances from the words.

Once the local descriptors are assigned, the next and the last step of BoW is to sum-pool the descriptors for each of the I image, which is equivalent to a histogram computation in case of a hard feature assignment.

D. Reducing Resource Requirements: Scalability

Clustering algorithms are computationally expensive. *K-means* has a time complexity of O(KTN), where K is the desired number of clusters, T is the number of iterations, N is the number of inputs [21], and a space complexity O(N) which is also high for large datasets with a huge number of features. After its first appearance by Mac Queen in 1967, *K-means* was widely applied in data mining; to fit the need of ever-growing data volumes, many efforts have been made to improve the scalability of the algorithm. Wang and Su [22] proposed an improved *K-means* algorithm. Outliers are detected and removed before clustering; this not only improves the accuracy of the algorithm, but also the running time. The idea pivots around improving initial cluster selection for the *K-means* algorithm. *In our approach, instead of removing spurious samples or outliers, we focus on selecting a small number of meaningful data points to create the initial clusters.* As the worst-case complexity of *k-means* is O(n^2), where n is the number of samples, we can see that reducing the number of data points for *k-means* algorithm by d times reduces the time complexity by d^2 times, which is a huge gain.

E. Classification Algorithm and Integration of Domain Knowledge

SVM classifiers are known for their good performance in classifying images. SVM classifiers are denoted by the kind of similarity measure, known as kernel [23] used to separate the feature space. For BoW features (or any histogram), chi-squared kernels are known to be most suitable as the chi-square distance measure is used in discrete probability distribution [24].

Chi-squared distance between two BoW histograms x and y can be symmetrically computed as

$$\frac{1}{2} \sum \frac{(x_i - y_i)^2}{(x_i + y_i)}. \tag{2}$$

The chi-squared kernel takes this form of distance measure and is defined as

$$k(x,y) = \exp\left(-\gamma \sum_i \frac{(x[i] - y[i])^2}{x[i] + y[i]}\right). \tag{3}$$

However, we experimented with several kernel types and BoW dictionary sizes. The rational behind experimenting with BoW dictionary sizes is that a vaster dictionary would better capture diversity in the data. Although the tradeoff here is the resource consumption. After Experiments with four kernel types of SVM classifiers radial basis function, linear kernel, and chi-squared kernel [23], we finally discovered the idea of a cascade of rules-based binary classifiers that integrates a class-grouping knowledge discovered in a previous article with the same dataset.

F. Domain-Knowledge-Aware Rule-Based SVM Classifier With a Chi-Squared Kernel

An agreement with literature was found about the accuracy of chi-squared kernel with BoW features. In comparison with a multiclass SVM classifier with a chi-squared kernel, a domain rule-based classifier with the same kernel type gave better performance. The domain rule for the classifier was derived in a previous article with the same dataset. The researchers of [11] already analyzed the same region as used in this article and found a metric of similarity of the surface cover classes sufficient to compile a class grouping dendrogram. We used this grouping described by the dendrogram to apply a hierarchical binary classification method. This rule-based classifier enables the integration of domain knowledge in the classification method, which is an important finding of this approach. An example is given in the results section.

G. Evolution Structures in Semantically Labeled SAR SITS

While browsing the visible SITS, a human user is able to find and record evolution patterns for a small dataset. However, the same does not work for huge datasets and images made of invisible spectrum like the radar images. To this end, LDA is applied to discover hidden patterns in the semantically classified SAR SITS. LDA observes the sequence to semantic labels over the 24 study months and finds abstract classes called "topics." The idea is described in Fig. 2. The SITS containing N images which are tiled into small patches of size $n \times n$ pixels. The backscatter values stored in these patches are then retrieved and subjected to a modified *k-means* clustering (as described above) to create a BoW representation which we refer to as BoW$_{n \times n}$. The rule-based classifier fed this representation and generated dataset with S unique semantic label. The labels for each image time is then stacked in the order of observation time, which generates a sequence of 24 semantic labels. Further, these sequences gain are represented as BoW over the S semantic labels, and denoted as BoW$_{s \times s}$. LDA is applied to this BoW$_{s \times s}$ representation.

LDA originated as a model for text processing, hence, we provide an analogy for terms being used with LDA, such as *corpus*, *document*, *words*, *dictionary*, and *topics* as follows.

1) *Word:* A word w_n corresponds to the semantic label of a macropatch (see Table II) (this is distinct from the word described in the BoWs models used for classification, where a word is a cluster label assigned to each $n \times n$-pixel image patch though a *K-means* clustering).
2) *Dictionary*: The set of all words is called dictionary.
3) *Document*: A sequence of words w_n is considered as a document. $D = (w_1, w_2, \ldots w_n)$. Here, n ranges from 1 to 24.
4) *Corpus*: A corpus $C = \{D_1, \ldots, D_M\}$ is the image dataset.

The surface cover evolution of the areas labeled with these change classes are further visualized using change signatures.

Next, we discuss the work mechanism of LDA.

5) *Work Mechanisms of LDA*

LDA is a generative probabilistic model of a corpus. The basic idea is that documents are represented as random mixtures over

TABLE II
IDs OF SEMANTIC LABELS USED IN THIS ARTICLE

0	2	3	5	6	7	8	1-4-9-10
Black border	Glacier	Marine glacier	Mountains	Old ice	First-year ice	Young ice	Water group: A group combining four classes of the reference classification: water body (1), melted snow (4), water current (9), floating ice (10)

latent topics, where each topic is characterized by a distribution over words.

According to the original research in [25], LDA assumes the following generative process for each document D in corpus C.

Choose $N \sim \text{Poisson}(\xi)$.

Choose $\theta \sim \text{Dir}(\alpha)$.

For each of the N words w_n:

Choose a topic $z_n \sim \text{Multinomial}(\theta)$.

Choose a word w_n from $p(w_n|z_n, \beta)$, a multinomial probability conditioned on the topic z_n.

Given the parameters α and β, the joint distribution of a topic mixture θ, a set of N topics z, and a set of N words D is given by

$$p(\theta, z, D|\alpha, \beta) = p(\theta|\alpha) \prod_{n=1}^{N} p\left(\frac{z_n}{\theta}\right) p\left(\frac{w_n}{z_n, \beta}\right) \quad (4)$$

where the marginal probability of a document D is obtained as

$$p(D|\alpha, \beta) = \int p(\theta|\alpha) \left(\prod_{n=1}^{N} \sum_{z_n} p\left(\frac{z_n}{\theta}\right) p\left(\frac{w_n}{z_n, \beta}\right) \right) d\theta. \quad (5)$$

LDA works on a BoWs framework and finds topics as probability distributions of words and documents as distribution of topics. In our case, each document was assigned the maximum probable topic to obtain the final change analysis map.

H. Interpretable Visualizations: Change Map and Color-Coded Change Signatures

LDA discovers latent structures that represent classes of evolution, which we commonly mention as "change classes." To have a consolidated visual delineation of the change classes. As mentioned before, the generative process of LDA finds the maximum probable topic for each document, i.e., each image patch. We visually present each such "topic" with a unique color to cover the whole SITS. This method summaries 24 months of observation into a single visualization. This visualization works as a gateway of communication with the domain expert for verification of the findings.

The change classes in this map is further described with color-coded matrices of semantic labels that clearly depicts a relationship between the SITS input to LDA and the change class discovered against that input. The color-coded matrix is contextually called "change signature." The S semantic labels are denoted with S unique colors. For each change class, a matrix entry

$M_{i,j} \in S$, where $M_{i,j}$ denotes the surface cover semantic labels for the ith macropatch and jth month of study.

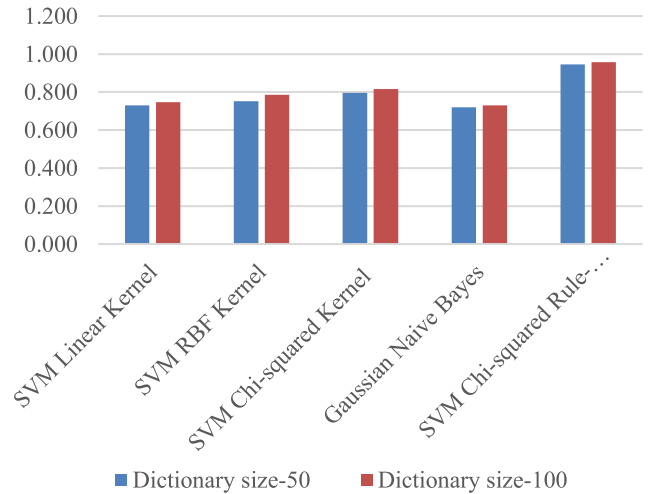Comparison of classifiers and dictionary size for SAR image classification

Fig. 4.   Comparison of classification accuracy.

VI. RESULTS

As mentioned before, our proposed framework contains two modules: classification and change modeling. The classification of SITS extends some limited amount of domain knowledge to create a larger set of domain knowledge; next, the change modeling algorithm discovers patterns of changes.

For the classification, we experimented with several classification algorithms and found that a domain-knowledge-aware rule-based SVM classifier gives the best accuracy.

A. Classification Results

We experimented with two dictionary sizes, 50 and 100 and four SVM kernel types—linear kernel, radial basis function, chi-squared kernel, and Gaussian Naïve Bayes kernel. Finally, the domain-knowledge-aware rule-based SVM kernel gave the best accuracy (Fig. 4).

1) Domain-Knowledge-Aware Rule-Based Classifier: We devised an algorithm involving class grouping knowledge and a group of binary SVM classifiers, which we refer to as a rule-based classifier. The classifier checks the interclass distance measure computed in [9] and predicts the semantic label for every sample. The pseudocode is presented in the appendix.

The binary classifiers were trained with 75% of the labeled samples from [7] *classifier 1, classifier 2, classifier 3, classifier 4, classifier 5, classifier 6,* and *classifier 7* achieving 100%, 91%, 99%, 91%, 93%, 96%, and 92% accuracy, respectively (Table III

TABLE III
CLASSIFIERS AND CLASS GROUPS

Classifier	Participating classes/class-groups	Accuracy (weighted average)
Classifier1	Black border vs. Group A	100%
Classifier2	Group A vs. Group B	91%
Classifier3	Glacier vs. Group C	99%
Classifier4	Young ice vs. Group E	91%
Classifier5	Group D vs. Group F	93%
Classifier6	Marine glacier vs. mountains	96%
Classifier7	Old ice vs. first-year ice	92%
Class groups	**Description**	
Group A	Floating ice, water current, water bodies, melted snow	
Group B	Glacier + Group C	
Group C	Young ice + Group E	
Group D	Marine glacier, mountains	
Group E	Marine glacier, mountain, old ice, first-year ice	
Group F	Old ice, first-year ice	

TABLE IV
SEMANTIC EVOLUTION OF MACROPATCH ID-81

2018		
Jan.–Apr.	**May.–Aug.**	**Sep.–Dec.**
Mountains	Mountains	Water group
Mountains	Water group	Mountains
Mountains	Water group	Mountains
Mountains	Water group	Mountains
2019		
Jan.–Apr.	**May–Aug.**	**Sep.–Dec.**
Mountains	Mountains	Water group
Mountains	Water group	Mountains
Mountains	Young ice,	Mountains
Mountains	Water group	Mountains

TABLE V
SEMANTIC EVOLUTION OF MACROPATCH ID-6071

2018		
Jan.–Apr.	**May–Aug.**	**Sep.–Dec.**
Marine glacier	Marine glacier	Marine glacier
Marine Glacier	Water group	Young ice Marine
marine Glacier	Water group Old	glacier Marine
marine glacier	ice	glacier
2019		
Jan.–Apr.	**May–Aug.**	**Sep.–Dec.**
Marine glacier	Marine glacier	Young ice Marine
Marine glacier	Water group	glacier Young ice
Marine glacier	First-year ice	Marine glacier
Young ice	First-year ice	

We can notice a change of surface cover classes during the summer months of both years; the mountain area is classified as *water group*, which is explained by the melting of snow in the summer months. We can also verify this fact with the quick-look image of the patch (Fig. 5). We see that the quick-look image appears darker when the patch is classified as water group or young ice. The readers may look into the other examples in Fig. 5 and observe similar behavior. We found that the surface cover classes like *mountains*, *glacier*, and *marine glacier* (mostly in the right half of the study area, roughly) show less changes than the sea cover on the other side. For macropatch ID 6071 (Fig. 5, column 7), which lies in the *marine glacier* area, we observe a sequence as in Table V. We again observe the appearance of the *water group* and the types of ice in the summer months. The quick-look images appear darker where the patch is classified as *water group*. Not only that, we see same semantic labels correspond to similar quick-look images, with an exception in Nov. 2019, where the patch has been classified as *young ice*, but the quick-look looks similar to the ones classified as *marine glacier*. This can be attributed to minor classification errors.

B. Results of Surface Cover Dynamics Analysis

The surface cover dynamics were analyzed using LDA, running over the semantic labels for 24 elements. The results from LDA were visualized using a change map and 10 change signatures.

1) Change Map: The change map depicted in Fig. 7(a) shows the evolution of the study area defined by change classes. The change classes are latent layers (called topics in LDA terminology) in the data discovered by LDA. *Areas labeled with the same color have undergone the same level of change, i.e.,* they belong to the same change class. LDA discovered 10 change classes. Each of the change classes is further described with change signatures in Fig. 7(b). The change signatures show identifying colors for each surface cover class. Each signature has 100 rows and 24 columns. The 100 rows show randomly selected 100 macropatches for the respective change class and the column numbers correspond to the month of observation. Thus, $S_{i,j}$ shows the surface cover semantic labels for the *i*th macropatch and *j*th month of study.

and Fig. 12 in the Appendix). The 24 scenes are tiled into patches of size of 256×256 pixels, referred as macropatches. Each such macropatch is considered as a sample whose semantic label is to be predicted.

This workflow creates a 24-element sequence of semantic labels. The results are explainable with respect to domain knowledge. For demonstration, we have chosen six locations (six macropatches designated by their respective macropatch IDs) in the study location as shown in Fig. 5; the time series of these locations are elaborated in Fig. 6. The captions below the quick-look images of the patches in Fig. 6 (vertically placed in order of observation months) list the sequence of semantic labels for each macropatch for 24 months in their order of time.

The first example is patch-ID 81 (see the first column in Fig. 5), which evolves into a sequence as in Table IV.
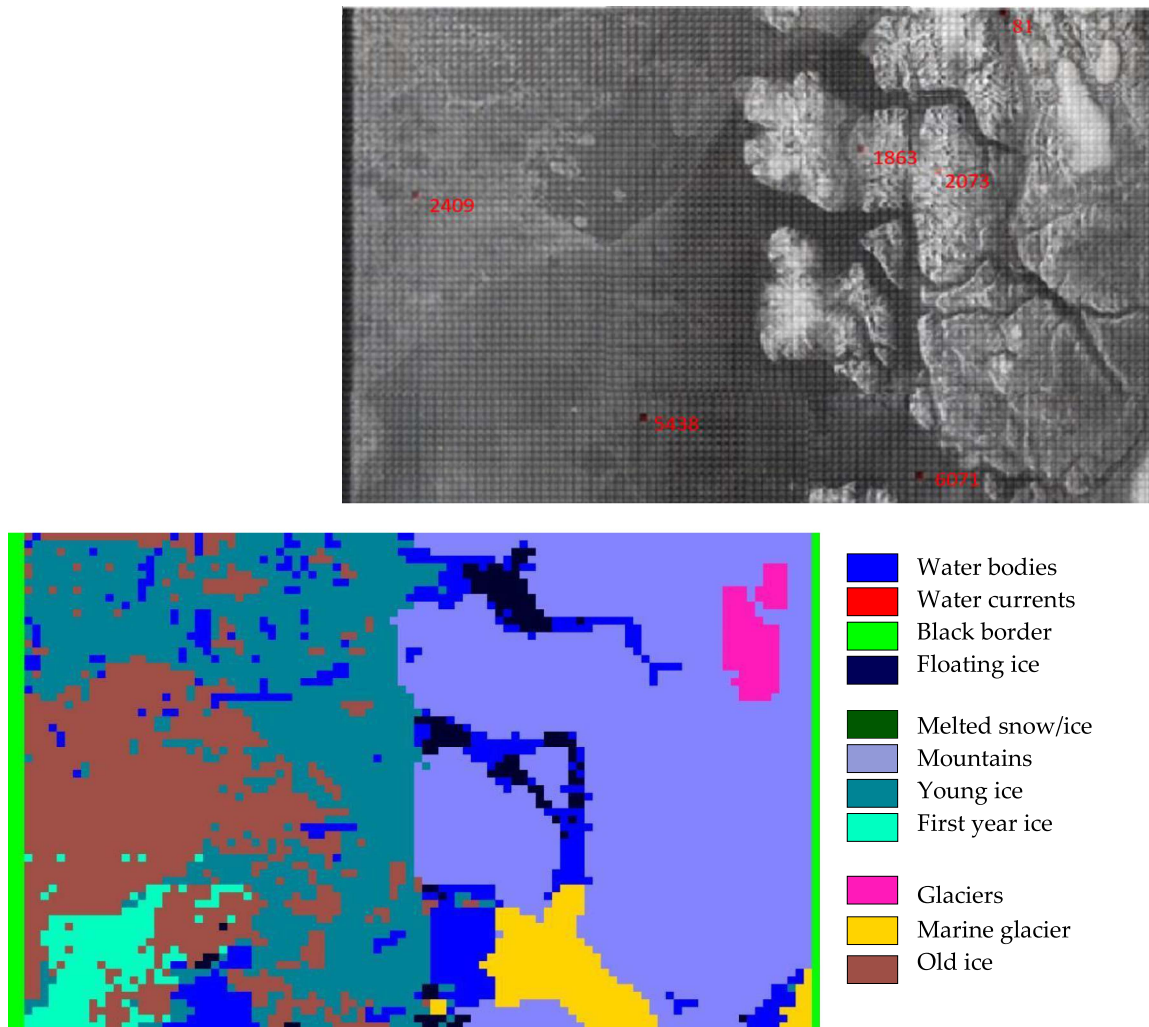
Fig. 5.   Selected patches to demonstrate our SITS classification. Locations are chosen based on classification map for April 2018 by [2]. Macropatch ID 81, ID 1863, and ID 2073 are in mountains area, macropatch ID 2409 and ID 5438 are old ice, and macropatch ID 6071 is in marine glacier.

2) Change Classes: The change map segregates a scene into distinguishable change regions by assigning identifying change class colors to the macropatches. In our case, LDA found 10 change classes (henceforth called as classes), we identified them as C.C-0, C.C-1, C.C-2, C.C-3, C.C-4, C.C-5, C.C-8, C.C-9, C.C-10, and C.C-11. Looking at the change map in Fig. 7, and comparing it with the quick-look image and the classification map from [2] in Fig. 6, we clearly see that the evolution of the entire *mountains* area falls into the class C.C-1. The dynamics of the *marine glacier* area is modeled by C.C-11, while for the other surface cover classes, no such one-to-one mapping is visible. This can be explained by the fact that the other surface cover classes went through more diverse changes that could not be distinctly modeled by one particular class. The size of each class in terms of number of macropatches is listed in Table VI.

3) Change Signatures: Change signatures, simply called *signatures* henceforth, are color-coded patterns showing the evolution of each change class in terms of surface cover classes over the study period. The amount of change in the surface cover classes is indicated by the homogeneity of the signature patterns. We plotted these patterns for each class by listing in color the sequence of the 24 semantic labels of the macropatches belonging to that particular class. There can be a maximum of eight unique labels in the 24-label sequences. These eight labels are being identified by eight unique colors as displayed in the colormap in Fig. 7(b). The 10 signatures are C.S-0 (for class C.C-0), C.S-1 (for class C.C-1), C.S-2 (for class C.C-2), C.S-3 (for class C.C-3), C.S-4 (for class C.C-4), C.S-5 (for class C.C-5), C.S-8 (for class C.C-8), C.S-9 (for class C.C-9), C.S-10 (for class C.C-10), and C.S-11 (for class C.C-11).

The 24-label sequences are plotted row-wise for each macropatch. In Fig. 7(b), we display only the first 100 macropatches and their 24-label sequences. The colors in the signatures are easy to follow and we can observe any prevalent pattern in the way of appearance of the surface cover classes in the change classes over the 24 months. An enlarged part from C.S-0 is shown Fig. 7(c), which shows the 24-label sequence of the 10th, 11th, and 12th macropatch of change class C.C-0.

Please note that no two signatures are exactly the same, and by similar we mean a) being composed of most of the same semantic labels in similar proportions and b) they follow a similar order of occurrence of the semantic label. Rather than this qualitative

Patch_ID Month	81	1863	2073	2409	5438	6071
Jan. 2018	month= 0 lbl= 5.0 Mountains	month= 0 lbl= 5.0 Mountains	month= 0 lbl= 5.0 Mountains	month= 0 lbl= 1491 Water group	month= 0 lbl= 6.0 Old ice	month= 0 lbl= 3.0 Marine glacier
Feb. 2018	month= 1 lbl= 5.0 Mountains	month= 1 lbl= 5.0 Mountains	month= 1 lbl= 5.0 Mountains	month= 1 lbl= 6.0 Old ice	month= 1 lbl= 6.0 Old ice	month= 1 lbl= 3.0 Marine glacier
Mar. 2018	month= 2 lbl= 5.0 Mountains	month= 2 lbl= 5.0 Mountains	month= 2 lbl= 5.0 Mountains	month= 2 lbl= 6.0 Old ice	month= 2 lbl= 6.0 Old ice	month= 2 lbl= 3.0 Marine glacier
Apr. 2018	month= 3 lbl= 5.0 Mountains	month= 3 lbl= 5.0 Mountains	month= 3 lbl= 5.0 Mountains	month= 3 lbl= 6.0 Old ice	month= 3 lbl= 6.0 Old ice	month= 3 lbl= 3.0 Marine glacier
May 2018	month= 4 lbl= 5.0 Mountains	month= 4 lbl= 5.0 Mountains	month= 4 lbl= 5.0 Mountains	month= 4 lbl= 7.0 First- year ice	month= 4 lbl= 7.0 First- year ice	month= 4 lbl= 3.0 Marine glacier
Jun. 2018	month= 5 lbl= 1491 Water group	month= 5 lbl= 5.0 Mountains	month= 5 lbl= 5.0 Mountains	month= 5 lbl= 6.0 Old ice	month= 5 lbl= 1491 Water group	month= 5 lbl= 1491 Water group
Jul. 2018	month= 6 lbl= 1491 Water group	month= 6 lbl= 1491 Water group	month= 6 lbl= 5.0 Mountains	month= 6 lbl= 7.0 First year ice	month= 6 lbl= 8.0 Young ice	month= 6 lbl= 1491 Water group
Aug. 2018	month= 7 lbl= 1491 Water group	month= 7 lbl= 1491 Water group	month= 7 lbl= 5.0 Mountains	month= 7 lbl= 8.0 Young ice	month= 7 lbl= 1491 Water group	month= 7 lbl= 6.0 Old ice
Sep. 2018	month= 8 lbl= 1491 Water group	month= 8 lbl= 5.0 Mountains	month= 8 lbl= 5.0 Mountains	month= 8 lbl= 8.0 Young ice	month= 8 lbl= 6.0 Old ice	month= 8 lbl= 3.0 Marine glacier
Oct. 2018	month= 9 lbl= 5.0 Mountains	month= 9 lbl= 5.0 Mountains	month= 9 lbl= 5.0 Mountains	month= 9 lbl= 6.0 Old ice	month= 9 lbl= 8.0 Young ice	month= 9 lbl= 8.0 Young ice
Nov. 2018	month= 10 lbl= 5.0 Mountains	month= 10 lbl= 5.0 Mountains	month= 10 lbl= 5.0 Mountains	month= 10 lbl= 7.0 First year ice	month= 10 lbl= 6.0 Old ice	month= 10 lbl= 3.0 Marine glacier
Dec. 2018	month= 11 lbl= 5.0 Mountains	month= 11 lbl= 5.0 Mountains	month= 11 lbl= 5.0 Mountains	month= 11 lbl= 6.0 Old ice	month= 11 lbl= 6.0 Old ice	month= 11 lbl= 3.0 Marine glacier

Fig. 6. Our SITS classification results for selected macropatches. The quick-look images for the respective patches are placed vertically, followed by their predicted semantic label.

description of similarity, a mathematical index can be computed to measure the similarity of signatures. However, we save that as a future activity. The signatures C.S-0, C.S-2, C.S-4, C.S-5, and C.S-8 are visibly similar, whereas C.S-1, C.S-9, and C.S-11 are different from all others.

C. Visualization Tool

A user-interface tool was developed to efficiently view and analyze changes in the area of each macropatch. Initially, the tool loads the quick-look images of the chosen month from the
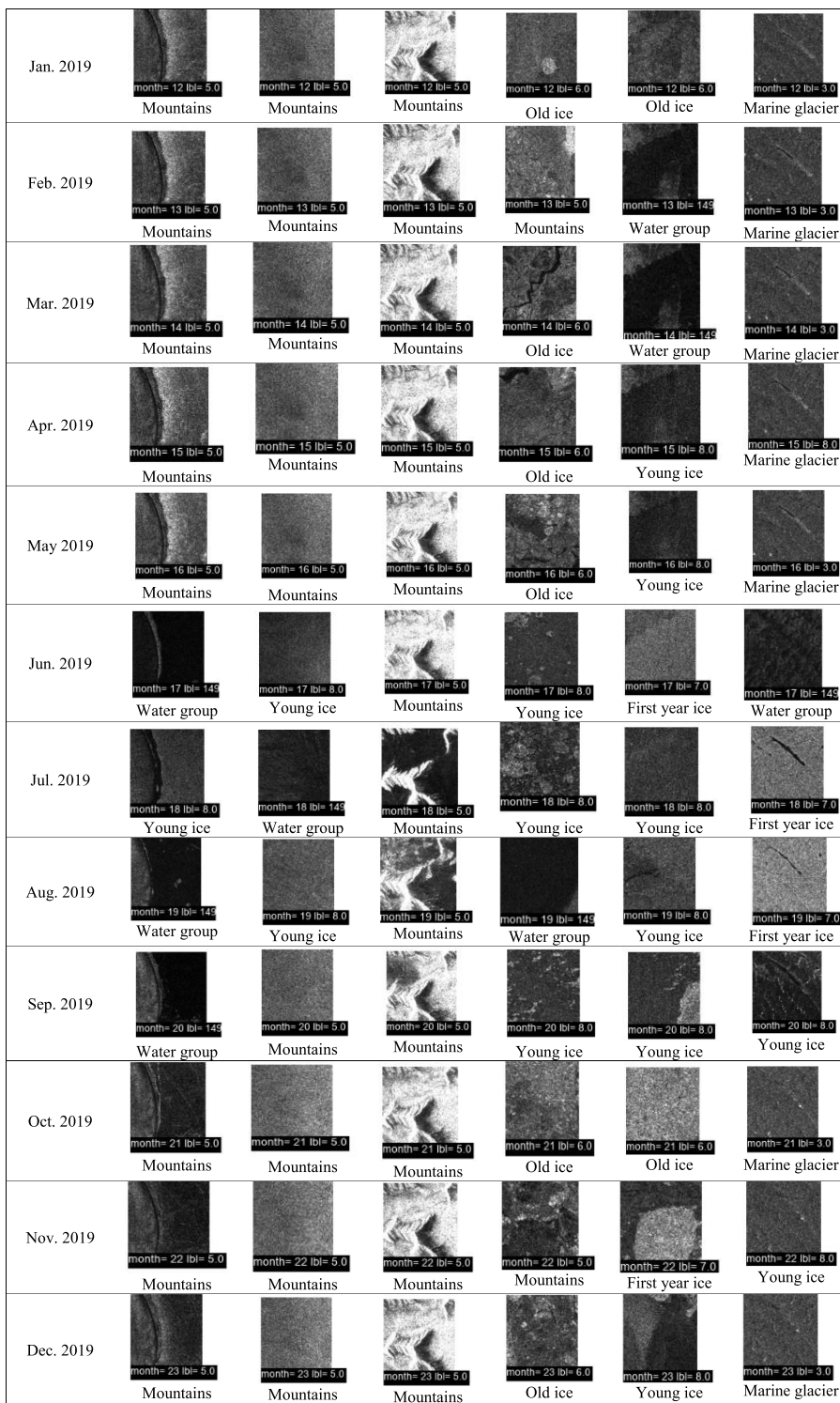
Fig. 6. *(Continued.)*

24 months of study, overlayed by 6400 clickable positions (the number of clickable positions is chosen to match the number of image patches, and can be set differently by users for a different application. This enables the user to click on any macropatch in the study area. After clicking, the tool displays the series of 24 quick look images in a larger scale for that specific macropatch, accompanied by the predicted label for each macropatch and a diagram showing the Kullback–Leiber (KL) divergence between consecutive months. As a result, it becomes very easy to follow a change in semantic label or a physical change in the view of the area shown in the macropatch time series. For a clear and efficient visualization of the changes in the study area over the study period, a database-backed visualization tool was implemented. The tool checks the macropatch size (256×256
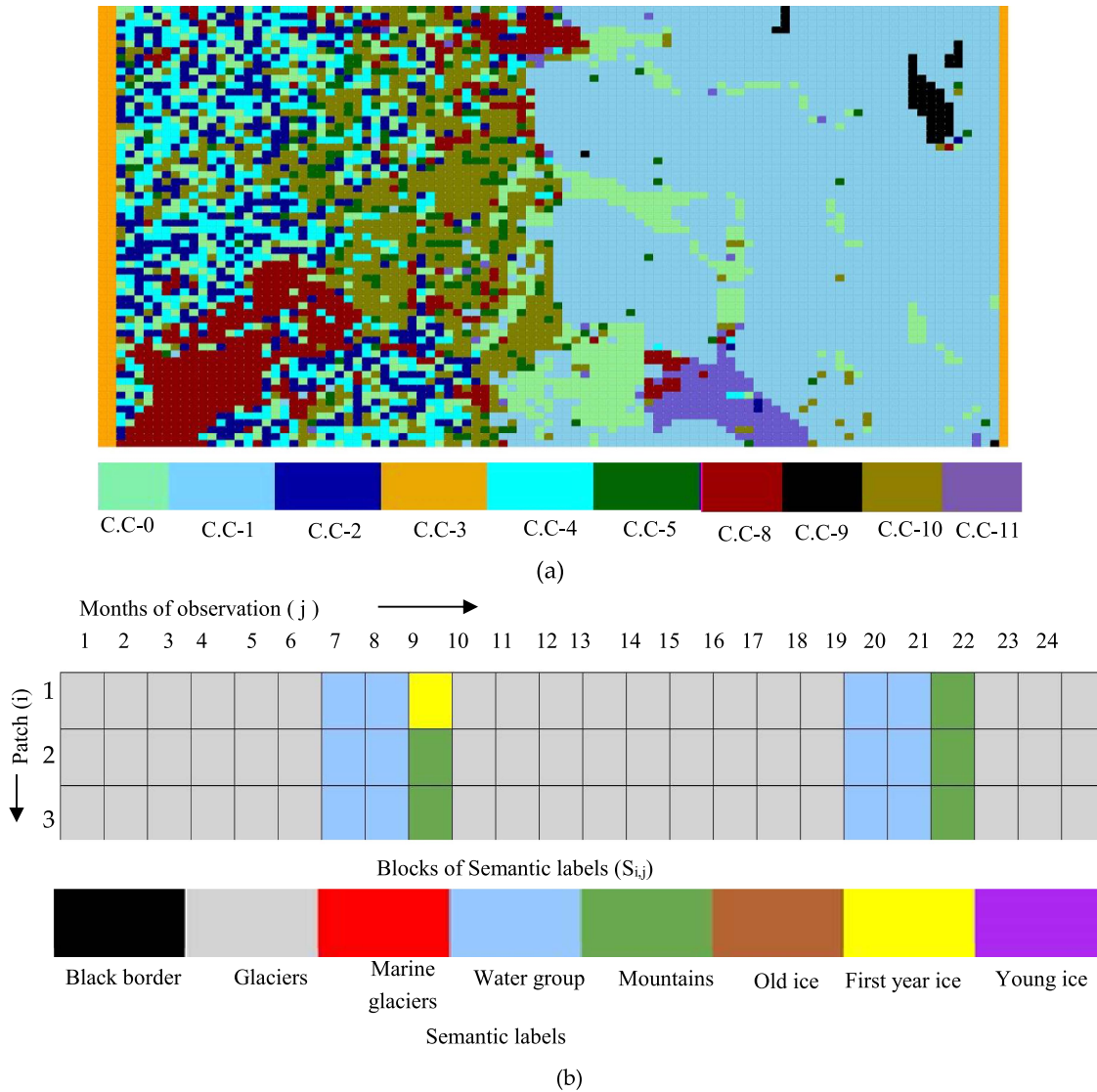
Fig. 7. (a) Change map with classes obtained from the LDA model (C.C). (b) Change signatures (C.S) explained, each block $S_{i,j}$ shows the predicted semantic label from month i and macropatch j (retrieved row-wise) for the respective change class.

TABLE VI
SIZE OF CHANGE CLASSES IN TERMS OF NUMBER OF MACROPATCHES

Change class	Number of macropatches (percent of the study area)
C.C-0	730 (11.4 %)
C.C-1	2645 (41.3%)
C.C-2	536 (0.08%)
C.C-3	192 (0.03%)
C.C-4	535 (0.08%)
C.C-5	205 (0.03%)
C.C-8	542 (0.08%)
C.C-9	49 (0.007%)
C.C-10	820 (12.8%)
C.C-11	146 (0.02%)
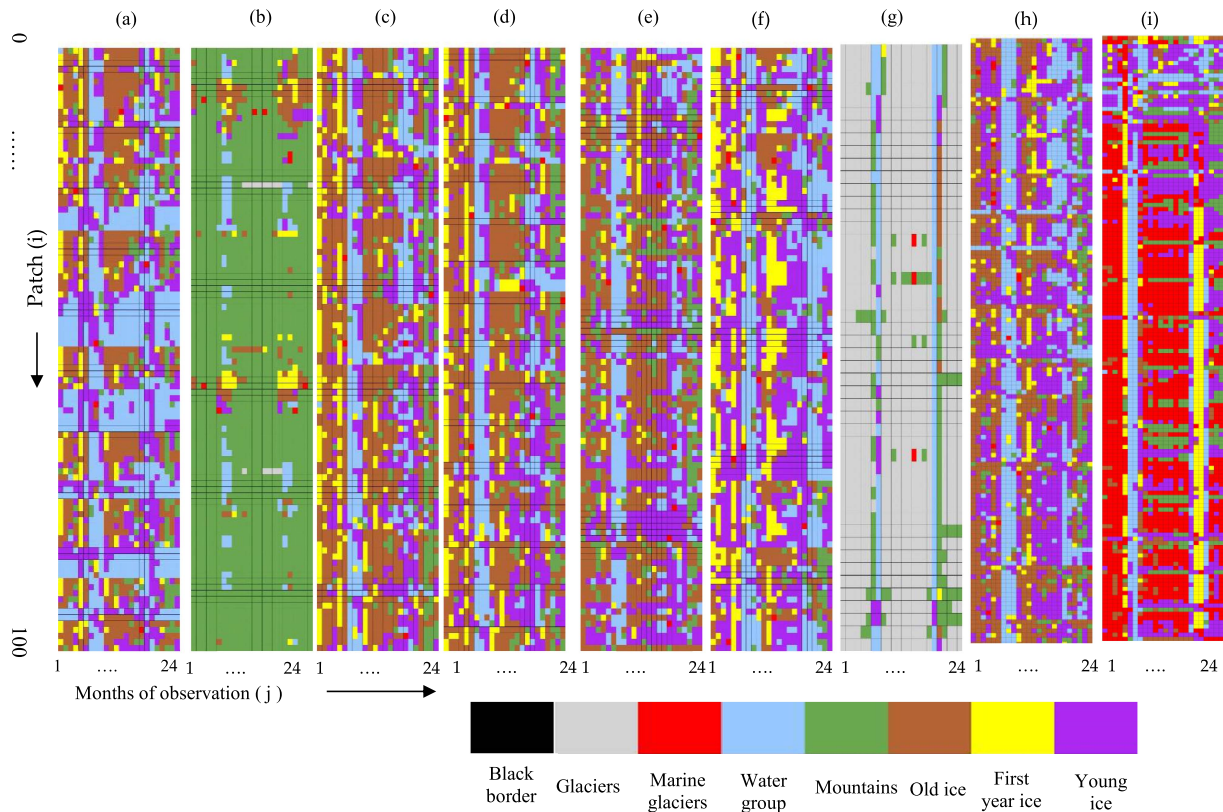
Fig. 8.   Change signatures describing the change classes (see Fig. 6). (a) Change signature C.S-0 describes change class C.C-0, a fair portion of the patches in C.C-1 changed to "water group" during summer months. (b) Change signatures C.S-1 describes the changes in C.C-1, which is vividly different from and more homogeneous than other change classes. (c) C.S-2 describes change class C.C-2. (d) C.S-4 describes C.C-4. (e) C.S-5 describes C.C-5. (f) C.S-8 describes C.C-8. (g) C.S-9 describes C.C-9. (h) C.S-10 describes class C.C-10. (i) C.S-11 describes class C.C-11.

pixels) used for the experiment, displays the quick-look chosen by the user, and allows the user to click on any macropatch of the displayed quick-look image of the whole scene. Clicking on a macropatch (each macropatch is a clickable button) sends a query to the database with respective *patch_ID* (used as an identifying field in the database). The architecture is shown in Fig. 7.

A sample result is shown in Fig. 9. The user chooses to show the quick-look image of June 2019. When the user clicks on a macropatch (in red), the tool shows the ID of the patch, and on clicking OK, the system loads the times series of the macropatch, along with semantic labels and a diagram showing the KL divergence of the BoWs features of the macropatch for 24 months (see Fig. 10). An ice expert's knowledge was used to assess the quality of the results as stated below.

D. Domain Expert's Feedback

Examining the spatial distribution of change classes (C.C.s) and their temporal signatures (C.S.s), we see that the SITS analysis clearly delineates different sea-ice and surface cover regimes that can be identified through their geographical location in Fig. 7, and their temporal pattern in Fig. 8.

C.S.-0 [Fig. 8(a)] is located in areas of sea-ice and open water associated with fjord environments. The presence of old ice in

these areas still needs further investigation, we expect to see more young and first-year ice.

C.S.-1 [Fig. 8(b)] over land areas shows regions covered by bare rock and dry snow during most of the year. The summer melt period leads to a wet snow surface which cause the open water class to occur in some of these areas during summer months.

C.S.-2 [Fig. 8(c)], C.S.-4 [Fig. 8(d)], and C.S.-5 [Fig. 8(e)] show a similar mixture of sea-ice classes and open water associated with areas dominated by drifting sea-ice. However, it is not clear to us why there is such an unusual occurrence of *mountain* surface class in the autumn of 2019 (also in C.S.-0), a further investigation may be needed here. We suspect that the phenomenon can be caused by some classification error.

C.S.-8 [Fig. 8(g)] has more open water and new ice. These categories are located in areas (Fig. 7) that do not have landfast and drifting sea-ice. Fig. 8 is shown with south at the top and north at the bottom. The northern area is the north-east water polynya that opens up to the north of the Norske Øer Ice Barrier. The southern area is open water associated with a later freeze-up and earlier break-up of sea-ice that is formed *in situ* (new and first year ice classes) over the deep water of the Norske Øer trough. *Old ice* and *marine glaciers* can also drift through these areas.

C.S.-9 [Fig. 8(h)] shows mountain glacier areas, with similar summer melt pattern to C.S.-1 [Fig. 8(b)].
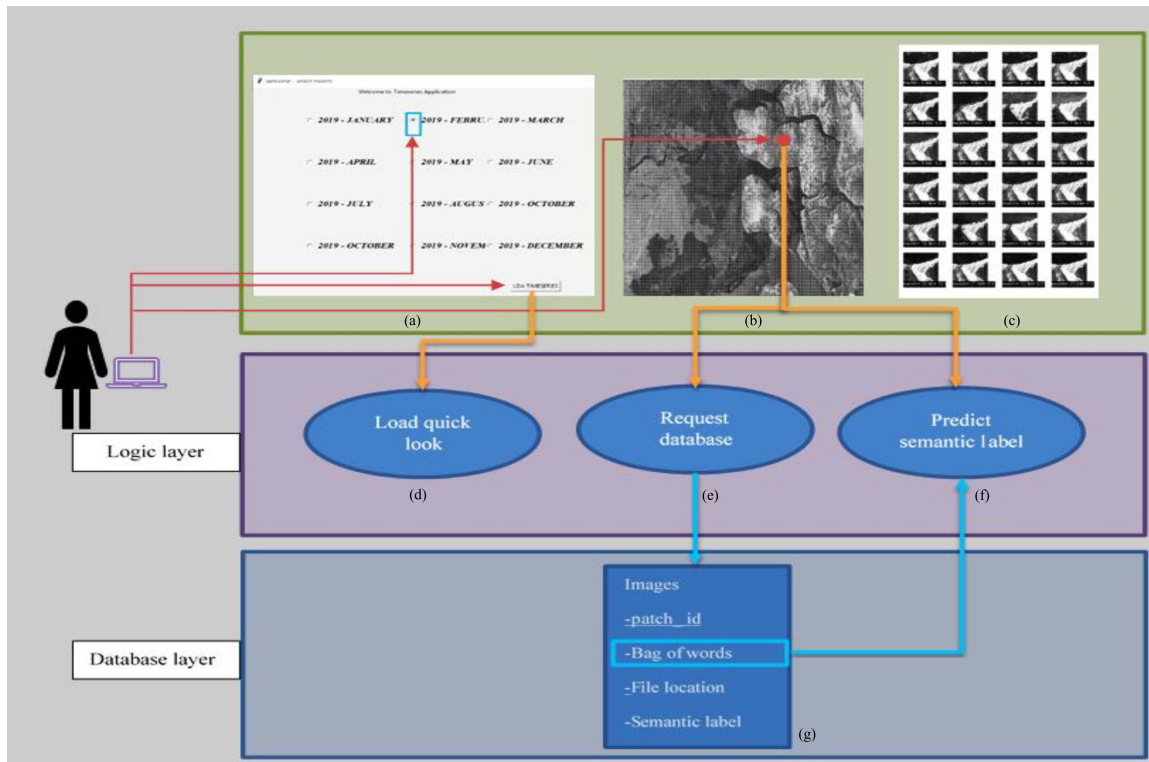
Fig. 9. Human interaction in the proposed visualization tool. (a) On the first screen, the user chooses the month. (b) Respective quick-look image of the chosen month, user selects a macropatch. (c) 24-month quick-look images of the chosen macropatch. (d) Software module to load the quick-look images chosen by the user. (e) Software module for the database requests for further details of the macropatch chosen for the chosen month. (f) Software module to the predict land cover class. (g) Schema of the database table, identified by patch_ID as the primary key.

C.S.-10 [Fig. 8(h)] is an area of landfast sea-ice associated with the Norske Øer Ice Barrier. This should show very little change throughout most of the year, with open water and mixed ice classes during late summer and autumn.

C.S.-11 [Fig. 8(i)] is the area of marine glacier ice associated with the Nioghalvfjerdsbrae (79 N Glacier). This should show a similar temporal pattern to Fig. 8(b).

We can evaluate the periods of melting conditions by reference to the Danish Meteorological Institute weather station at Henrik Krøyer Holme. It is located at 80°38′0″N, 13°43′ 0″W and data can be obtained through https://www.dmi.dk/vejrarkiv/. This shows that the period of melt identified do correspond to high temperatures in the area.

VII. DISCUSSION

The predicted semantic image labels obtained from SITS by applying our proposed workflow created a 24-label sequence of semantic labels for each macropatch which was then analyzed using LDA. A careful observation of the classification results and the change map shows that different surface cover classes experienced different degrees of changes during the study months. For example, the change map obtained from LDA, when compared with the classification map from [2] clearly segregates the mountain regions (surface cover class Mountains) from the rest. The classification results and visualizations obtained from the tool confirm the fact that the mountain

regions are the most unchanging regions during the entire year.

Please note that there are much less changes in the class C.C-1 and in class C.C-9 (Fig. 7). We also know from our classification results that most macropatches in regions labeled as C.C-1 were classified as *mountains*, and most macropatches in regions of C.C-2 were labeled as *glacier*. These finding are verified by the signatures C.S-1 and C.S-9. C.S-1 shows a large portion of the semantic label *mountains*, in parallel with the appearance of *water group*, *first-year ice*, and *old ice* in the summer months. The appearance of *water group* is most obvious in C.S-9; again, during the summer months of both years, the light blue spots almost forming two straight lines through the plot of C.S-9 indicate that the macropatches were classified as *water group* in the summer of both study years. Similar kinds of patterns are observable in other signatures as well. This not only demonstrates the dynamics in the change classes, but also verifies the classification of SITS with physical phenomena. C.S-3 (not shown here) shows a completely uniform color as there are no changes in the black border class.

When we combine the information on the size of each change class and its signature, we obtain another level of understanding. The change class C.C-1 spans 41% of the study area (labeled as *mountains* by the authors of [2]) and its signature C.S-1 [see Fig. 8(b)] shows a rather uniform pattern meaning that the mountain areas did not change much, due to static bare rock and snow cover, except in the summer months. During summer
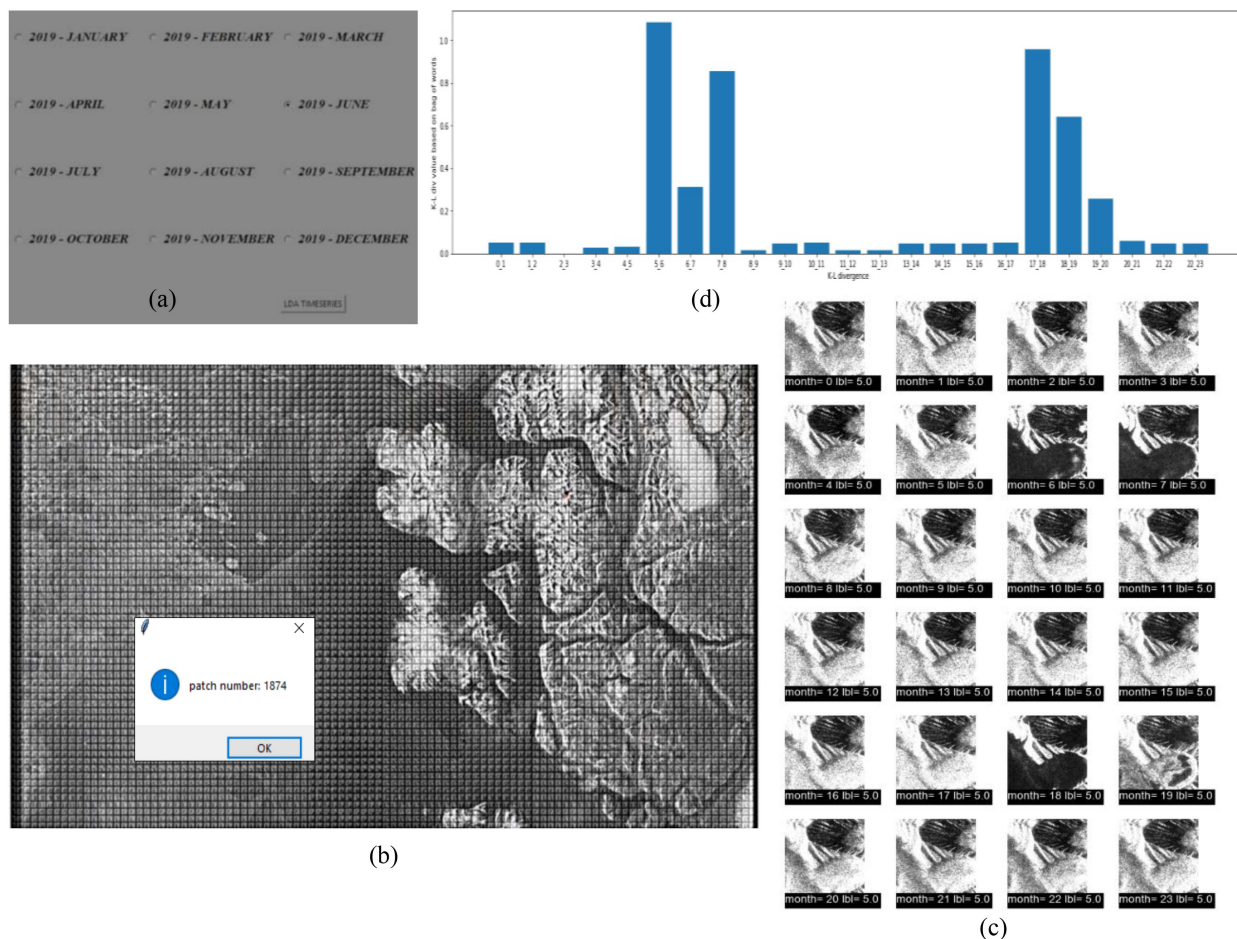
Fig. 10. Flow of user interaction in verifying model results. (a) First screen appears after starting the program, asking the user to choose one of the study months (only 1 year is shown due to space limitations). (b) Corresponding quick-look of the chosen month is loaded with clickable grid distinctly covering each macropatch, when the user clicks on a chosen location, the system shows the macropatch-ID. (c) Quick-look image of 24 months (for the selected patch) are displayed, with predicted semantic labels below the image. (d) KL-divergence between contiguous months (they are computed and displayed by the system).

of both years, the change signature shows occurrence of water group, due to melting of the surface.

The second largest class is C.C-10 [Fig. 8(a)]. C.C-10 covers the areas that were found to be *first-year ice*, *water body*, and *floating ice* in the reference classification. In our result, signature C.S-10 describes the area as sequences of *old ice*, *first-year ice*, and *water group* definitely occurring during the summer months.

Change class C.C-0 spans areas including those labeled as *first-year ice*.

We can guess that the ice areas would undergo more changes than the areas covered by mountains; the signature C.S-0 supports our intuition by showing less homogeneous patterns.

There, we see a great amount of water group even beyond the summer months, they appear all over the year, which means, the areas that were classified as first-year ice in April 2018 underwent some evolution as described by C.S-0 over the 24 months of study. Thus, we can explain the land cover changes and also verify them with the visualizations obtained from the tool.

An example can be given to demonstrate how to draw a conclusion with the visualizations from the user-interface. Fig. 9

shows some patches. We have selected these patches to cover different semantic classes. Patch _ID-81 is a macropatch in the mountain region. The results from the user-interface tool shows the quick-look of the patch alongside the 24 semantic labels predicted from the classier. We see from the results of the user-interface in Fig. 9 that the quick-look of patch-81 does not undergo much changes except in summer months, the semantic labels are also consistent with the quick-look images, i.e., the same semantic label does not appear with much differing quick-look images. This serves as a verification for the classification results.

Next, if we compare the change map in Fig. 8 and find that patch-81 would be in change class C.C-1 which is described by signature C.S-1. Signature C.S-1 is mostly of a homogeneous color indicting very less changes compared to other change classes. This example demonstrates how the tool helps us to verify the results. A more detailed discussion of such examples is presented in Section III.

C.S.-2, C.S.-4, and C.S.-5 have a very similar pattern of sea-ice and open water classes, and can be identified as drifting sea-ice.

```
#Initial setting:
        # M scenes are tiled into N macropatches of size (256×256)
        # The BoW function converts a macropatch into its bag-of-words representation
        #The Classify function predicts the semantic label of the given sample by applying the given pre-trained classifier
FOR i IN (0, N):
        sample = BoW(macropatchᵢ)
        label = Classify (classifier1, sample)
        IF label == 0:
predicted_label='Black Border'
Else :
label = Classify(classifier2,sample)
IF (label == 0) :
predicted_label ='Water Group'
ELSE:
label = Classify (classifier3, sample)
IF (label == 0):
predicted_label ='Glacier'
ELSE:
label = Classify(classifier4, sample)
IF (Label == 0) :
predicted_label = 'Young Ice in Sea'
ELSE:
label= Classify (classifier5, sample)
IF (label == 0): #sample is in group (Icebergs, Mountains)
label =Classify (classifier6, sample)
IF (label ==0):
predicted_label ='Icebergs'
ELSE:
predicted_label ='Mountains'
ELSE:
label= Classify (classifier7, sample):
IF (label ==0):
predicted_label = 'Old Ice'
ELSE:
predicted_label = 'First Year Ice'

RETURN predicted_label
```

Fig. 11. Semantic class prediction algorithm.

VIII. CONCLUSION

The framework presented by the authors has offered a new approach for SITS analysis, which aims to solve some unaddressed issues by existing researches. The specific problems addressed were the lack of labeled information in SITS studies and the lack of transparency occurred in such studies. Another issue is scalability of such methods as SITS datasets are usually huge. We propose a framework for SITS analysis that is explainable, in terms that every ML model used in the chain is transparent [9], integrates domain knowledge [6], and the intermediate and final outcomes are made to be interpretable with the help of visualizations. Not just that, considering the practical problems with large datasets in SITS studies, we have also proposed a scalable dictionary computation strategy for LDA change modeling.

The idea behind the proposed workflow is to first extend the limited set of labeled data in hand and then to discover latent spatiotemporal patterns. A classification algorithm extends the existing labeled dataset to create labeled dataset for a much longer period of time. The resulting semantic labels are then stacked to create temporal sequences of semantic labels for each spatial unit. LDA being a textual model, discovers patterns in these sequences, which becomes the new spatiotemporal label for each image pixel (see Fig. 8). In essence, the image pixels are no more static information but contains evolution dynamics over the study period.

The case study is conducted in the Polar region. We extended scenes from three study months to create an SITS spanning 24 months. A change map resulting from the LDA change model shows the evolution of the study area. Interestingly, it reflects the geophysical truth that sea areas have undergone more diverse changes while the mountain regions have stayed rather stable. The results are compared with classification map from [2] and also verified by a domain expert (see Section VI).

However, there are challenges emerging in the proposed framework. First, it may need a lot of manual effort to visually verify the intermediate results; an intelligent tool may help reduce effort. Second, the gigantic amount of engineering tasks in implementing the framework and the visualization tool has to be justified with the importance of the study.

In the frame of the ExtremeEarth project, we plan to increase the volume of data and to select different locations. As future work, we plan to incorporate a statistical approach to measure the differences between the signatures. However, given the importance of visualization in RS and also in explainability of artificial intelligence methods, our method may support further articles in this direction, without much regard to the area of study.
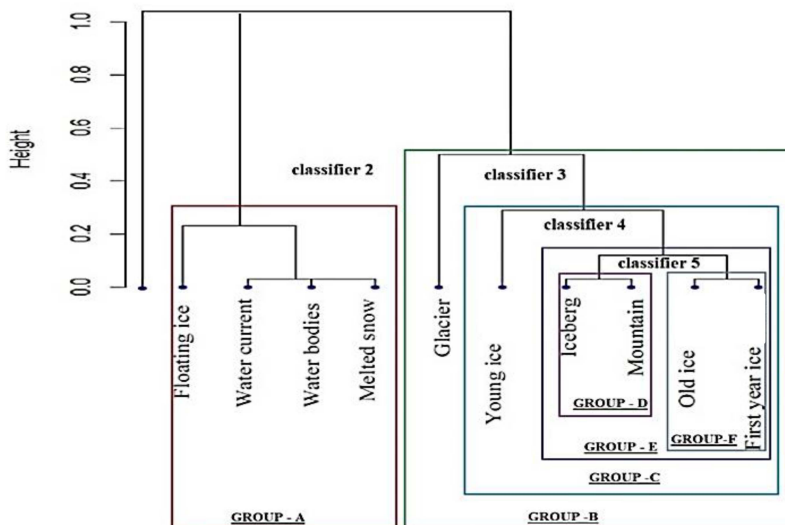
Fig. 12. Class grouping domain knowledge.

	precision	recall	f1-score	support
1	0.95	0.93	0.94	1784
2	0.89	0.92	0.91	1076
accuracy			0.93	2860
macro avg	0.92	0.93	0.92	2860
weighted avg	0.93	0.93	0.93	2860

Fig. 13. Classification report of classifier 5 in Fig. 12.

Predicting land cover class on the test set

	precision	recall	f1-score
Mountains	0.87	0.83	0.85
Water_bodies	0.91	0.80	0.85
Icebergs	0.92	0.85	0.88
Sea3_young_ice	0.69	0.60	0.64
Sea2_First_year_ice	0.76	0.87	0.81
Sea1_Old_ice	0.78	0.78	0.78
Water_Ice_Currents	0.86	0.78	0.82
Floating_Ice	0.78	0.95	0.86
Mealted_Snow_Ice_in_Mountains	0.60	0.75	0.67
Glacier	0.78	0.64	0.70
accuracy			0.80
macro avg	0.79	0.79	0.79
weighted avg	0.81	0.80	0.80

Fig. 14. Classification report of multiclass SVM.

Predicting land cover class on the test set

	precision	recall	f1-score	support
Mountains	0.83	0.79	0.81	24
Water_bodies	0.91	0.80	0.85	25
Icebergs	0.96	0.85	0.90	27
Sea3_young_ice	0.59	0.67	0.62	15
Sea2_First_year_ice	0.79	0.73	0.76	15
Sea1_Old_ice	0.65	0.83	0.73	18
Water_Ice_Currents	0.68	0.74	0.71	23
Floating_Ice	0.83	0.86	0.84	22
Mealted_Snow_Ice_in_Mountains	0.79	0.75	0.77	20
Glacier	0.70	0.64	0.67	11
accuracy			0.78	200
macro avg	0.77	0.77	0.77	200
weighted avg	0.79	0.78	0.78	200

Fig. 15. Classification scores of a random forest classifier.

APPENDIX

We designed a rule-based classification concept that uses class distinction knowledge from another research project. The logic of the classifier is presented as pseudocode in Fig. 11. The class distinction diagram and our nomenclature of class groups is depicted in Fig. 12 and as pseudocode in Fig. 11.

ACKNOWLEDGMENT

REFERENCES

[1] T. John and J. Overland, "Contrasting climate change in the two polar regions," *Polar Res.*, vol. 28, no. 2, pp. 146–164, 2009.

[2] C. O. Dumitru, V. Andrei, G. Schwarz, and M. Datcu, "Machine learning for sea ice monitoring from satellites. International archives of the photogrammetry," *Remote Sens. Spatial Inf. Sci.*, vol. XLII-2/W16, pp. 83–89, 2019.

[3] S. Cui, G. Schwarz, and M. Datcu, "Remote sensing image classification: No features, no clustering," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 11, pp. 5158–5170, Nov. 2015.

[4] A. Julea et al., "Unsupervised spatiotemporal mining of satellite image time series using grouped frequent sequential patterns," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 4, pp. 1417–1430, Apr. 2011.

[5] D. Sacha et al.et al., "What you see is what you can change: Human-centered machine learning by interactive visualization," *Neurocomputing*, vol. 268, pp. 164–175, 2017.

[6] R. Roscher, B. Bohn, M. F. Duarte, and J. Garcke, "Explainable machine learning for scientific insights and discoveries," *IEEE Access*, vol. 8, pp. 42200–42216, 2020.

[7] C. McCue, *Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis*. Oxford, U.K.: Butterworth-Heinemann, 2014.

[8] C. Vaduva, T. Costachioiu, C. Patrascu, I. Gavat, V. Lazarescu, and M. Datcu, "A latent analysis of earth surface dynamic evolution using change map time series," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2105–2118, Apr. 2013, doi: 10.1109/TGRS.2012.2219316.

[9] C. Karmakar, C. O. Dumitru, G. Schwarz, and M. Datcu, "Feature-free explainable data mining in SAR images using latent dirichlet allocation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 676–689, 2021.

[10] K. Ekaterina, D. Ienco, J. Sublime, and M. Trocan, "Unsupervised change detection analysis in satellite image time series using deep learning combined with graph-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1450–1466, 2020.

[11] S. Singha, M. Johansson, N. Hughes, S. M. Hvidegaard, and H. Skourup, "Arctic sea ice characterization using spaceborne fully polarimetric L-, C-, and X-Band SAR with validation by airborne measurements," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 3715–3734, Jul. 2018.

[12] W. Tan, J. Li, L. Xu, and M. A. Chapman, "Semiautomated segmentation of sentinel-1 SAR imagery for mapping sea ice in labrador coast," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1419–1432, May 2018.

[13] U.; Maulik and D. Chakraborty, "Remote sensing image classification: A survey of support-vector-machine-based advanced techniques," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 1, pp. 33–52, Mar. 2017.

[14] X. M. Li, Y. Sun, and Q. Zhang, "Extraction of sea ice cover by sentinel-1 SAR based on support vector machine with unsupervised generation of training data," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3040–3053, Apr. 2021.

[15] M. Lienou, H. Maitre, and M. Datcu, "Semantic annotation of satellite images using latent dirichlet allocation," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 1, pp. 28–32, Jan. 2010.

[16] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Proc. IEEE 9th Int. Conf. Comput. Vis.*, 2003, pp. 1470–1477.

[17] M. Lienou, H. Maitre, and M. Datcu, "Semantic annotation of satellite images using latent dirichlet allocation," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 1, pp. 28–32, Jan. 2010.

[18] S. Xu, T. Fang, D. Li, and S. Wang, "Object classification of aerial images with bag-of-visual words," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 2, pp. 366–370, Apr. 2010.

[19] H. Sun, X. Sun, H. Wang, Y. Li, and X. Li, "Automatic target detection in high-resolution remote sensing images using spatial sparse coding BoW model," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 1, pp. 109–113, Jan. 2012.

[20] L. Gueguen and M. Datcu, "Image time-series data mining based on the information-bottleneck principle," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 827–838, Apr. 2007.

[21] M. K. Pakhira, "A linear time-complexity k-means algorithm using cluster shifting," in *Proc. IEEE Int. Conf. Comput. Intell. Commun. Netw.*, 2014, pp. 1047–1051.

[22] J. Wang and X. Su, "An improved K-means clustering algorithm," in *Proc. IEEE 3rd Int. Conf. Commun. Softw. Netw.*, 2011, pp. 44–46.

[23] T. Hofmann, B. Schölkopf, and A. Smola, "Kernel methods in machine learning," *Ann. Statist.*, vol. 36, no. 3, pp. 1171–1220, 2008.

[24] A. Hadid and M. Pietikäinen, "Face recognition with local binary patterns," in *Computer Vision - ECCV 2004. ECCV 2004. Lecture Notes in Computer Science*, vol. 3021, T. Pajdla and J. Matas, Eds. Berlin, Germany: Springer.

[25] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.